Smooth Trading with Overconfidence and Market Power

Albert S. Kyle, Anna A. Obizhaeva, and Yajun Wang

University of Maryland

Quantitative Finance Retrospective Workshop
Fields Institute, Toronto
October 28, 2013

Main Idea

We present a dynamic model of informed trading which derives endogenously the speed with which traders trade and study the properties of markets when all traders smooth out their trading.

We model a trading game of oligopolistic traders with overconfidence and market power, where speed of trading is governed by trade-off:

- Reducing market impact motivates traders to trade slowly.
- Decaying private information motivates them to trade quickly.

Practical Implications

Order shredding significantly changes equilibrium properties.

- Model explains why dumping large quantities on the market too quickly has large temporary price impact. This is relevant for understanding flash crashes.

- Model explains why even though prices immediately reveal average signal of traders, traders continue to trade in equilibrium.

- Model explains why even though prices are fully revealing, an economist will be finding anomalies.

Key Feature

We develop an equilibrium model where speed matters. For each trader, **price** is linear function of **traders inventory** (permanent price impact) and **derivative of inventory** (temporary impact).

Trading Costs and Speed: Empirical Evidence

A conventional wisdom holds that speed of trading affects transaction costs. There is a link between trading pressure and asset prices.

- Holthausen, et al. (1990) measure temporary and permanent price effects associated with block trades and find most of the adjustment occurring in the very first trade.
- Chan and Lakonishok (1995) find high demand for immediacy tends to be associated with larger market impact.
- Keim and Madahvan (1997) find that more aggressive trades of index funds and technical traders have larger costs than trades of more patient value investors.
- Dufour and Engle (2000) find that the price impact of trades increases when duration between transaction decreases.
- Almgren et al. (2005) calibrate a specific model of transaction costs as function of shares traded and speed of trading.

Speed of Trading and Theoretical Models

In contrast, the speed of trading usually plays a limited role in most theoretical models.

Price impact costs of changing inventory levels continuously by a given amount usually do not depend on the derivative of the trader's inventory levels; there is no temporary price impact.

Speed of Trading and Kyle (1985)

- The **informed trader** privately observes the liquidation value of a risky asset and spreads large trades out over time taking his market impact into account.
- **Noise traders** dump quantities into the market.
- In continuous time version, order flow consists of the "smooth" order flow ($\sim dt$) from informed trader and the "diffusion" order flow ($\sim dB_t$) from noise traders.
- Competitive **market makers** provide liquidity for orders of all sizes X by offering linear supply curve $P(X) = P_{-} + \lambda \cdot X$ with the constant market depth λ and break-even.
- Equilibrium prices follow a martingale.

Speed of Trading and Kyle (1985)

In Kyle (1985), the speed of trading is important neither for informed trader nor noise trader:

- The **informed trader** trades smoothly, but his profits do not depend on the rate of trading.
- The **noise trader** would be better off by smooth out their trading over time, as he would walk up or down the demand curve as a price-discriminating monopolist.

Instantaneous Liquidity Disappears When All Traders Smooth Trading

In unrestricted model, each trader tries to "walk the demand curve" of all the other traders. This slows down trading.

Order flow becomes predictable. The market offers no instantaneous liquidity, and the equilibrium collapses.

Developing intuition about dynamic properties of liquidity in markets with smooth trading is the main focus of our paper.

Theoretical Model Assumptions

We consider dynamic oligopoly model of informed trading in continuous time, to make transparent the idea that each trader trades "smoothly."

- **Overconfidence:** Trade based on "relative" agreement to disagree. Each trader believes his flow of information is more precise than other traders believe it to be.
- **Market Power:** Each trader learns from prices and restricts trading to reduce impact.
- **Symmetry:** No noise traders; no market makers; no rational uninformed traders.

Our model is a fully-fledged dynamic version of Kyle (1989), but trading is based on **agreement-to-disagree**, not noise trading.

Continuous-Time Model: Assumptions

- There are N risk-averse oligopolistic traders, who trade a risky asset with a zero net supply against a risk-free asset.
- Risk-free rate is r.
- A risky asset pays out dividends $D(t)$ growing at a rate $G^*(t)$.

$$dD(t) = -\alpha_D \cdot D(t) \cdot dt + G^*(t) \cdot dt + \sigma_D \cdot dB_D,$$

$$dG^*(t) = -\alpha_G \cdot G^*(t) \cdot dt + \sigma_G \cdot dB_G.$$

Dividends D are observable. Growth rate G^* is unobservable.

Continuous-Time Model: Information Flow

- Let $E_t^n\{\dots\}$ and $var_t^n\{\dots\}$ use all information and n's beliefs:

$$G_n(t) := E_t^n\{G^*(t)\}, \quad \Omega := var_t^n\left\{\frac{G^*(t)}{\sigma_G}\right\}$$

- Each trader observes the **public information** flow $dI_0(t)$ from dividends ($\tau_0 = \sigma_G^2/\sigma_D^2$):

$$dI_0(t) = \tau_0^{1/2} \cdot \frac{G^*(t)}{\Omega^{1/2}\sigma_G} \cdot dt + dB_0.$$

- Each trader n has continuous flow of **private information** $dI_n(t)$ about the unobserved growth rate $G^*(t)$:

$$dI_n(t) = \tau_n^{1/2} \cdot \frac{G^*(t)}{\Omega^{1/2}\sigma_G} \cdot dt + dB_n, \qquad n = 1, \dots, N,$$

- Trader n infers from prices the **average of other signals**

$$I_{-n} := \frac{1}{N-1}\Sigma_{m \neq n} I_m(t).$$

Continuous-Time Model: Bayesian Updating

- Each trader n thinks that his own information has high precision $\tau_n = \tau_H$ and other traders have low precision $\tau_m = \tau_L$, $m \neq n$, with $\tau_H > \tau_L \geq 0$.
- Trader n constructs signals $H_i, i = 0, ..N$ as weighted average of information flow:

$$H_n(t) := \int_{u=-\infty}^{t} e^{-(\alpha_G+\tau)\cdot(t-u)} \cdot dI_n(u),$$

$$\textit{where} \quad \tau = \tau_0 + \tau_H + (N-1)\cdot\tau_L, \qquad \Omega^{-1} = 2\cdot\alpha_G + \tau.$$

$$G_n(t) = \sigma_G \cdot \Omega^{1/2} \cdot \left(\tau_0^{1/2} \cdot H_0(t) + \tau_H^{1/2} \cdot H_n(t) + \sum_{m\neq n} \tau_L^{1/2} \cdot H_m(t) \right)$$

- The importance of each bit of information about G decays exponentially at a rate $\alpha_G + \tau$, the same for every trader.

Continuous-Time Model: Optimization

Each trader n chooses consumption path c_t and trading rate x_t to maximize CARA utility function $U(c_t) = -e^{-A \cdot c_t}$,

$$V(M, S, D, H_0, H_n, H_{-n}) = \max_{\{c_t, x_t\}} \quad E_t \left[\int_{s=t}^{\infty} -e^{-\rho s} \cdot U(c_s) \cdot ds \right].$$

Inventory: $dS(t) = x_t \cdot dt.$

Cash: $dM(t) = (r \cdot M(t) + S(t) \cdot D(t) - c_t - P(x_t) \cdot x_t) \cdot dt.$

Dividends: $dD(t) = -\alpha_D \cdot D(t) \cdot dt + G^*(t) \cdot dt + \sigma_D \cdot dB_D.$

Information Flow Dynamics: $d\hat{H}_n$ and $d\hat{H}_{-n} = \sum_{m \neq n} dH_m$ incorporate H_0 into H_n.

Conjectured Linear Strategies

Trader n conjectures the other $N-1$ traders, $m = 1, \ldots N, m \neq n$, submit symmetric linear demand schedules of the form

$$X_m(t) = \gamma_D \cdot D(t) + \gamma_H \cdot \hat{H}_m(t) - \gamma_S \cdot S_m(t) - \gamma_P \cdot P(t).$$

Let $x_n(t) := dS_n(t)/dt$. Assume continuous single price auction like Kyle (1989). Market clearing quantities are *derivatives* of inventories.

$$P(x_n(t)) = \frac{\gamma_D}{\gamma_P} \cdot D(t) + \frac{\gamma_H}{\gamma_P} \cdot \hat{H}_{-n}(t) + \frac{\gamma_S}{\gamma_P} \frac{1}{N-1} \cdot S_n(t) + \frac{1}{(N-1)\gamma_P} \cdot x_n(t).$$

All traders exercise monopoly power optimally, with "no regret pricing," taking into account how their own trading affects other traders' beliefs about H_{-m} inferred from "fully revealing" prices.

Conjecture: Quadratic Value Function

Value function depends on nine *psi*-parameters:

$$
\begin{aligned}
V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n}) = \\
& -\exp\Big[\psi_0 + \phi_M \cdot M_n + \phi_{SD} \cdot S_n \cdot D \\
& \quad + \frac{1}{2}\psi_{SS} \cdot S_n^2 + \psi_{Sn} \cdot S_n \cdot \hat{H}_n + \psi_{Sx} \cdot S_n \cdot \hat{H}_{-n} \\
& + \frac{1}{2}\psi_{nn} \cdot \hat{H}_n^2 + \frac{1}{2}\psi_{xx} \cdot \hat{H}_{-n}^2 + \psi_{nx} \cdot \hat{H}_n \cdot \hat{H}_{-n}\Big].
\end{aligned}
$$

Note: Wealth does not enter value function; instead have separate components cash $M(t)$ and security holdings $S(t)$.

Equilibrium problem is to solve for nine ψ-parameters and four γ-parameters which sustain a symmetric equilibrium.

Theorem: Symmetric Linear Flow Equilibrium

- Equilibrium defined by solution to six mostly quadratic polynomials in six unknowns.
- Numerical results consistent with "derived" existence condition

$$\frac{\tau_H^{1/2}}{\tau_L^{1/2}} > 2 + \frac{2}{N-2}.$$

Theorem: Symmetric Linear Flow Equilibrium

- Market clearing quantities are derivatives of inventories, implying partial adjustment towards "target inventory", depending on constant C_L,

$$x_n(t) = dS_n(t)/dt = \gamma_S \cdot \Big(C_L \cdot (H_n(t) - H_{-n}(t)) - S_n(t) \Big).$$

- Equilibrium price is dampened average, i.e., $0 < C_G < 1$, of buy and hold valuations based on Gordon's growth formula:

$$P^*(t) = \frac{D(t)}{r + \alpha_D} + \frac{C_G \cdot \frac{1}{N} \sum_{n=1}^{N} G_n(t)}{(r + \alpha_D)(r + \alpha_G)}.$$

Implications

The model explains several realistic features of trading and prices in speculative markets.

Implications: Prices Adjust Immediately

- Prices immediately reveal the average estimate of a growth rate and the average of signals $\sum_{n=1,..N} H_n/N$.

$$P^*(t) = \frac{D(t)}{r + \alpha_D} + \frac{C_G \cdot \frac{1}{N} \sum_{n=1}^{N} G_n(t)}{(r + \alpha_D)(r + \alpha_G)}.$$

- Trading on information continues *after* signals revealed in prices. This is different from Milgrom and Stockey (1982).

Implications: Quantities Adjust Slowly

- Market power with private information makes quantities adjust slowly. Traders "shred orders."
- Trading strategy is "partial adjustment" towards "steady-state" target inventory.

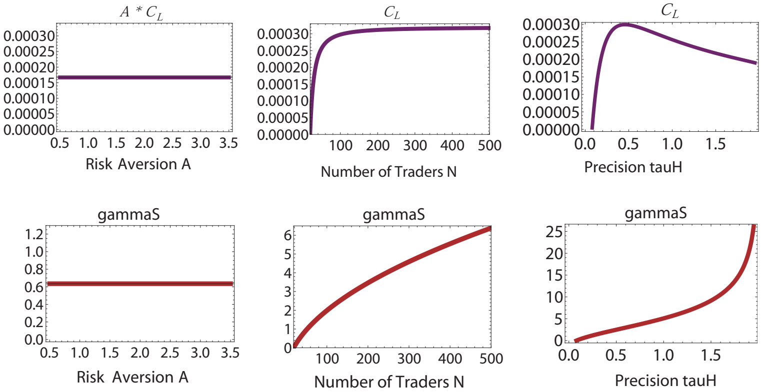
$$x_n^*(t) = \gamma_S \cdot \Big(S^{TI}(t) - S_n(t) \Big).$$

$$S^{TI}(t) = C_L \cdot (H_n(t) - H_{-n}(t)).$$

- "Half-life" of trading $1/\gamma_S$ depends on flow of information into the market. Size of target inventories C_L depends risk-aversion and overconfidence.

Coefficients C_L and γ_S

Coefficients C_L related to target inventories and γ_S related to speed of converging to target inventories as functions of risk aversion A, degree of competition N and overconfidence τ_H, keeping total precision τ constant.



Short-term Trading on Long-Term Information

- Traders acquire private information about fundamentals unfolding over long-term horizons.
- Traders build and reduce their positions over much shorter horizons, when their information become known to other traders.

Implications: Price Impact Functions

- From perspective of each oligopolistic trader, price is linear function of traders inventory (permanent price impact) and derivative of inventory (temporary impact).

$$P(S(t), x(t)) = \lambda_0 + \lambda_S \cdot S(t) + \lambda_x \cdot x(t),$$

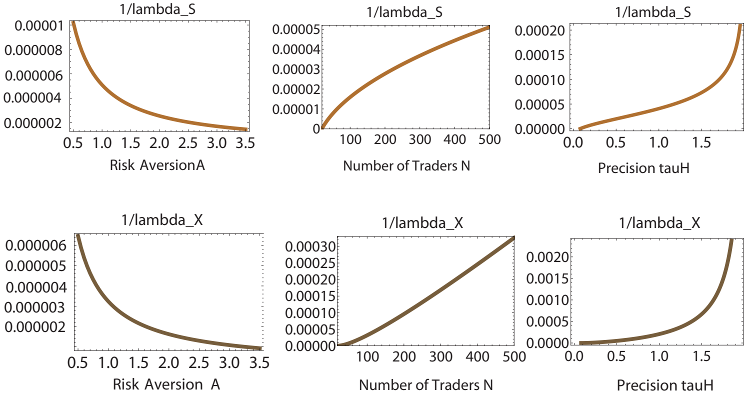
  where constants $\lambda_S = \gamma_s \cdot \lambda_x$ and $dS(t) = x(t) \cdot dt$. Price impact is linear in the first and second derivatives of inventory.

- Permanent and temporary impact is based on "adverse selection," even though price reveals the average of private signals.

- Block traders are infinitely expensive.

Coefficients λ_S and λ_x

Permanent price impact λ_S and temporary price impact λ_x as functions of risk aversion A, degree of competition N and overconfidence τ_H, keeping total precision τ constant.



Implied Execution Cost

- The cost of buying $\tilde{B}$ over time T at uniform rate:

$$E\{\tilde{C}\} = \left(\lambda_S + \frac{\lambda_x}{T/2}\right) \cdot \frac{\tilde{B}^2}{2} = \lambda_S \cdot \left(1 + \frac{1}{\gamma_S} \cdot \frac{1}{T/2}\right) \cdot \frac{\tilde{B}^2}{2}.$$

- The cost of buying $\tilde{B}$ at a constant rate γ:

$$E\{\tilde{C}\} = \left(\lambda_S + \gamma \cdot \lambda_x\right) \cdot \frac{\tilde{B}^2}{2} = \lambda_S \cdot \left(1 + \frac{\gamma}{\gamma_S}\right) \cdot \frac{\tilde{B}^2}{2}.$$

Recall $\lambda_S = \gamma_s \cdot \lambda_x$.

Implied Execution Cost - Summary

- Infinitely fast block trades are infinitely expensive due to temporary price impact costs.
- Infinitely slow trades incur only permanent price impact costs.
- In the equilibrium $(\gamma = \gamma_S)$, the permanent cost $\lambda_S \cdot \tilde{B}^2/2$ is equal to the temporary cost $\lambda_S \cdot \tilde{B}^2/2 \cdot \gamma/\gamma_S$; both are equal to one half of the total cost $E\{C\} = \lambda_S \cdot \tilde{B}^2$.

  Each trader expects to break even and pay out his potential monopoly profits to others in form of temporary price impact.

Literature of Optimal Execution

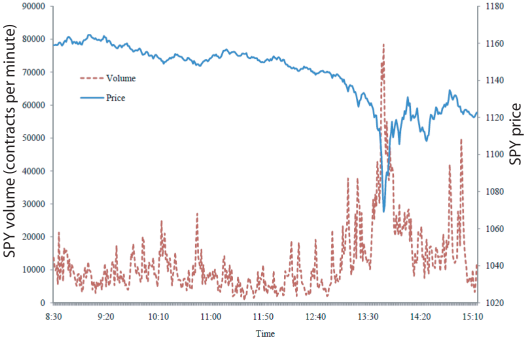
A conventional wisdom is that speed of trading affects incurred transaction costs. Papers on optimal execution study optimal strategies given exogenously specified price impact functions, with a lot of attention on costs due to temporary price impact.

Grinold and Kahn (1999) and Almgren and Chriss (2000) proposed to model price impact functions similar to ours and derived the optimal execution strategy of buying $\bar{S}$:

$$S(t) = \bar{S} \cdot \left(1 - \frac{sinh(k \cdot (T - t))}{sinh(k \cdot T)}\right),$$

where $1/k$ is the "half-life" of executing an order in the absence of any external time constraint.

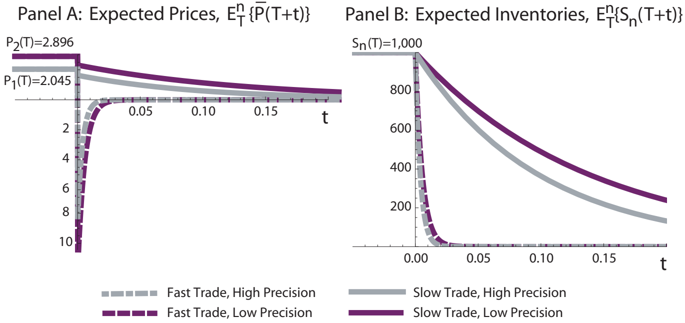
Fast Trading Can Lead to Flash Crashes



- Joint CFTC-SEC report identified a large sale of 75,000 contracts as a trigger for Flash crash on May 6, 2010.
- Kyle and Obizhaeva (2013): invariance implies impact would be 0.60%, which is smaller than the actual decline of 5%; costs were inflated by too fast execution. The order was executed faster than other large orders of similar magnitude.

Fast Trading Can Lead to Flash Crashes

Price path and inventories deviate from the equilibrium if the speed of trading γ deviates from the equilibrium rate γ_S:



$$E_T^n\{\bar{P}(T+t)\} = -\frac{\gamma - \gamma_S}{(N-1)\gamma_P} \cdot e^{-\gamma t} \cdot S_n(T), \qquad E_T^n\{\bar{S}_n(T+t)\} = e^{-\gamma \cdot t} \cdot S_n(T).$$

Speed of Trading and Price Patterns

- Speeding up selling leads to sharp price decline following by V-shaped recovery.
- Slowing up leads to price increase and then convergence to equilibrium level.
- Speeding up execution twice leads to twice bigger price decline relative to the equilibrium price change.
- Speed of recovery depends on price resilience, $\alpha_G + \tau$.
- Selling may occur after prices crash, while the market recovers.

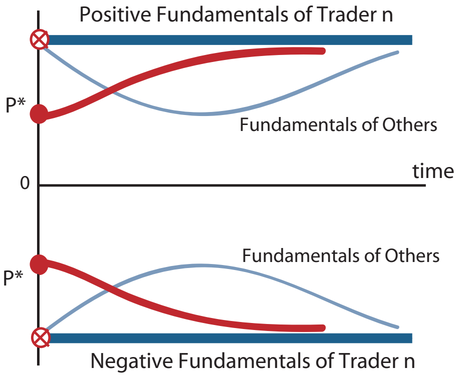
Implications: Dampened Price Reflects a Keynesian "Beauty Contest"

- Our model captures the intuition of the beauty contest:

  For most of these persons are, in fact, largely concerned, not with making superior long-term forecasts of the probable yield on an investment over its whole life, but with foreseeing changes in the conventional basis of valuation a short time ahead of general public.

- In contrast to Keynes(1936), short-term trading dynamics in our paper dampens price volatility.

“Dampening Effect”



Even if all traders agree on the fundamental value today, each trader believes that other traders will find out their information is incorrect in the future. Despite fundamentals, each trader “agrees” on a dampened equilibrium price. This effect does not exist in one-period model.

Keynes (1936) and "Dampening Effect"

Keynes:

- *"It is not sensible to pay 25 for an investment of which you believe the prospective yield to justify value of 30, if you also believe that the market will value it at 20 three months hence."*
- When a trader purchases a stock, he is *"attaching his hopes, not so much to its prospective yield, as to a favorable change in the conventional basis of valuation."*
- *Speculation predominates over enterprise.*

In our dynamic model, traders also seem to be preoccupied with **"short-term price dynamics"** rather than **"hold-to-maturity"** values. The market internalizes overconfidence of traders and corrects equilibrium prices, so instead price deviations are dampened.

Dampened Price Fluctuations

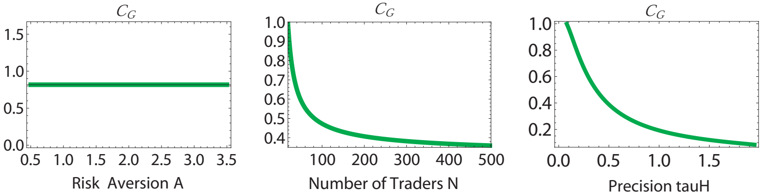
Equilibrium price are similar to Gordon's formula with growth rate being the average of all estimates G_n:

$$P^*(t) = \frac{D(t)}{r + \alpha_D} + \frac{C_G \cdot \bar{G}(t)}{(r + \alpha_D)(r + \alpha_G)}.$$

Since $C_G < 1$, prices are not equal to the average estimates of fundamentals. In the model with agreement to disagree, beauty contest dampens price fluctuations.

Coefficient C_G

Coefficient C_G does not change with risk aversion A. Coefficient C_G decreases with degree of competition N and overconfidence τ_H, keeping total precision $\bar{\tau}$ constant.



Implications: Anomalies

- Price patterns have to be studied from the perspective of the economist, who may assign different precisions to information sources and disagree with traders on how to construct signals.
- Regardless of beliefs, everybody finds anomalies such as mean-reversion and momentum in the model.
- Given reasonably calibrated of parameters, our model provides insights on horizons and magnitudes of anomalies; see Kyle, Obizhaeva, Wang (2014) soon.

Implications: Valuation of Risky Asset

Value function parameters define how the trader values security holdings as compromise between "buy-and-hold" value and "mark-to-market" value.

$$V(M_n, S_n, D, \hat{H}_n, \hat{H}_{-n}) = -\exp\Big[\psi_0 - rA \cdot (M_n + P_n(t) \cdot S_n) + \frac{1}{2}\psi_{nn} \cdot \hat{H}_n^2 + \frac{1}{2}\psi_{xx} \cdot \hat{H}_{-n}^2 + \psi_{nx} \cdot \hat{H}_n \cdot \hat{H}_{-n}\Big].$$

where $P_n(t) = P(t) + \lambda_S \cdot S_n^{TI}(t) - \frac{1}{2}\lambda_S \cdot S_n(t)$.

The "mark-to-market" value $P(t)$ is adjusted for

- the difference between the market price and trader n's appraisal value, $\lambda_S \cdot S_n^{TI}(t)$,
- the cost of liquidating his inventories, $-\frac{1}{2}\lambda_S \cdot S_n(t)$.

H's terms capture the value of trading opportunities.

Implications: Market Liquidity

- Symmetry of strategies does not define liquidity parameter γ_P in $X_n(t) = \gamma_D \cdot D(t) + \gamma_H \cdot H_n(t) - \gamma_S \cdot S_n(t) - \gamma_P \cdot P(t)$.

- Liquidity is "somewhat indeterminate." It reflects a delicate and unstable balance between demand and supply of liquidity.

- Similar delicacy in determination of equilibrium arises in Kyle (1985), where market depth is not determined from the optimization of informed trader either. The second order condition requires the market depth to be constant.

  The level of market depth is determined from the martingale condition.

Flow Equilibrium and Real Markets

The idea that securities markets offer a flow equilibrium rather than a stock equilibrium may seem far-fetched. Yet recent trends are consistent with the way our model predicts liquidity to be supplied and demanded.

- Electronic processing of orders has reduced the fixed costs of executing an order.
- "Order shredding" strategies are similar to strategies in our model. In the future, the market may offer explicit flow equilibrium trading mechanisms.

Flow Equilibrium and Real Markets

- Our model predicts vanishingly small market depth available at any point of time. Market depth is predicted to be available only over time.

- In real markets, market depth available on the "top of the book" is influenced by rules of time and price priority. The smaller the minimum tick size, the smaller is the externality created by those rules and the less of liquidity is expected to be available.

Black (1995)

We formulate in a precise mathematical model ideas that Fischer Black has formulated intuitively in his last paper called "Equilibrium Exchanges," where he outlined his thoughts about how would people trade in the equilibrium, if there are no restrictions on trading strategies or on exchanges.

There is no conventional liquidity available for market orders and conventional limit orders. Traders use indexed limit orders at different levels of urgency. Price moves by an amount increasing in level of urgency. In our model of smooth trading, we formally prove the intuition of Black (1995).

Conclusions

This paper is meant to be a realistic model of the asset management industry, not a “behavioral finance” model.

- We presented a dynamic oligopoly model of informed trading in continuous time with imperfect competition and agreement to disagree.
- In flow equilibrium, speed of trading is derived endogenously.
- Transaction costs depend on the speed of trading.
- Agreement to disagree generates momentum, mean-reversion, and other anomalies.