Flow Trading

Eric Budish
University of Chicago

Peter Cramton
University of Cologne

Albert S. Kyle
University of Maryland

Mina Lee
Washington Univ. St. Louis

David Malec
University of Maryland

Finance Seminar

Princeton University
February 19, 2025

Motivation for a New Market Design

Description of current financial exchanges for equities, futures, options, bonds, etc.:

- Use mostly standard limit orders—a price, quantity, and direction: "Buy 1000 shares of AAPL at $150.00 per share or better."
- Orders are typically for an individual asset rather than portfolios.
- Orders are processed continuously, one-at-a-time in order of arrival, with incoming "executable" orders matched with "resting" orders in limit order book.
- Orders respect a minimum price increment ("tick size") and a minimum quantity increment ("lot size"). For example, a tick size of $0.01 and lot size of 100 shares.

Historical basis for this mechanism: In the past, trading was face to face with trades registered, cleared, settled with pencil and paper.

- Specialists, trading pits, all the way back to the buttonwood tree.
- Humans can run a continuous limit order book if orders arrive slowly enough.
- Computers now run continuous limit order book markets at modern speeds and order volumes.
- On the order of $1 quadrillion per year.

Motivation for a New Market Design

Clearly passes a "market test." However, there are multiple ways that this market design creates unnecessary costs and complexity, especially given the possibilities enabled by modern computers.

1. Costs related to speed.
   (a) Stale orders risk being "sniped" by high-frequency traders. 20% of trading volume, 17-33% of cost of liquidity.
   (b) Discrete tick size induces queuing and races for time priority.

2. Costs related to trading large quantities over time.
   (a) Institutional investors must place and cancel thousands or even millions of small orders to trade meaningful quantities.
   (b) Trade execution and institutional prime brokerage industries provide these services at significant cost.

3. Costs related to trading portfolios of assets, both long-only and long-short.
   (a) Trading portfolios and maintaining arbitrage pricing relationships requires constantly placing and canceling orders as prices change.
   (b) Indirect evidence of value: ETFs. Now 40% of volume and ~$10 trillion of assets.

Motivation for a New Market Design

One additional “canary in the coalmine” statistic that is suggestive that the modern exchange design may merit a rethink:

Over 50% of trading volume in the US equities market now takes place *off*-exchange.

Flow Trading: Combination of Three Ideas

This paper proposes a new market design for financial assets. Our design takes advantage of advances in modern computing—It could not be run under the buttonwood tree—and addresses these three sources of unnecessary costs and complexity.

Our design combines three ideas, complementary in a "nesting doll" sense:

1. Frequent Batch Auctions: markets are cleared in discrete-time batch auctions, held at intervals such as once per second (Budish, Cramton and Shim, 2015).

2. Continuous-scaled limit orders: Piecewise-linear, downward-sloping demand curves, continuous in price and quantity, with quantity expressed as "flows" rather than as discrete-quantity changes (Kyle and Lee, 2017).

   (a) "Buy a maximum of 1 share per second at $150.00 or better, declining to 0 shares per second at $151.00 or worse, until 1000 shares are bought."

3. Flow orders: investors and other market participants can directly trade arbitrary portfolios of assets.

   (a) Portfolio = any user-defined linear combination of assets. Asset weights can be arbitrary positive or negative real numbers.

Flow Trading: Combination of Three Ideas

Portfolio = any user-defined linear combination of assets.

- Lots of use cases for finance: market or sector index, customized long-only portfolio (customized ETF), pairs trade (buy A, sell B), customized long-short portfolio (factor model), ESG screens, etc.
- Includes both complements (same sign) and substitutes (opposite signs).
- Important: prices in our design are asset-by-asset "item prices", jointly determined each batch interval to clear the market. Does not require unrealistic coincidence of wants.
  - Existence of such prices not ex-ante obvious given complementarities
  - Ex: a buyer of portfolio of AAPL, GOOG, MSFT might trade with
    - seller of AAPL, AAPL+MSFT portfolio, S&P 500 index
    - trader with a long AMZN short AAPL pairs trade
    - trader with a long value short growth factor model
    - etc.
- We call the design "Flow Trading" because investors and other market participants can directly implement their desired trading strategies as flows.

Flow Trading vs. Traditional Exchange

Flow Trading	Traditional Exchange
Batch auctions once per second	Sequential matching one at a time
Downward-sloping piecewise-linear supply and demand curves for flows	Discontinuous step functions for discrete quantities
Orders for portfolios (linear combinations)	Orders for one asset

Table 1: Comparison of Flow Trading with Traditional Exchange

Benefits

Flow Trading has several sets of benefits relative to the status quo

1. Inherits benefits of BCS's and KL's prior work.
    - BCS: addresses latency arbitrage and the arms race for trading speed.
    - KL: builds in the ability to trade gradually directly into the market design. This is what large investors want practically and what theory suggests is optimal (Black 1971; Kyle 85; Vayanos 99).

2. BCS+KL = prices and quantities can be approximately continuous without issue.
    - Ex: nano-shares, micro-dollars.
    - Addresses many of the complexities and inefficiencies caused by tick-size and lot-size constraints in modern markets:
        - (i) races for queue position;
        - (ii) complex order types;
        - (iii) proliferation of exchanges
    - In status quo: continuous prices and quantities would cause an explosion of message traffic (Edelman and Ostrovsky, 2007; Baruch and Glosten, 2013). Artificial constraints instead.

Benefits

3. Market participants can directly express many common trading demands.
    - Ex: TWAP, customized ETFs, factor portfolios, ESG tilts, pairs trades.
    - Indicators of value: size of ETF industry, size of industries around fast arbitrage and short-horizon predictive analytics.

4. Market participants can directly provide liquidity across related assets.
    - Ex: buy A, sell B where A and B are highly correlated assets or portfolios.
    - Reduces economic rents, fixes “correlation breakdown.”
    - Like a string that ties prices together.

Benefits

5. Transparency and fairness.
    - All orders that are executable at the market-clearing prices are executed at rate (full or partial) determined by order parameters and public prices.
    - All participants can thus check whether they received exactly correct execution.
    - May not sound radical, but this is a significant transparency improvement over the current market design (Tyc, 2014).

Caveat:

- Flow Trading is not designed to mitigate market failures related to market power or private information (see Rostek and Yoon 2020 for a recent survey).
- Market Participants still must think strategically about how to trade on private information and how to manage their price impact, just as in the status quo.

Technical Foundations

Four sets of technical results:

1. Applied Theory Model
   (a) BCS-style model to articulate the three sources of inefficiency in the status quo that Flow Trading addresses.
   (b) Kyle (1985)-style model to provide a richer analysis of the most novel part of our market design, trading portfolios of assets.

2. Existence and Uniqueness of Equilibrium Prices
   (a) Formally define flow orders and market clearing conditions.
   (b) Market-clearing quantities exist and are unique.
   (c) Market-clearing prices exist, convex set. Can be made unique with tie-breaking rule.

Technical Foundations

3. Computational Feasibility
   (a) Walrasian tatonnement guaranteed to converge (easier than Scarf's problem).
   (b) For speed: add exchange as an epsilon "market maker of last resort." Allows us to use interior point methods which are much faster.
   (c) Simulation results: can compute large problems using public-domain solvers and ordinary workstation in <1 second.
   (d) Ex: 500 assets, 1M orders, 0.56 seconds.
   (e) Not just that we are in P versus NP, but actually fast enough for real-world use.

4. Microfoundation for Flow Orders
   (a) CARA-Normal framework, show that a trader's optimal demand can be expressed using Flow Trading Orders.
   (b) If assets have factor structure of rank K, need K Flow Trading orders.
   (c) Flow Trading orders are not fully general, this result microfounds our approach.

Plan for the Talk

1. Introduction
2. Related Literature
3. Applied Theory Model
4. Existence and Uniqueness of Equilibrium Prices
5. Computational Feasibility
6. Microfoundation for Flow Orders
7. Conclusion

Plan for the Talk

1. Introduction
2. **Related Literature**
3. Applied Theory Model
4. Existence and Uniqueness of Equilibrium Prices
5. Computational Feasibility
6. Microfoundation for Flow Orders
7. Conclusion

Literature

- Our market design combines orders for portfolios with ideas from Budish, Cramton, Shim (2015) and Kyle and Lee (2017).
- Flow orders are motivated by theoretical models (Vayanos (1999); Du and Zhu (2018); Kyle, Obizhaeva, Wang (2018)) as well as empirical evidence (popularity of TWAP and VWAP trading).
- Sophisticated expressions of preferences over multiple objects: Lahaie and Parkes (2004); Sandholm and Boutilier (2006); Milgrom (2009); Klemperer (2010); Cramton (2017); Budish et al (2017).
- Wittwer (2021) and Rostek and Yoon (2020a,b) discuss welfare implications of clearing assets jointly versus separately.
- Growing literature on financial market design: Duffie and Zhu (2017), Hortacsu, Kastl and Zhang (2018), Zhang (2020), Chen and Duffie (2020), Duffie and Dworczak (2021), Clark Houde and Kastl (2021), Budish Lee and Shim (2024), many others.
- Indivisible goods and existence of competitive eqm: Kelso and Crawford (1982), Hatfield and Milgrom (2005), Hatfield et al (2013, 2019), Baldwin and Klemperer (2019).

Plan for the Talk

1. Introduction
2. Related Literature
3. **Applied Theory Model**
4. Existence and Uniqueness of Equilibrium Prices
5. Computational Feasibility
6. Microfoundation for Flow Orders
7. Conclusion

Model of the Status Quo Market Design

- Plan: BCS-style model of a continuous limit order book market, with:
  - multiple correlated assets
  - multiple time periods
  - large and small investors
- Caveat:
  - This model will do a reasonable job of articulating the critique of the status quo and the comparison of status quo to flow trading.
  - At the cost of making some assumptions that are clearly quite stylized.
  - Kyle-1985 style model below will have a richer treatment of quantities and portfolios, with different stylized assumptions.

Model of the Status Quo Market Design

- 2 assets, A and B. Assume fundamentals are perfectly correlated.
- 2 time periods, 1 and 2.
- Investors
  - γ_s = probability that a small investor arrives to market, wants to trade exactly 1 unit
  - γ_l = probability that a large investor arrives to market, wants to trade exactly 4 units
  - could generalize the "1" and the "4" to have distributions
- Small investors: equally likely to "land" in 1A, 1B, 2A, 2B. Trade 1 unit wherever they land.
  - Write as $\gamma_s^{1A} = \gamma_s^{1B} = \gamma_s^{2A} = \gamma_s^{2B} = \frac{\gamma_s}{4}$

Model of the Status Quo Market Design

- Large investors
  - Like small, equally likely to land in 1A, 1B, 2A, 2B if they do not have technology.
  - If they have portfolio technology: then they can spread their trading across A and B.
  - If they have routing technology: then they can spread their trading across 1 and 2.
  - If they have both technologies: then they can spread their trading across 1A, 1B, 2A, 2B.

- Technologies
  - η_p = proportion of large investors with portfolio technology over assets. Costs c_p.
  - η_r = proportion of large investors with routing technology over time. Costs c_r.
  - η_{pr} = proportion of large investors with both portfolio and trade routing technology.
  - η_0 = proportion of large investors without technology ($\eta_p + \eta_r + \eta_{pr} + \eta_0 = 1$).
  - Treat these proportions as exogenous and discuss equilibrium implications.

Model of the Status Quo Market Design

- Adverse selection and sniping
  - γ_{public}: probability that there is public news that triggers a sniping race (speed costs c_s)
  - $\gamma_{private}$: probability that there is private news that enables an informed trader to profitably trade
  - J_{public}, $J_{private}$: jump distributions of public and private news, respectively, with upper bound $\bar{J}_{public}$, $\bar{J}_{private}$
  - (See Budish, Lee and Shim 2024 for modeling definitions)

- Timing assumptions
  - We assume there is at most one event per time period – the arrival of an investor, an informed trader, or a sniping race.
  - We focus on equilibrium in period 1, which we think of as the stationary equilibrium.
  - We are not going to try to capture any conditioning of period 2 behavior on actions in period 1 (in a richer model, there would always be a next period).

Equilibrium of the Status Quo

- The equilibrium bid-ask spread s_1 for the 1st unit of liquidity in a given asset-time market, say market 1A, is characterized by:

$$(\gamma_s^{1A} + \eta_0\gamma_l^{1A} + \eta_r\gamma_l^{A} + \eta_p\gamma_l^{1} + \eta_{pr})\frac{s_1^*}{2} = \gamma_{private}L_{private}(\frac{s_1^*}{2}) + \gamma_{public}L_{public}(\frac{s_1^*}{2}) \quad (1)$$

  where

$$L_{private}(x) = \int_x^{\bar{J}_{private}} (J_{private} - x)dJ_{private}$$

$$L_{public}(x) = \int_x^{\bar{J}_{public}} (J_{public} - x)dJ_{public}.$$

- LHS = revenues from investors. RHS = costs of adverse selection + sniping.
- On LHS of (1): get revenues from investors if
  - Small arrives to 1A: γ_s^{1A}
  - Large without technology arrives to 1A: $\eta_0\gamma_l^{1A}$
  - Large with routing technology arrives to A: $\eta_r\gamma_l^{A}$
  - Large with portfolio technology arrives at 1: $\eta_p\gamma_l^{1}$
  - Large with both technologies arrives: η_{pr}

Equilibrium of the Status Quo

- The equilibrium bid-ask spread s_2 for the 2nd unit of liquidity in a given asset-time market, say market 1A, is characterized by

$$(\eta_0\gamma_l^{1A} + \eta_r\gamma_l^{A} + \eta_p\gamma_l^{1})\frac{s_2^*}{2} = \gamma_{private}L_{private}(\frac{s_2^*}{2}) + \gamma_{public}L_{public}(\frac{s_2^*}{2}) \tag{2}$$

  because the only investors who trade the 2nd unit are

  - Large without technology $(\eta_0\gamma_l^{1A})$
  - Large with only routing technology who arrives to A $(\eta_r\gamma_l^{A})$
  - Large with only portfolio technology who arrives at 1 $(\eta_p\gamma_l^{1})$

- Small investors don't trade enough quantity to need the 2nd unit.
- Large investors with both technologies don't trade the 2nd unit anywhere – they just buy 1 unit in each asset-time market.

Equilibrium of the Status Quo

- The equilibrium bid-ask spread s_4 for the 4th unit of liquidity in a given asset-time market, say market 1A, is characterized by

$$(\eta_0 \gamma_l^{1A}) \frac{s_4^*}{2} = \gamma_{private} L_{private}(\frac{s_4^*}{2}) + \gamma_{public} L_{public}(\frac{s_4^*}{2}) \tag{3}$$

  because the only investor who trades the 4th unit is a large investor without technology who arrives specifically to 1A.

- Note: $s_3^* = s_4^*$ in this model, anybody who trades the 4th unit also trades the 3rd unit. Could generalize but doesn't add much.

Insights from the Status Quo Equilibrium

1. Liquidity and market depth – increasing cost of liquidity for large orders. The spread is smaller for the 1st unit than the 2nd unit than the 3rd and 4th units.
   - That is: $s_1 < s_2 < s_3 = s_4$, this is from comparison of (1), (2), (3).
   - Intuition: LHS is better for liquidity providers the more investors who arrive and want to trade at least that quantity.

2. Incentive to adopt trading technology. This market depth result in turn gives large investors the incentive to buy the trading technology.
   - If I don't have trading technology, I pay $s_1 + s_2 + s_3 + s_4$ wherever I show up.
   - If I only have one technology, I pay $2(s_1 + s_2)$.
   - If I have both technologies, I pay $4s_1$.

Insights from the Status Quo Equilibrium

3. Strategic complementarity in investing in trading technology:
   - There is a strategic complementarity in adopting trading technology, because s_3, s_4 goes to infinity as $\eta_0 \to 0$.
   - More generally, as η_0 decreases, s_3, s_4 increases. So as others adopt technology, I need to also.
   - We could call this a trading technology trap.

4. Trading technology spillovers to small investors.
   - Note also: small investors are helped by large investors spreading their trading out over time and assets, i.e. the small investors are helped by the trading technology.
   - Because s_1 is decreasing in $\eta_r \gamma_l^A + \eta_p \gamma_l^1 + \eta_{pr}$, the proportion of large investors who would not have traded in this asset (here, 1A) if they didn't have the technology, who now do trade in this asset because they do have the technology.

Insights from the Status Quo Equilibrium

Model is clearly quite stylized, but it does a reasonable job of capturing several important features of the status quo:

- Captures the increasing cost of trading large quantities in a simple way. (Could generalize from "1" and "4").
- Captures the strategic complementarity of trading technology in a simple way (idea of strategic complementarity in spreading out trading is in Vayanos 1999).
- Captures that large investors are trapped – have to invest, especially as others do.
- Captures that small investors are actually helped by large investors having the trading technology.
- Captures that most trades are "1 unit at a time" across time and space.
- Captures adverse selection, sniping, trading technology, all in a simple-ish framework.

Modeling Flow Trading

- What do the three ingredients of Flow Trading do?
  - Batching: eliminates the $\gamma_{public} L_{public}(\frac{s}{2})$ sniping expense on the RHS of (1), (2) and (3). That is, all quantities of trade are cheaper.
  - Flow: spreads the investor demand smoothly across time periods – all large investors get the "routing" technology built in to the design, for free.
    - As if $\eta_r = 1$, $c_r = 0$.
  - Portfolio: spreads the investor demand smoothly across assets – all large investors get the "portfolio" technology built in to the design, for free.
    - As if $\eta_p = 1 c_p = 0$.

- So the equilibrium condition for the bid-ask spread s_1 for the first unit of liquidity under Flow Trading, for a given asset-time market (say 1A), is characterized by:

$$(\gamma_s^{1A} + \gamma_l)\frac{s_1^*}{2} = \gamma_{private} L_{private}(\frac{s_1^*}{2}). \tag{4}$$

  - revenue on the LHS is higher because all large investors trade smoothly everywhere. γ_l is <u>all</u> large investors.
  - costs on the RHS are lower because batching eliminates the $\gamma_{public} L_{public}(\cdot)$ sniping expense term.

Welfare Effects of Flow Trading

In the model, flow trading improves liquidity for investors and eliminates deadweight costs from technology.

1. Small investors gain from better liquidity.
   (a) This is from comparison of (1) and (4).

2. Large investors gain from better liquidity.
   (a) Large without technology – get much better liquidity.
      i. This is from comparison of (2) and (3) to (4).
   (b) Large with technology – also get better liquidity, but smaller gain than the large investors without technology.
      i. This is from comparison of (1) to (4).

Welfare Effects of Flow Trading

3. Technology cost savings.
   (a) Portfolio and routing technology "built in", saves $(\eta_{pr} + \eta_p)c_p$, $(\eta_{pr} + \eta_r)c_r$.
   (b) Speed technology not needed, saves c_s.
   (c) These cost savings manifest in better liquidity in equilibrium.

- Economic sources of these inefficiencies: (i.e., "what is the market failure?")
  - inefficient trading by large investors absent expensive technology. technology trap in the η's .
  - prisoner's dilemma for speed investment for the HFTs.

- Informed traders have effects in different directions.
  - They save on trading costs.
  - They pay lower bid-ask spreads for the quantities they trade.
  - There is less quantity for them to trade against (Baldauf and Mollner, 2020).

- HFTs can no longer snipe stale quotes, but they no longer pay c_{speed}.

Kyle (1985) Style Model

For a richer analysis of the most novel part of our market design, trading portfolios, let's use a Kyle (1985) style model.

- As above, two assets A and B, perfectly correlated fundamentals
- Liquidation value $v \sim N(0, \sigma_v^2)$
- One-shot Kyle (1985) model:
  - Informed trader: knows v
  - Noise traders, who trade one of the two assets exogenously
    - $u_A \sim N(0, \sigma_u^2)$
    - $u_B \sim N(0, \sigma_u^2)$
    - v, u_A, u_B all independently distributed
  - As in Kyle (1985), the informed trader and noise traders submit *market orders*, choosing only the quantity to trade at the prevailing market-clearing price.
  - $M \geq 3$ risk neutral market makers who strategically submit demand schedules. In the limit as $M \to \infty$ aggregate profits of market makers are zero.

Equilibrium of the Separate-Asset Kyle 85 Market

Equilibrium for each asset, if the assets trade separately:

Proposition. In equilibrium, the informed trader's traded quantity (x) and each of the market makers' demand schedules $y(p)$ are given by:

$$x = \left(\frac{\sigma_u}{\sigma_v}\sqrt{\frac{M-2}{M}}\right) v \qquad \text{and} \qquad y(p) = -\left(\frac{2\sigma_u}{M\sigma_v}\sqrt{\frac{M-2}{M}}\right) p, \tag{5}$$

respectively. And the equilibrium price is given by:

$$p = \frac{v}{2} + \left(\frac{\sigma_v}{2\sigma_u}\sqrt{\frac{M}{M-2}}\right) u. \tag{6}$$

Equilibrium with Flow Trading

To model Flow Trading

- We allow the informed trader and the market makers to trade two portfolios: $A + B$ and $A - B$.
- Let u_+ and u_- represent the aggregate noise traders' demand for portfolios $A + B$ and $A - B$.
- For the $A + B$ portfolio, we can apply the single-asset equilibrium after substituting v with $2v$, σ_v with $2\sigma_v$, and σ_u with $\sigma_u/\sqrt{2}$:

$$x_+ = \left(\frac{\sigma_u}{\sqrt{2}\sigma_v}\sqrt{\frac{M-2}{M}}\right) v \qquad \text{and} \qquad y_+(p_+) = -\left(\frac{\sigma_u}{M\sqrt{2}\sigma_v}\sqrt{\frac{M-2}{M}}\right) p_+, \tag{7}$$

  The equilibrium price is given by:

$$p_+ = v + \left(\frac{\sigma_v}{\sqrt{2}\sigma_u}\sqrt{\frac{M}{M-2}}\right)(u_A + u_B). \tag{8}$$

- For the $A - B$ portfolio, in equilibrium market makers provide infinite depth to noise traders.
  - The informed trader chooses not to participate.

Ex-ante utility	Informed trader	Noise traders	Market makers
Baseline	$\sigma_v\sigma_u\sqrt{\frac{M-2}{M}} - c$	$-\sigma_v\sigma_u\sqrt{\frac{M}{M-2}}$	$\frac{2\sigma_v\sigma_u}{\sqrt{M(M-2)}} - c$
Flow trading	$\frac{\sigma_v\sigma_u}{\sqrt{2}}\sqrt{\frac{M-2}{M}}$	$-\frac{\sigma_v\sigma_u}{\sqrt{2}}\sqrt{\frac{M}{M-2}}$	$\frac{\sqrt{2}\sigma_v\sigma_u}{\sqrt{M(M-2)}}$

Table 2: Welfare comparisons

- Noise traders lose less.
  - Intuition: market makers efficiently transform A demand into B demand or vice versa.
- No correlation breakdown.
  - The two assets always have the same price in equilibrium despite noise traders.
  - $A - B$ order like a string that ties correlated prices together.
- Informed traders, market makers have offsetting effects.
  - Reduction in aggregate noise trading reduces their profits. $\frac{\sqrt{2}}{2}$ factor based on variance of noise trader demand for the $A + B$ portfolio.
  - Save technology costs c of the status quo: speed, routing technology, etc. (These could be different for Informed versus MM's, write as c_I and c_{MM}). This is a social welfare gain.

Plan for the Talk

1. Introduction
2. Related Literature
3. Applied Theory Model
4. **Existence and Uniqueness of Equilibrium Prices**
5. Computational Feasibility
6. Microfoundation for Flow Orders
7. Conclusion

How Flow Orders Work

There are N assets indexed n, I orders indexed i.
A "flow order" is specified by a tuple $(\boldsymbol{w}_i, p_i^L, p_i^H, q_i, Q_i^{max})$.

- Description of portfolio: vector of portfolio weights $\boldsymbol{w}_i \in \mathbb{R}^N$:
  - Individual asset: One nonzero weight to buy (+) or sell (-) one asset
  - Substitutes: One positive weight, one negative weight for pairs trade
  - Complements: 500 positive index weights to buy the S&P 500
  - Market making: pair of orders with weights $\boldsymbol{w}_i$ and $-\boldsymbol{w}_i$
- Two limit prices for the portfolio ($p_i^L = \$50.30$ and $p_i^H = \$50.40$ per share)
  - Negative portfolio weight (-1 share) and negative portfolio limit prices $p_i^L = -\$50.40$ and $p_i^H = -\$50.30$ for sell order
  - $p_i^L < p_i^H$ whether buying or selling (with both in $\mathbb{R}$)
- Maximum execution rate ($q_i = 1.00$ portfolio unit per second)
- Cumulative quantity to be executed ($Q_i^{max} = 10,000$ portfolio units)

How Flow Orders Work

Order executes

- At maximum rate (fully executable) q_i if price weakly below p_i^L
- At zero rate (nonexecutable) if price weakly above p_i^H
- At linearly interpolated rate (partially executable) if price in $[p_i^L, p_i^H]$.

Buying vs. Selling

- "Selling" an asset is buying a portfolio with negative weight on that asset at a negative price.
- Whether buying or selling, always $p_i^L < p_i^H$.
- Think of p_i^L as analogous to the limit price in a limit order, whether positive or negative: "trade my full quantity at any price weakly better than p_i^L".
- Think of p_i^H as expressing the price at which indifferent between trading and not: "do not trade if the price is p_i^H or worse."
- For portfolios with positive and negative weights, there may not be a natural buying versus selling direction to the order.

Math: One Flow Order

Flow order i is described by the tuple $(\boldsymbol{w}_i, p_i^L, p_i^H, q_i, Q_i^{max})$.

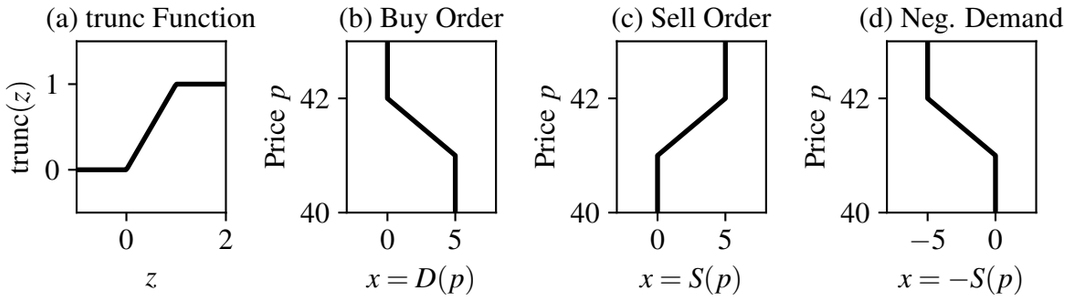
Let $\boldsymbol{\pi} = (\pi_1, \ldots \pi_N)$ denote a vector of N market-clearing asset prices. The price of the portfolio is the weighted sum of asset prices:

$$p_i = \boldsymbol{\pi}^\top \boldsymbol{w}_i. \tag{9}$$

Assume the order's cumulative purchased quantity is not within q_i of Q_i^{max}. The execution rate x_i of order i is given by:

$$x_i = D_i(p_i) = q_i \cdot \text{trunc}\left(\frac{p_i^H - p_i}{p_i^H - p_i^L}\right), \qquad \text{where} \qquad \text{trunc}(x) := \begin{cases} 1, & \text{for } x \geq 1 \\ x, & \text{for } 0 < x < 1 \\ 0, & \text{for } x \leq 0 \end{cases}. \tag{10}$$

Illustration of Buying and Selling



For an order to sell 5 shares of a single asset n between $\$41$ and $\$42$:

- $\boldsymbol{w}_i$ is sparse vector with $w_{in} = -1$ share, zero otherwise
- $p_i^L = -42.00 < p_i^H = -41.00$ dollars per share
- $q_i = 5$ shares

Flexible Preference Expression for Finance

Some key financial use cases:

- Standard limit order (set $q_i = Q_i^{max}$ and $p_i^H \to p_i^{L^+}$)
- Time-weighted average price (TWAP) order (p_i^L sufficiently aggressive). Our analog of a market order
- Pairs trades ($(+, -, 0, \ldots, 0)$)
- Portfolio trades ($(+, +, \ldots)$ or $(-, -, \ldots)$)
- General long-short strategies
- Market making strategies (two orders with weights $\boldsymbol{w}_i$, $-\boldsymbol{w}_i$)

Flexibility of the language (shoe analogy):

- Buy or sell a left shoe or right shoe separately
- Substitutes: Swap a left shoe for a right shoe (or vice versa)
- Complements: Buy left shoe and right shoe together
- Urgency expressed by maximum execution rate
- Arbitrary continuous downward-sloping portfolio demand function can be approximated with piecewise-linear orders

Limitations of the Preference Expression Language

Limitations

- Demand schedules must be downward sloping.
- Trading demands are only defined at exactly the ratio of portfolio weights specified in the order. (Contrast to consumer theory).
- Trading demands are linear within each order.
- Language does not allow for indivisibilities (Ex: at least 100 shares, or none).
- Demand depends only on portfolio prices. Can't condition demand for asset A on realized price of asset B: "buy whichever of Left Shoe or Right Shoe gives me more surplus given my reported valuations". (Contrast: Klemperer (2010) or Milgrom (2009)).
- Relatedly, no in-order contingencies ("buy A if the price of B is high enough").

Informally: "sweet spot" where expressive enough for finance, some limitations, but with guaranteed existence and fast computation.

Math: Market Clearing

Market clears in N assets. (Not portfolios, specified by I orders, with $N \ll I$)

At price vector $\boldsymbol{\pi} \in \mathbb{R}^N$, the exchange converts each order i's demand for portfolio units to demand for underlying assets by multiplying by its portfolio weights $\boldsymbol{w}_i$. Summing over the I orders yields the excess demand vector:

$$\text{Excess Demand Vector} = D(\boldsymbol{\pi}) := \sum_{i=1}^{I} D_i \left(\boldsymbol{w}_i{}^{\mathsf{T}} \boldsymbol{\pi} \mid \boldsymbol{w}_i, q_i, p_i^L, p_i^H \right) \cdot \boldsymbol{w}_i. \tag{11}$$

The exchange seeks to find a market clearing price vector

$$D(\boldsymbol{\pi}) = \mathbf{0}, \qquad (N \text{ equations in } N \text{ unknowns}) \tag{12}$$

in which case each order executes at rate

$$x_i = D_i(\boldsymbol{w}_i{}^{\mathsf{T}} \boldsymbol{\pi}). \qquad (\text{scalar equation in portfolio units } x_i) \tag{13}$$

Existence and Uniqueness

Questions:

- Do equilibrium prices and quantities exist?
- If they exist, are they unique?

Idea for proof:

- Treat orders as expressions of preferences, implying quasi-linear (dollar) quadratic utility for order i for quantities in range $[0, q_i]$.
  - Note: order i's preferences are only defined exactly on the $\boldsymbol{w}_i$ vector.
  - Note: order i's preferences are satiated at q_i.
- Exchange solves the problem of maximizing sum of dollar utility across orders subject to market clearing constraint.
- Interpret Lagrange multipliers for market clearing constraint as price vector $\boldsymbol{\pi}$.

Setting up the Optimization Problem

- Let $V_i(x_i)$ denote the dollar utility of order i from a trade rate of $x_i = D_i(p_i)$ portfolio units per batch.
- To find $V_i(x_i)$ first define the marginal utility function $M_i(x_i)$ as the inverse demand curve $p_i = M_i(x_i)$.
  - In words, this function maps from a trade rate $x_i \in [0, q_i]$ into prices $p_i \in [p_i^L, p_i^H]$.

- Rearranging from earlier we have

$$M_i(x_i) = p_i^H - \frac{p_i^H - p_i^L}{q_i} x_i, \; x_i \in [0, q_i].$$

- Utility is the integral of marginal utility:

$$V_i(x_i) = \int_0^{x_i} M_i(u) du.$$

- Since marginal value is linear in x_i, total value is quadratic and thus strictly concave in x_i :

$$V_i(x) = p_i^H x - \frac{p_i^H - p_i^L}{2q_i} x^2. \tag{14}$$

Setting up the Optimization Problem

Exchange solves the problem of finding quantities $\boldsymbol{x} = (x_1, \ldots, x_I)$ to solve

$$\max_{\boldsymbol{x}} \sum_{i=1}^{I} V_i(x_i) \qquad \text{subject to} \begin{cases} \sum_{i=0}^{I} x_i \boldsymbol{w}_i = \boldsymbol{0} & \text{(market clearing)} \\ 0 \leq x_i \leq q_i \text{ for all } i & \text{(order execution rate).} \end{cases} \tag{15}$$

This is a quadratic optimization problem with:

- N linear equality constraints enforcing market clearing of N assets
- $2I$ linear inequality constraints enforcing no overfilling or underfilling of I orders

In matrix form: Let $\boldsymbol{D}$ denote the $I \times I$ positive definite diagonal matrix whose ith diagonal element is $(p_i^H - p_i^L)/q_i$.

$$\max_{\boldsymbol{x}} \left[\boldsymbol{x}^\mathsf{T}\boldsymbol{p}^H - \tfrac{1}{2}\boldsymbol{x}^\mathsf{T}\boldsymbol{D}\,\boldsymbol{x}\right] \qquad \text{subject to} \qquad \boldsymbol{W}\,\boldsymbol{x} = 0 \qquad \text{and} \qquad 0 \leq \boldsymbol{x} \leq \boldsymbol{q}. \tag{16}$$

Theorem: Existence and Uniqueness of Quantities

Theorem 1 (Existence of Market Clearing Prices)**.** *There exists a unique quantity vector* $\boldsymbol{x}^*$ *which solves the maximization problem* (15)*.*

Proof.

- Compactness: Inequality constraints on quantities. Market clearing conditions are linear constraints, which defines the intersection of hyperplanes. Thus the set of vectors of trade rates $\boldsymbol{x}$ that satisfies all constraints is compact and convex.
- Feasibility: No trade $\boldsymbol{x} = 0$ is feasible: satisfies constraints with finite value of objective.
- Strictly Concave Objective Function: Each function V_i is quadratic and therefore strictly concave. Since V is the sum of V_i across i, the function V is concave as well (on $\mathbb{R}^I$ and on the compact and convex subset defined by the constraints).
- Well-known principle of convex analysis that a strictly concave objective function on a non-empty compact and convex set has a unique maximizing vector $\boldsymbol{x}^*$ (Reference: Bertsekas (2009), Propositions 3.1.1, 3.2.1).

□

Why is Existence so Simple?

The language allows users to define arbitrary portfolios, including complements and substitutes.

In GE and Indivisible Goods literatures, complements especially make existence hard. (Arrow-Debreu-McKenzie, Starr 1969, Hatfield-Kominers-Westkamp 2021, Baldwin-Klemperer 2019)

Yet here the existence proof is simple. Why?

- Goods are infinitely divisible.
- Portfolio demand schedules are downward sloping.
- Utility for each order is defined only on the line segment associated with the portfolio weights (not defined off diagonal).
- Quantities are bounded. Zero trade is feasible.
- No in-order contingencies or linkages across multiple orders. (This limits the kinds of complementarities and substitutabilities)

A sweet spot? Expressive enough to be useful, and existence is guaranteed.

Existence of Market-Clearing Prices

To prove that market-clearing prices exist, we exploit the duality between the problems of finding optimal prices and quantities.

- Start with the Lagrangian function of the vector of trade rates $\boldsymbol{x}$ with three constraints: market clearing; trade rates for each portfolio are greater than 0; trade rates for each portfolio are less than their defined maximum. In vector notation:

$$L(\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := \boldsymbol{x}^{\mathsf{T}} \boldsymbol{p}^{\boldsymbol{H}} - \frac{1}{2} \boldsymbol{x}^{\mathsf{T}} \boldsymbol{D} \boldsymbol{x} - \boldsymbol{\pi}^{\mathsf{T}} \boldsymbol{W} \boldsymbol{x} + \boldsymbol{\mu}^{\mathsf{T}} \boldsymbol{x} + \boldsymbol{\lambda}^{\mathsf{T}} (\boldsymbol{q} - \boldsymbol{x}). \tag{17}$$

- Prices $\boldsymbol{\pi}$ are positive or negative Lagrange multipliers enforcing market clearing.
- $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are Lagrange multipliers enforcing order execution rate constraints.

The dual problem associated with the primal problem of maximizing aggregate utility is:

$$\hat{G}(\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) := \max_{\boldsymbol{x}} \; L(\boldsymbol{x}, \boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \qquad \text{for} \qquad \boldsymbol{\pi} \in \mathbb{R}^N, \qquad \boldsymbol{\mu} \geq 0, \qquad \boldsymbol{\lambda} \geq 0. \tag{18}$$

$$g := \inf_{\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}} \hat{G}(\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \qquad \text{subject to} \qquad \boldsymbol{\pi} \in \mathbb{R}^N, \qquad \boldsymbol{\mu} \geq 0, \qquad \boldsymbol{\lambda} \geq 0 \tag{19}$$

Result: Existence of Market Clearing Prices

Theorem 2 (Existence of Market Clearing Prices)**.** *There exists at least one optimal solution* $(\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *to the dual problem* (19)*. The solutions* $\boldsymbol{x}$ *and* $(\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu})$ *are a primal-dual pair which satisfies the strict duality relationship*

$$g = V(\boldsymbol{x}). \tag{20}$$

Proof. The quadratic program has these properties:

- Concavity: The objective function $V(\boldsymbol{x})$ is concave.
- Finite solution: The primal objective is the sum of a finite number of concave quadratic functions, each of which is bounded above.
- Linear constraints: The market clearing constraints $\boldsymbol{W}\boldsymbol{x} = 0$ and the the order execution rate constraints $\boldsymbol{x} \geq 0$ and $\boldsymbol{x} \leq q$ are all linear.
- Feasibility: No trade ($\boldsymbol{x} = 0$) is feasible: clears markets, satisfies order execution rates.

Standard result from convex programming that these conditions guarantee that a solution to the dual problem exists and has the same value as the supremum to the primal (Bertsekas 2015, Proposition 5.3.4). □

The set of market clearing prices is convex, but may be unbounded (simple example: a single order).

Plan for the Talk

1. Introduction
2. Related Literature
3. Applied Theory Model
4. Existence and Uniqueness of Equilibrium Prices
5. **Computational Feasibility**
6. Microfoundation for Flow Orders
7. Conclusion

Computation

Question: prices and quantities exist. Can we compute them?

- Many economic settings where prices are known to exist but hard to find (Scarf and Hansen, 1973)
- Many economic settings where prices are trivial to compute—one asset version of our problem is an easy example!
- Our problem lies in between

- Plan [most details in paper]
  - Show that gradient method works. “Easier than Scarf’s problem.”
  - Result: gradient method convergence slow (confirmed in simulations).
  - Add “exchange as market maker” which enables interior point methods.
  - Result: faster in theory, and also in simulations.
  - Goal in mind: solve large problems in less than one second.

Computation: Tatonnement and Gradient Method

Walrasian tatonnement is an intuitive approach for calculating market-clearing prices

- Auctioneer announces tentative prices
- Traders respond with quantities
- Prices adjust in the direction of excess demand
- Process continues until the market clears

In our setting, Tatonnement can be implemented by applying gradient optimization to an objective function of prices whose first-order conditions correspond to market clearing. Called a Gains function:

$$G(\boldsymbol{\pi}) := \inf_{\boldsymbol{\lambda},\boldsymbol{\mu}} \hat{G}(\boldsymbol{\pi}, \boldsymbol{\lambda}, \boldsymbol{\mu}) \qquad \text{subject to} \qquad \boldsymbol{\mu} \geq \mathbf{0}, \qquad \boldsymbol{\lambda} \geq \mathbf{0}. \tag{21}$$

- Theorem 2 implies that this function's first order conditions, $G'(\boldsymbol{\pi}) = \mathbf{0}$, correspond to market clearing.
- Bad news: Although convergence can be guaranteed, convergence is too slow for our purposes (see Nesterov 2004, Corollary 2.1.2).
  - Reducing error by proportion ϵ takes $O(\frac{1}{\epsilon})$ iterations. E.g. $O(10^6)$ iterations for 0.000001 improvement.

Interior Point Method

- Literature shows that interior point methods are computationally superior to the more intuitive gradient point method.
- Theoretically, interior point method requires a feasible allocation on the interior of the constraint set. That is, clears the market and strictly satisfies inequality constraints, $0 < \boldsymbol{x} < \boldsymbol{q}$.
- No natural candidate for us. We know no trade is feasible ($\boldsymbol{x} = 0$) but not interior.
- Idea: exchange as a "small market maker". Linear demand curve for each asset n

$$\epsilon_n(\pi_{0n} - \pi_n), \tag{22}$$

  where $\epsilon_n > 0$ is small, π_{0n} is exchange's best guess at clearing price (e.g., last price or from some initial computation).
- Can be implemented with two flow orders for each asset, one to buy at prices below π_{0n}, one to sell at prices above π_{0n}.
- Now we have an interior point. Exchange can soak up epsilons of demand.
- Reducing error by proportion ϵ takes $O(log\frac{1}{\epsilon})$ iterations. E.g. $O(13.8)$ iterations for 0.000001 improvement.

How Interior Point Method Works

Adding exchange as a trader, modified primal problem is

$$\max_{\boldsymbol{x},\boldsymbol{y}} \left[\boldsymbol{x}^{\mathsf{T}}\boldsymbol{p}^{H} - \tfrac{1}{2}\,\boldsymbol{x}^{\mathsf{T}}\boldsymbol{D}\,\boldsymbol{x} + \boldsymbol{y}^{\mathsf{T}}\boldsymbol{\pi}_0 - \tfrac{1}{2}\,\boldsymbol{y}^{\mathsf{T}}\boldsymbol{\epsilon}^{-1}\,\boldsymbol{y}\right] \quad \text{subject to} \quad \boldsymbol{W}\boldsymbol{x} + \boldsymbol{y} = 0, \quad 0 \leq \boldsymbol{x} \leq \boldsymbol{q}. \tag{23}$$

Interior point method changes problem by replacing inequality constraints with penalty functions:

$$\max_{\boldsymbol{x},\boldsymbol{y}} \left[\boldsymbol{x}^{\mathsf{T}}\boldsymbol{p}^{H} - \tfrac{1}{2}\boldsymbol{x}^{\mathsf{T}}\boldsymbol{D}\,\boldsymbol{x} + \boldsymbol{y}^{\mathsf{T}}\boldsymbol{\pi}_0 - \tfrac{1}{2}\boldsymbol{y}^{\mathsf{T}}\boldsymbol{\epsilon}^{-1}\,\boldsymbol{y} + \bar{\nu}\log(\boldsymbol{x})^{\mathsf{T}}1 + \bar{\nu}\log(\boldsymbol{q}-\boldsymbol{x})^{\mathsf{T}}1\right], \tag{24}$$

Solution of modified problem with inequality constraints converges to solution of modified problem without inequality constraint in the limit $\bar{\nu} \to 0$.

Simulation Assumptions

Try to pose a difficult problem:

- 500 assets and 100,000 orders in base case scenario.
  - 50% for individual assets, 25% for various index portfolios, 25% for pairs trades.
  - Also 1000 exchange orders, one buy and one sell for each asset, with tiny slope (1 dollar per 1% change in price).
- Huge variation in liquidity across assets (Invariance assumptions).
- Indices: market index, "size" indices, "industry" indices, each value-weighted and equal-weighted. 32 in base case, up to 202 in high case.
- Limit prices are distributed around an arbitrary initial price, normalized to \$100 per share or index unit. Lognormal with mean \$97 for buys, mean \$103 for sells, st. dev of 10%.
- Very small (one basis point) expected difference between p_i^H and p_i^L on average, with large variance.
- Overall: trying to capture that liquidity and trading volume vary enormously across assets and indices.

Simulation Results

- Panel A varies orders: $t \approx 1$ sec for 500 assets and 2,000,000 orders.
- Panel B varies assets: $t \approx 1$ sec for 2000 assets and 100,000 orders.

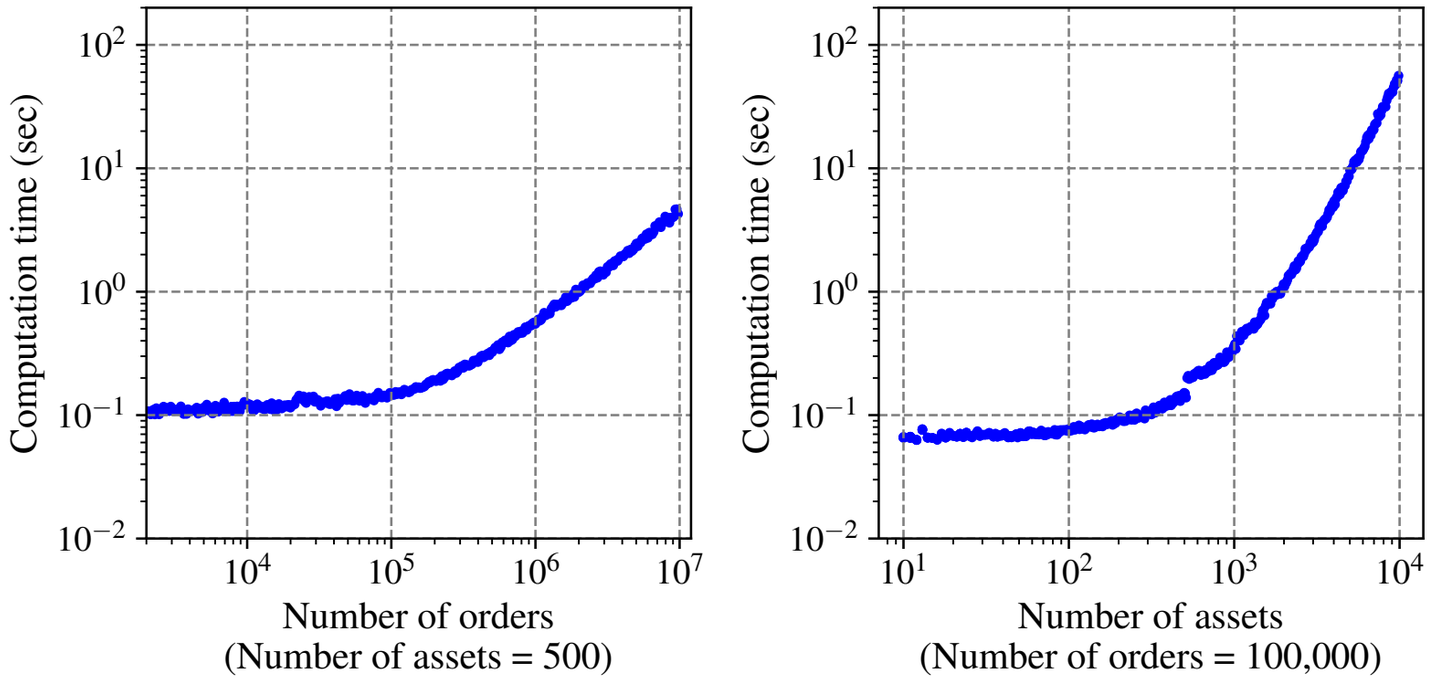


Figure 1: Execution time for 2×500 simulated market outcomes.

Discussion

- Overall, market-clearing prices and quantities are computed quickly and accurately with minimal trading by exchange.
  - (Exchange averages $3.2 per million dollars, or 0.032 basis points of trading volume)
- Computation times grow with number of orders and assets, as expected.
- Reassuringly, growth looks linear.
- Several thousands of assets, large numbers of orders, in <1 to 10 seconds, on an ordinary workstation with open-source software.
- Interpret as “proof of concept.”
- If adopted for use in practice, we would expect more powerful computers and more refined algorithms.

Plan for the Talk

1. Introduction
2. Related Literature
3. Applied Theory Model
4. Existence and Uniqueness of Equilibrium Prices
5. Computational Feasibility
6. **Microfoundation for Flow Orders**
7. Conclusion

Microfoundation for Flow Orders

- Our flow orders are more general than traditional limit orders but the language still imposes restrictions (as discussed above).
- In particular, demand for a portfolio depends on
  - In our language: the price of the portfolio (scalar)
  - In general: the complete vector of asset prices
- We provide a modest microfoundation for our approach using the workhorse CARA-Normal model.

Microfoundation for Flow Orders

- Assume trader has CARA (exponential) utility with risk aversion A.
- Assume trader has subjective beliefs that liquidation values $\boldsymbol{v}$ are jointly normally distributed with subjective mean $\boldsymbol{m}$ and subjective variance Σ.
- At a given set of prices $\boldsymbol{\pi}$ the trader's portfolio optimization problem, given her beliefs, risk preferences, and prices:

$$\max_{\boldsymbol{\omega}} E\left[-\exp\left(-A(\boldsymbol{v}-\boldsymbol{\pi})^{\mathsf{T}}\boldsymbol{\omega}\right)\right] \tag{25}$$

Joint normality allows us to transform this to to

$$\max_{\boldsymbol{\omega}} \left[(\boldsymbol{m}-\boldsymbol{\pi})^{\mathsf{T}}\boldsymbol{\omega} - \frac{1}{2}A\boldsymbol{\omega}^{\mathsf{T}}\Sigma\boldsymbol{\omega}\right]. \tag{26}$$

The first order condition implies that the optimal portfolio demand is

$$\boldsymbol{\omega} = (A\Sigma)^{-1}(\boldsymbol{m}-\boldsymbol{\pi}). \tag{27}$$

- Observe: Quantity of asset n, given by ω_n is a seemingly arbitrary linear function of entire price vector $\boldsymbol{\pi}$, not just price of asset n itself.

Implementing the Optimum with Flow Orders

If the prices $\boldsymbol{\pi}$ are known and fixed, the trader can implement their optimum with a single flow order.

- Set weights $\boldsymbol{w}_i$ such that $\boldsymbol{w}_i = \boldsymbol{\omega}$
- Set maximum rate $q_i = 1$
- Set $p_i^H > p_i^L > \boldsymbol{\pi}^\mathsf{T} \boldsymbol{w}_i$ so that the order is fully executable at the portfolio price

What if the trader does not know the asset prices? This might capture that prices are changing over time and traders trade gradually.

Theorem 4. Consider a static CARA-normal framework in which a trader believes that the variance-covariance matrix of the asset payoffs has rank K. Then the trader's optimal portfolio demand (equation (27)) can be implemented as the sum of K downward-sloping demand schedules for portfolios, each of which depends only on that portfolio's price.

- In practice, traders may select a few portfolios which they perceive to have sufficiently attractive Sharpe ratios.

Adding Price Impact

- The same logic applies to a model with strategic trading as captured by a linear price impact matrix.

Theorem 5. Consider a static CARA-normal framework in which a trader believes that her price impact, denoted by Λ, is linear and positive semidefinite. Then the strategic trader's optimal portfolio can be represented as the sum of K' downward-sloping demand curves for portfolios, where K' denotes the rank of $A\Sigma + \Lambda$.

Plan for the Talk

1. Introduction
2. Related Literature
3. Applied Theory Model
4. Existence and Uniqueness of Equilibrium Prices
5. Computational Feasibility
6. Microfoundation for Flow Orders
7. **Conclusion**

Conclusion I: A New Market Design for Finance

- This paper has introduced a new market design for trading financial assets, such as stocks, bonds, futures, options, currencies, etc.
- Financial markets have used continuous limit order books since the 18th century ... we think there is room for improvement given the possibilities afforded by modern computers.
- Three elements:
  - frequent batch auctions from Budish, Cramton and Shim (2015)
  - continuous-scaled limit orders from Kyle and Lee (2017)
  - novel language for trading portfolios = arbitrary linear combinations of assets
- Enables investors and market participants to directly trade their desired portfolios and arbitrage trades as flows — implement their trading needs directly.
- Technical foundations: applied theory model; existence and uniqueness results; computational results; microfoundations for flow orders.

Conclusion II: Review of Benefits

- Benefits of BCS, KL in isolation
- Benefits of combining BCS+KL: a market design in which time is discrete and prices and quantities are continuous.
  - Status quo market design has these reversed.
  - Continuous time and discrete prices/quantities are the cause of significant complexity, inefficiency, and rent seeking.
  - Policy debates that relate: arms race for trading speed; complex order types; proprietary market data and access; internalization of retail investors' order flow.
- Benefits of novel language for flow orders:
  - Rich enough to directly express many important kinds of trading demands (custom ETFs, factor models, long-short arbitrage, etc.).
  - Fixes correlation breakdown, improves liquidity across correlated assets.
  - Allows for guaranteed existence of prices and fast computation.
  - "Point on the frontier of language design"—tradeoff btwn expressiveness and computability.

Conclusion III: Directions for Future Research

- On flow trading specifically
  - Study the potential efficiency benefits in greater detail. Our applied theory model just scratches the surface.
  - Study price impact and optimal use of flow orders in a fully dynamic model. (Hard).
- On financial market design more generally ... there is so much important work to do!
- Hard to understate the importance and size of financial markets. Literally >$1 quadrillion volume per year.
- So if you can improve efficiency by even 1 basis point, that's >$100bn annually, >$1 trillion NPV.
- Lots of institutional details ... barrier to entry ... pay the fixed costs!