Unwinding Stochastic Order Flow

SSRN: 4609588

Marcel Nutz
Columbia University

January 2024
Joint work with

Kevin Webster

Long Zhao
Motivation: Central Risk Book (CRB)

- Centralized trading unit recently established in many large banks, trading firms

CRB aggregates order flow from other business units within the organization (in-flow)

- In-flow is a stochastic process. Uncontrolled.
- CRB nets opposite orders (internalization)
- Unwinds outstanding positions in the market (out-flow)

→ Optimal execution problem for a stochastic position

- Another use case: market maker aggressively unwinding inventory
Some Stylized Facts to Capture

- FX dealers achieve high double-digit internalization rates
  Bank of England ’14

- For institutional size orders in large cap stocks,
  ≈ 30 bps of price impact costs
  ≈ 5 bps of instantaneous costs
  Nasdaq Guide for Trading Interns ’22

- Intraday impact decay: half of a trade’s impact dissipates over 1-2 h
  Horst et al. ’19, Muhle-Karbe et al. ’22

- Price impact parameters vary intraday
  Cont et al. ’13, Fruth et al. ’13

- Volume in closing auction is a double-digit percentage of total
  Bouchaud et al. ’18
Deliverables

- Flexible model that can be calibrated to a particular flow
- Principled, implementable unwind strategy
  Guarantee for no-price-manipulation
  Formula for expected cost
- Sensitivity to in-flow characteristics
  Misspecification analysis
Outline

1 Model and Synopsis of Results

2 Solution

3 Numerical Simulations
Model

- Finite horizon $T > 0$ (one trading day)
- Cumulative in-flow orders: process $Z$, 
  \[ \frac{dZ_t}{dt} = -\theta Z_t + \sigma dW_t, \quad Z_0 = z \]
- Can be mean-reverting or trending. Historical internal data
- Model is tractable for general martingale driver

- Cumulative out-flow process $Q$ is our control
- Choose $Q$ such as to minimize expected transaction cost
- Liquidation constraint $Q_T = Z_T$
Transaction Costs

Two reasons to hold inventory:
- Increase chance of netting, reduce spread cost
- Unwind slowly to reduce impact

Transaction cost modeling:
- $S_t$ “unaffected" market price
- Assumption: $S$ is a martingale
- $P_t = P_t(Q, S)$ actual execution price
- Execution cost
  $$\int_0^T P_t \, dQ_t$$
Instantaneous cost

- Execution price \( P_t = S_t + \frac{1}{2} \varepsilon \dot{Q}_t \) with \( \varepsilon > 0 \)
- Execution cost
  \[
  \int_0^T P_t \, dQ_t = \int_0^T S_t \, dQ_t + \frac{1}{2} \varepsilon \int_0^T \dot{Q}_t^2 \, dt.
  \]

- Imposes absolutely continuous trading
  
  Bertsimas & Lo ’98, Almgren & Chriss ’01, …

Persistent (decaying) impact

- Impact process \( dY_t = -\beta Y_t \, dt + \lambda \, dQ_t \)
- Execution price \( P_t = S_t + \frac{1}{2} (Y_t^- + Y_t) \)
- Execution cost
  \[
  \int_0^T P_t \, dQ_t = \int_0^T S_t \, dQ_t + \int_0^T Y_t^- \, dQ_t + \frac{1}{2} [Y, Q]_T.
  \]

  Obizhaeva & Wang ’13, …
Transaction Cost Model

- **Regular LOB trading** for \( t \in (0, T) \)
- **Both** instantaneous cost and persistent impact cost (permanent cost)
  Graewe & Horst ’17
- **Absolutely continuous trading during the day**, \( t \in (0, T) \)

- Opening auction (cross) at \( t = 0 \)
- Closing auction (cross) at \( t = T \)
- No instantaneous cost (no spread), only impact cost
- \( \rightarrow \) Block trades in the auctions

- Generic trading strategy \( Q \) is a triplet \( (J_0, (q_t), J_T) \):
  \[
  J_0 = \Delta Q_0, \quad q_t = \dot{Q}_t, \quad t \in (0, T), \quad J_T = \Delta Q_T
  \]
Impact process

\[ dY_t = -\beta Y_t \, dt + \lambda \, dQ_t, \quad Y_{0-} = y \]

Execution price

\[ P_t := \begin{cases} 
S_t + \frac{1}{2}(Y_{t-} + Y_t), & t \in \{0, T\} \\
S_t + Y_t + \frac{1}{2} \varepsilon q_t, & t \in (0, T). 
\end{cases} \]

Expected execution cost

\[ C(J_0, q, J_T) = \mathbb{E} \left[ \int_0^T P_t \, dQ_t \right] = \mathbb{E} \left[ P_0 J_0 + \int_0^T P_t q_t \, dt + P_T J_T \right] \]

Problem: minimize \( C(J_0, q, J_T) \) subject to \( Q_T = Z_T \)
Synopsis

- Semi-explicit optimal strategy
- Intra-day trading speed has feedback-form

\[ q_t = f_t X_t + g_t Y_t + h_t Z_t \]

\( X_t \) ... inventory, \( Y_t \) ... impact state, \( Z_t \) ... cum. in-flow

- Time-dependent coefficients \( f_t, g_t, h_t \) determined by ODE
  (closed form for constant liquidity parameters)
- \( h_t \) is the adjustment to stochastic in-flow
Synopsis

Having a computable model allows us to study (model-dependent) trading metrics from an input–output perspective.

Core trading metrics strongly depend on in-flow characteristics:

- Momentum requires more aggressive trading, increasing costs. Reversion leads to more internalization and more warehousing.
- Misspecification: overestimating momentum sharply increases costs.
- Expected trading costs (per order notional) are minimized at a particular in-flow volatility.
Example Paths

Two realizations of the same model ($\theta = 0$) illustrating extreme regimes.
Outline

1 Model and Synopsis of Results

2 Solution

3 Numerical Simulations
Solution Strategy

- Fix arbitrary $J_0$ and $q$: outstanding position before closing auction has to be the final block trade: $J_T = Q_T - Z_T$
- Replace liquidation constraint by cost of that block trade

- After reformulation, that cost is a function of the terminal states: standard LQ control problem

$\rightarrow$ Solution $q^*$ and cost $C(J_0, q^*, J_T^*)$

- Finally, optimize that cost $C(J_0, q, J_T)$ over $J_0$
Rewriting the Execution Cost

Let \((J_0, q, J_T) \in A\). Define state processes \((X_t, Y_t^c, Z_t)_{t \in [0, T]}\) by

\[
\begin{align*}
    dX_t &= q_t \, dt - dZ_t, & X_0 &= J_0 - z & \text{outstanding position} \\
    dY_t^c &= (-\beta Y_t^c + \lambda q_t) \, dt, & Y_0^c &= y + \lambda J_0 & \text{impact state} \\
    dZ_t &= -\theta Z_t \, dt + \sigma_t \, dW_t, & Z_0 &= z & \text{cumulative in-flow}
\end{align*}
\]

The expected execution cost is

\[
C(J_0, q, J_T) = \frac{1}{2} \mathbb{E} \left\{ \int_0^T \left[ \frac{2\beta}{\lambda} (Y_t^c)^2 + \varepsilon q_t^2 \right] \, dt + \frac{1}{\lambda} (Y_T^c - \lambda X_T)^2 \right. \\
\left. - \frac{y^2}{\lambda} + S_T Z_T \right\}
\]
Dynamic Programming

- Start at time $t$, initial states $(x, y, x)$:

$$
\begin{align*}
  dX_s &= q_s \, ds - dZ_s, & X_t &= x \quad \text{outstanding position} \\
  dY^c_s &= (-\beta Y^c_s + \lambda q_s) \, ds, & Y^c_t &= y \quad \text{impact state} \\
  dZ_s &= -\theta Z_s \, ds + \sigma \, dW_s, & Z_t &= z \quad \text{cumulative in-flow}
\end{align*}
$$

- Auxiliary LQ problem

$$
\nu(t, x, y, z) = \inf_{q} \frac{1}{2} \mathbb{E} \left\{ \int_t^T \left( \frac{\beta}{\lambda} (Y^c_s)^2 + \frac{1}{2} \varepsilon q_s^2 \right) \, ds + \frac{1}{2\lambda} (Y^c_T - \lambda X_T)^2 \right\}
$$

- Quadratic value function $\nu$

- Linear optimal feedback control: $q^*(t, x, y, z) = f_t x + g_t y + h_t z$
Time-varying Market Parameters

- Liquidity increases over the course of the trading day (regular hours)
  \[ \lambda_t \]
- Auctions empirically more liquid than adjacent regular trading
- Also make \( \varepsilon_t, \beta_t \) time-dependent (bounded)
- \( \lambda_t, \varepsilon_t \) bounded away from zero

Fast increase in liquidity can give rise to arbitrage ("price manipulation")
Huberman & Stanzl '04

- E.g., suppose \( \beta, \varepsilon \equiv 0 \) and \( \lambda_{t_0} > \lambda_{t_1} \) for some \( t_0 < t_1 \)
  \[ \text{Buy at } t_0 \text{ and sell at } t_1. \text{ Or: Sell at } t_0 \text{ and buy at } t_1. \]
- In general, trade-off between \( \beta \) and \( \lambda \)
Parameter Restrictions

- \( \lambda \) differentiable (regular hours)
- \( \lambda_0^- \) for opening auction

**Assumption:**

\[
2\beta_t + \dot{\gamma}_t > 0 \quad (\dot{\gamma}_t := \frac{\dot{\lambda}_t}{\lambda_t}) \quad \text{and} \quad \lambda_0^- \leq \lambda_0
\]

Fruth, Schöneborn & Urusov '14

- We use \( \lambda_T \) for the closing auction (liquidity jump causes arbitrage)
Execution Cost

For any \((J_0, q, J_T) \in \mathcal{A}\), the expected execution cost is

\[
C(J_0, q, J_T) = \frac{1}{2} \mathbb{E}\left\{ \int_0^T \left[ \frac{2\beta_t + \dot{\gamma}_t}{\lambda_t} (Y^c_t)^2 + \varepsilon_t q_t^2 \right] dt + \frac{1}{\lambda_T} (Y^c_T - \lambda_T X_T)^2 \right. \\
+ \left. (Y^c_0)^2 \left[ \frac{1}{\lambda_{0-}} - \frac{1}{\lambda_0} \right] - \frac{y^2}{\lambda_0} + S_T Z_T \right\},
\]

where \((X_t, Y^c_t, Z_t)_{t \in [0, T]}\) are defined by

\[
\begin{aligned}
\begin{cases}
dX_t = q_t \, dt - dZ_t, & X_0 = J_0 - z \\
dY^c_t = (-\beta_t Y^c_t + \lambda_t q_t) \, dt, & Y^c_0 = y + \lambda_{0-} J_0 \\
dZ_t = -\theta_t Z_t \, dt + \sigma_t \, dW_t, & Z_0 = z
\end{cases}
\end{aligned}
\]

\(\lambda_{0-} \leq \lambda_0\) and \(2\beta_t + \dot{\gamma}_t > 0\) make problem (strictly) convex
(No) Price Manipulation

There is no “price manipulation” strategy:

**Corollary:** Let $Z \equiv 0$ and $y = 0$. Then $Q \equiv 0$ is the only admissible strategy with zero cost.

**Remark:**

- Starting with **non-zero inventory**, round trips could nevertheless be profitable: “transaction-triggered price manipulation”
  
  Alfonsi, Schied & Slynko ’11, Gatheral, Schied & Slynko ’12

- Starting with **non-zero impact state**, it will often be optimal to trade, even if $Z \equiv 0$
Solution on \((0, T)\) for fixed \(J_0\)

Fix \(t \in [0, T)\) and consider \((X_s, Y_s^c, Z_s)_{s \in [t, T]}\) started at \((x, y, z)\). Then

\[
v(t, x, y, z) := \inf_q \frac{1}{2} \mathbb{E} \left\{ \int_t^T \left[ \frac{2\beta_s + \gamma_s}{\lambda_s} (Y_s^c)^2 + \varepsilon_s q_s^2 \right] ds \right. \\
+ \left. \frac{1}{\lambda_T} (Y_T^c - \lambda_T X_T)^2 \right\}
\]

is of the form

\[
v(t, x, y, z) = \frac{1}{2} A_t x^2 + B_t x y + \frac{1}{2} C_t y^2 + D_t x z + E_t y z + \frac{1}{2} F_t z^2 + K_t
\]

where \(A_t, B_t, \ldots, K_t\) are defined below.
Riccati ODE System

**Proposition:** The Riccati ODE system

\[
\begin{align*}
\dot{A}_t &= \varepsilon_t^{-1}(A_t + \lambda_t B_t)^2, & A_T &= \lambda_T \\
\dot{B}_t &= \varepsilon_t^{-1}(A_t + \lambda_t B_t)(B_t + \lambda_t C_t) + \beta_t B_t, & B_T &= -1 \\
\dot{C}_t &= \varepsilon_t^{-1}(B_t + \lambda_t C_t)^2 + 2\beta_t C_t - \lambda_t^{-1}(2\beta_t + \gamma_t), & C_T &= \lambda_T^{-1} \\
\dot{D}_t &= \varepsilon_t^{-1}(A_t + \lambda_t B_t)(D_t + \lambda_t E_t) - \theta_t(A_t - D_t), & D_T &= 0 \\
\dot{E}_t &= \varepsilon_t^{-1}(B_t + \lambda_t C_t)(D_t + \lambda_t E_t) - \theta_t(B_t - E_t) + \beta_t E_t, & E_T &= 0 \\
\dot{F}_t &= \varepsilon_t^{-1}(D_t + \lambda_t E_t)^2 - 2\theta_t(D_t - F_t), & F_T &= 0 \\
\dot{K}_t &= -\sigma_t^2(A_t - 2D_t + F_t)/2, & K_T &= 0
\end{align*}
\]

has a unique solution \((A_t, B_t, C_t, E_t, F_t, K_t)_{t \in [0, T]}\).

Wonham '68, Graewe & Horst '17
Optimal Strategy

The unique optimal control (in feedback form) is

\[ q^*(t, x, y, z) = f_t x + g_t y + h_t z \]

where

\[ f_t := -\varepsilon_t^{-1}(A_t + \lambda_t B_t) \]
\[ g_t := -\varepsilon_t^{-1}(B_t + \lambda_t C_t) \]
\[ h_t := -\varepsilon_t^{-1}(D_t + \lambda_t E_t) \]

- \( f, g \) only depend on market parameters, not on in-flow characteristics
- \( h \) also depends on mean reversion \( \theta \), but not on \( \sigma \)
- \( h \) is the adjustment for random in-flows
- \( \sigma \) only affects the cost

Closed-form solution for time-independent liquidity parameters
Properties of the Optimal Strategy

\[ q^*(t, x, y, z) = f_t x + g_t y + h_t z \]

- \( h = 0 \) if \( \theta \equiv 0 \) (martingale in-flow)
- \( h \) is the adjustment to in-flows with reversion/momentum
- \( h \) is monotone decreasing wrt. in \( \theta \)
- \( h \geq 0 \) if \( \theta \leq 0 \) (momentum): overtrading
- \( h \leq 0 \) if \( \theta \geq 0 \) (mean-reversion)

Under the additional condition \( \beta + \dot{\gamma} > 0 \) (vs. \( 2\beta + \dot{\gamma} > 0 \)):

- \( g \leq 0 \): high impact state incentivizes selling and vice versa
Properties of the Optimal Strategy (2)

\[ q^*(t, x, y, z) = f_t x + g_t y + h_t z \]

Again, under $\beta + \dot{\gamma} > 0$:

- $f \leq 0$: high inventory incentivizes selling and vice versa

Whereas:

- If $\beta + \dot{\gamma} < 0 < 2\beta + \dot{\gamma}$, then $f_t > 0$ for $T - t$ small
- $q^*$ trades in the “wrong” direction even if $y = 0$ and no future in-flow
  “transaction-triggered price manipulation”
  Fruth, Schöneborn & Urusov ’14

- In practice, enforce $\beta + \dot{\gamma} > 0$ when using estimated parameters
Optimal Initial Block Trade

**Proposition:** The optimal block trade $J_0$ for the opening auction is

$$J_0 = -\frac{g_0^- + \eta_0^-}{f_0^- + \lambda_0^- (g_0^- + \eta_0^-)} y + \frac{f_0^- - h_0^-}{f_0^- + \lambda_0^- (g_0^- + \eta_0^-)} z,$$

where $\eta_0^- := -\varepsilon_0^{-1}(1 - \lambda_0^-/\lambda_0) \leq 0$,

$$g_0^- + \eta_0^- < 0 \quad \text{and} \quad f_0^- + \lambda_0^- (g_0^- + \eta_0^-) < 0.$$

If $\beta_t + \dot{\gamma}_t > 0$ for $t \in [0, T]$, then $f_0^- - h_0^- < 0$. 
1 Model and Synopsis of Results
2 Solution
3 Numerical Simulations
Trading Metrics

- Key metrics such as impact cost are **model-dependent**
- Having a reduced-form model allows us to study trading metrics from an input–output perspective
- Practitioners express many quantities *per order notional*

⇒ Our experiments use a finite-variation in-flow

- Internalization is a key metric:

  \[
  \text{internalization rate} = 1 - \frac{\text{total variation of out-flow}}{\text{total variation of in-flow}} \\
  \approx \text{fraction of in-flow orders netted}
  \]

- model-independent and client-centric; often used as a proxy for cost
In-Flow Autocorrelation: Sample Paths

Sensitivity to flow autocorrelation ($\theta = -1, 0, 1$) for an initial inventory ($z$) and daily flow volatility ($\sigma$) of 10% ADV: sample path.
### In-Flow Autocorrelation: Averages

<table>
<thead>
<tr>
<th>Parameter</th>
<th>( \theta )</th>
<th>In-flow (ADV%)</th>
<th>Out-flow (ADV%)</th>
<th>Spread (bps)</th>
<th>Cost (bps)</th>
<th>Impact cost (bps)</th>
<th>Closing trade (% total)</th>
<th>Internalization (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>momentum</td>
<td>-1</td>
<td>61</td>
<td>31</td>
<td>4.9</td>
<td>42.6</td>
<td>17</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>martingale</td>
<td>0</td>
<td>46</td>
<td>15</td>
<td>1.7</td>
<td>14.5</td>
<td>21</td>
<td>68</td>
<td></td>
</tr>
<tr>
<td>reversal</td>
<td>1</td>
<td>52</td>
<td>9</td>
<td>0.5</td>
<td>4.8</td>
<td>27</td>
<td>84</td>
<td></td>
</tr>
</tbody>
</table>

- Higher autocorrelation leads to more aggressive unwinding
- Internalization drops from 84% to 51%: the strategy trades 3x more volume on the market
- Impact costs increase from 5bps to 43bps: the strategy pays 8x more trading costs
- The closing jump trade shrinks from 27% to 17% of the outflow: the strategy doesn’t warehouse as much into the close
In-Flow Autocorrelation: Distributions

- **Total variation of in-flow (ADV%)**
- **Internalization (%)**
- **Impact cost per in-flow (bps)**
- **Total variation of out-flow (ADV%)**
- **Internalization regret (%)**
- **Spread cost per in-flow (bps)**

Legend:
- **momentum in-flow**
- **martingale in-flow**
- **reversal in-flow**
In-Flow Volatility

Average metrics’ sensitivity to in-volatility ($\sigma$) for an initial inventory ($z$) of 3% ADV and different autocorrelation values ($\theta$).
In-Flow Volatility

- *some* flow volatility reduces cost per order notional, by promoting internalization
- costs increase with *too much* flow volatility

<table>
<thead>
<tr>
<th>( \theta )</th>
<th>-1</th>
<th>-0.5</th>
<th>0</th>
<th>0.5</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>optimal ( \sigma/z ) (%)</td>
<td>99</td>
<td>79</td>
<td>55</td>
<td>35</td>
<td>17</td>
</tr>
<tr>
<td>Internalization (%)</td>
<td>55</td>
<td>65</td>
<td>72</td>
<td>77</td>
<td>82</td>
</tr>
</tbody>
</table>

Optimal in-flow volatility, as a percentage of inventory, across various \( \theta \)
Misspecification

The P&L depends on $\theta, \sigma$, but the strategy only depends on $\theta$. Therefore, there are no misspecification costs for $\sigma$.

Actual $\theta = 0$: true in-flow is martingale

Overly aggressive trading sharply increases transaction cost and misses netting opportunities. Preferable to underestimate momentum, causing less aggressive trading.
Thank you