Reading Group Presentation on “Size Discovery”
by Darrell Duffie and Haoxiang Zhu

David Hao Zhang

October 27, 2016
Introduction

Size-discovery refers to mechanisms by which large transactions can be quickly arranged at *fixed prices*. Examples:

1. Workup, a trading protocol by which buyers and sellers successively increase, or “work up,” the quantities of an asset that are exchanged at a fixed price. Used in market for Treasuries in BrokerTec after each CLOB trade.

2. “Matching sessions,” a variant of workup found in markets for corporate bonds and credit default swaps (CDS). Platform operator chooses workup price.

3. “Block-crossing dark pools,” such as Liquidnet and POSIT, which are predominantly used in equity markets. Concurrent with exchange trading.
Focus of this paper

- Focus of this paper is on how “work up” affects the **efficient allocation** of an asset given *imperfect competition*.
- Price discovery matters only in on its effect on the **allocation**.
- Context:
  - $n$ traders trading an asset which pays out a random per-unit amount $\pi$ with mean $\nu$ at economy’s ending time $T$. $T$ is exponentially distributed with parameter $r$.
  - $n$ is small enough that traders worry about *price impact*.
  - Inventory cost is quadratic $\gamma q^2$, $\gamma > 0$, so efficient allocation has equal inventory.
  - Initial distribution of inventory is not efficient. How to achieve efficient allocation?
Figure 1: *Inventory paths with and without a workup.*
Outline of paper


2. Adding an instantaneous “work up” stage before the double auctions, first use a simple model of Bilateral Trading (Propositions 3 and 4), then a complicated one with Multilateral Trading (Propositions 5 and 6).

3. Compute the “work up” equilibrium, compare with the “no work up” equilibrium.
Comment: “work up” is not instantaneous

Significant volume above 1 second (BrokerTec, cash Treasuries):

Schaumburg and Yang (2016), Liberty Street Economics, FRBNY.
Comment: Non-instantaneous work up may affect price discovery

Schaumburg and Yang (2016), Liberty Street Economics, FRBNY.
Dynamic Trading in Double Auctions, Context

- The trading periods are separated by some clock time ∆, so that the $k$-th auction is held at time $kΔ$.
- Right before auction $k + 1$, trader $i$ receives an incremental inventory shock $w_{i,k+1}$.
- Traders submit a demand schedule in each the $k$-th auction of the form $x_{ik}(\cdot) : \mathbb{R} \to \mathbb{R}$. The market clearing condition is:

$$\sum_i x_{ik}(p_k) = 0 \quad (1)$$

The inventory of trader $i$ thus satisfies the dynamic equation

$$z_{i,k+1} = z_{ik} + x_{ik}(p_k) + w_{i,k+1} \quad (2)$$

The total inventory is $Z_k = \sum_i z_{ik}$.

*Comment: Is this a good approximation for central limit order book trading?*
When choosing a demand schedule in period $k$, each trader maximizes his conditional mean of the sum of two contributions to his final net payoff:

1. The trading profit, which is the realization of the $\pi$ for positions held at time $T$, net of the total purchase cost of the asset in the prior double auctions. The second contribution is a holding cost for inventory.

2. The cost per unit of time of holding $q$ units of inventory is $\gamma q^2$, for a coefficient $\gamma > 0$ that reflects the costs to the trader of holding risky inventory.

For simplicity, set the discount rate to zero.
In summary, for given demand schedules $x_{i1}, x_{i2}, \ldots$, the ultimate net payoff to be achieved by trader $i$, beginning at period $k$, is

$$U_{ik} = \pi z_{i,K(T)} - \sum_{j=k}^{K(T)} p_j x_{ij}(p_j) - \int_{k\Delta}^{T} \gamma z_{i,K(t)}^2 dt$$  \hspace{1cm} (3)$$

where $K(t) = \max\{k : k\Delta \leq t\}$ denotes the number of the last trading period before time $t$. 

Dynamic Trading in Double Auctions, Traders’ incentives formalized
Dynamic Trading in Double Auctions, Traders’ dynamic incentives formalized

For given demand schedules, the continuation utility of trader \(i\) at the \(k\)-th auction, provided it is held before the time \(T\) at which the asset payout is realized, is thus

\[
V_{ik} = E(U_{ik}|\mathcal{F})
\]

where \(\mathcal{F}_{ik}\) represents the information of trader \(i\) just before the \(k\)-th auction. Therefore, the continuation utility of trader \(i\) satisfies the recursion

\[
V_{ik} = -x_{ik}p_k - \gamma \eta (x_{ik} + z_{ik})^2 + (1 - e^{-r\Delta})(x_{ik} + z_{ik})v + e^{-r\Delta}E(V_{i,k+1}|\mathcal{F}_{ik})
\]

where \(\eta\) is the expected duration of time from a given auction until the earlier of the next auction time and the payoff time \(T\):

\[
\eta = \int_{0}^{\Delta} rte^{-rt} dt + e^{-r\Delta} \Delta = \frac{1 - e^{-r\Delta}}{r}
\]
Proposition 1

In the game associated with the sequence of double auctions, there exists a stationary and subgame perfect equilibrium, in which the demand schedule of trader $i$ in the $k$-th auction is given by

$$x_{ik}(p) = a_{\Delta} \left( v - p - \frac{2\gamma}{r} z_{ik} \right)$$  \hspace{1cm} (6)

where

$$a_{\Delta} = \frac{r}{2\gamma} \frac{2(n-2)}{(n-1) + \frac{2e^{-r\Delta}}{1-e^{-r\Delta}} + \sqrt{(n-1)^2 + \frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}}}$$  \hspace{1cm} (7)

The equilibrium price in auction $k$ is:

$$p_k = v - \frac{2\gamma}{nr} Z_k$$  \hspace{1cm} (8)

Comment: Existence, but far from uniqueness.
Sketch of proof of Proposition 1

1. Conjecture a linear strategy: \( x_{ik} = av - bp_k + dz_{ik} \).
2. Plug this in to compute the value function conditional on competitor strategies.
3. Take the FOC of the resulting value function, which must be satisfied by the one-shot deviation principle.
Properties of Proposition 1

- The slope $a_\Delta$ of the equilibrium supply schedule is increasing in $\Delta$. That is, trading is more aggressive if double auctions are conducted at a lower frequency.
- The market-clearing price $p_k$ reveals the total inventory $Z_k$ at the moment of the $k$-th auction.
- The efficient allocation immediately assigns each trader the average inventory $Z_k/n$. The convergence rate is $a_\Delta \frac{2\gamma}{r}$, because by substitution, we have:

$$z_{i,k+1} = z_{ik} - a_\Delta \frac{2\gamma}{r} \left(z_{ik} - \frac{Z_k}{n}\right) + w_{i,k+1} \quad (9)$$
Proposition 2

Let $V_{i,0^+} = E(U_{i0}|z_{i0}, p_0)$ denote the initial utility of trader $i$, evaluated at time 0 after conditioning on the initial market-clearing price $p_0$, which reveals the initial total inventory $Z \equiv Z_0$. We have:

$$V_{i,0^+} = \left[ v \frac{Z}{n} - \frac{\gamma}{r} \left( \frac{Z}{n} \right)^2 \right] + \left( v - 2 \frac{\gamma}{r} \frac{Z}{n} \right) \left( z_{i0} - \frac{Z}{n} \right)$$

$$- \frac{\gamma}{r} \frac{1 - 2a\Delta \frac{\gamma}{r}}{n-1} \left( z_{i0} - \frac{Z}{n} \right)^2 + \Theta$$

Where $\Theta < 0$ is a constant.
Sketch of Proof of Proposition 2

$V_{i,0+}$ can be written in sequence form, which can be computed in two steps:

$$V_{i,0+} = \sum_{k=0}^{\infty} e^{-r\Delta k}$$

$$E \left[ -x_{ik} p^*(Z) + (1 - e^{-r\Delta}) \left( \nu(x_{ik} + z_{ik}) - \frac{\gamma}{r} (x_{ik} + z_{ik})^2 \right) \right] | z_{i0}, Z]$$

Step 1: calculate $V_{i,0+}$ under the assumption that $\sigma_w = 0$.

1. Recognize that without no periodic inventory shocks, inventory evolves deterministically given the strategy in Proposition 1.

   $$z_{i,k+1} = z_{ik} + x_{ik}(p^*(Z)) = \left( 1 - a_\Delta \frac{2\gamma}{r} \right) (z_{ik} - Z/n).$$

2. This gives the first three terms in the expression of $V_{i,0+}$.

Step 2: Calculate the last term $\Theta$ which results from $\sigma_w > 0$. 
Properties of Proposition 2

- The first term of is the total utility of trader $i$ in the event that trader $i$ already holds the initial efficient allocation $Z/n$.

- The second term of is the amount that could be hypothetically received by trader $i$ for immediately selling the entire excess inventory, $z_{i0} - Z/n$, at the market-clearing price, $v - 2\gamma Z/(rn)$.

- The third term is from traders strategically shade their bids to reduce the price impact of their orders.

- The constant $\Theta$ captures the additional allocative inefficiency caused by periodic inventory shocks.
Bilateral workups

1. Consider a setting in which each of an arbitrary number of bilateral workup sessions is conducted between an exogenously matched pair of traders, one with negative inventory, “buyer,” and one with positive inventory, “seller.”

2. Any trader not participating in one of the bilateral workup sessions is active only in the subsequent double-auction market.

3. New assumption: initial inventories have exponential distribution \( \mu \).

4. Any trader’s strategy in the subsequent double-auction market, solved in Proposition 1, depends only on that trader’s inventory level. Thus any public reporting, to all \( n \) traders, of the workup transaction volume plays no role in the subsequent double-auction analysis.
Comment: Do workups really have no price impact?

Figure 4: Permanent Price Impact of Trade

This figure plots the 20-day moving average of the price impact of $1 million buyer-initiated volume transacted during pre-workup versus workup phases. The price impact measures are first computed daily from a VAR(5) model of return and trade flows, and then averaged over rolling 20-day intervals. Estimation is based on BrokerTec data for the on-the-run 2-, 5-, 10- and 30-year Treasury securities over the period 2006-2011. Observations outside the [7:00-17:30] time window are excluded from model estimation.

1. The buyer allows the workup transaction size to increase until the time $T_b$ at which his residual inventory size $S^b + Q(T_b)$ is equal to some threshold $M_b \in \mathbb{R}_+$.

2. The seller likewise chooses a dropout time $T_s$ at which his residual inventory size $S^s - Q(T_s)$ reaches some $M_s \in \mathbb{R}_+$.

3. A threshold equilibrium is a pair $(M_b, M_s) \in \mathbb{R}_+^2$ with the property that $M_b$ maximizes the conditional expected payoff of the buyer given the seller’s threshold $M_s$ and conditional on the buyer’s inventory $S_b$, and vice versa.
Bilateral workups - conjectured equilibrium form illustrated

Illustration of thresholds of residual inventories.

Illustration of thresholds of residual inventories.
We define:

\[
C = \frac{1 - 2a\Delta \gamma}{n - 1} \quad (10)
\]

\[
M = \frac{n - 1}{n + n^2 \frac{C}{1-C} \mu} \quad (11)
\]

Suppose that the workup price $\bar{p}$ satisfies:

\[
|\bar{p} - v| \leq \frac{2\gamma M[C + (1 - C)\frac{3n-2}{n^2}]}{r} \quad (12)
\]
Proposition 3

The workup session has a unique equilibrium in deterministic dropout-inventory strategies. The buyer’s and seller’s dropout levels, $M_b$ and $M_s$, for residual inventory are given by

$$M_b = \frac{n - 1}{n + n^2 \frac{C}{1-C} \mu} + \delta = M + \delta$$  \hspace{1cm} (13)

$$M_s = \frac{n - 1}{n + n^2 \frac{C}{1-C} \mu} - \delta = M - \delta$$  \hspace{1cm} (14)

where $M$ is the dropout quantity for the unbiased price $\bar{p} = \nu$, and where

$$\delta = \frac{r}{2\gamma} \frac{\bar{p} - \nu}{C + (1-C) \frac{3n-2}{n^2}}$$  \hspace{1cm} (15)

That is, in equilibrium, the buyer and seller allow the workup quantity to increase until the magnitude of their residual inventories reach $M_b$ or $M_s$, respectively, or until the other trader has dropped out, whichever comes first.
Sketch proof of Proposition 3

1. Conjecture the equilibrium dropout strategy.
2. Conditional on this conjecture, compute the traders' expectation of the distribution of aggregate inventories $Z$.
   - More specifically, let $F_y$ be the event that the buyer’s requested quantity $y > 0$ is filled. By memoryless property of exponential distributions, the seller’s remaining excess inventory $z_{s0} - M_s - y | F_y$ is exponentially distributed with mean $\frac{1}{\mu}$.
   - Therefore,
     \[
     E(Z | F_y, z_{b0}) = E(z_{b0}) + E(z_{s0} | F_y) = z_{b0} + y + M_s + \frac{1}{\mu}.
     \]
3. Using this distribution, compute $E(U^b(y) | F_y, S^b)$.
4. Use first order condition $\frac{dE(U^b(y) | F_y, S^b)}{dy} = 0$ to find the maximum, yielding the equilibrium strategy.
Discussion of Proposition 3

- The two workup participants do not generally attempt to liquidate all of their excess inventories during the workup, because optimal target inventories are determined by two countervailing incentives.
  1. Because of the slow convergence of a trader’s inventory to efficient levels during the subsequent double-auction market, each trader has an incentive to execute large block trades in the workup.
  2. However, a trader faces winner’s curse regarding the total inventory \( Z \) and the double-auction prices \( p \).

- If \( p > v \), the buyer views the workup price to be less favorable than the expected double auction price. Thus, the buyer is more cautious than the seller in the workup. The opposite is true if \( p < v \). Uncertainty about \( p \) generates \( M_b, M_s > 0 \).

- Workup is strictly Pareto improving.
Figure 2: Immediate inventory imbalance reduction by workup.
Parameters: $n = 5$, $\mu = 1$, $r = 0.1$, $\gamma = 0.05$, $\Delta = 0$, $S^b = -1.5$, $S^s = 2$.
The outcomes of the inventories of traders not entering workup are zero.
Workup vs No Workup - Welfare improvements

The total welfare improvement achieved by workup between the buyer and the seller is 

\[ 2e^{-2M\mu(1 + M\mu)}\mu^2, \]  

which is also decreasing in \( M \) and invariant to the workup price \( \bar{p} \).

Further, based on calculations shown in Appendix C, the fraction of the total inefficiency costs of the buyer and the seller that is eliminated by their participation in the bilateral workup is 

\[ R = \frac{n(n-1)}{4(n-2)}e^{-\frac{2}{n}}\mu\frac{1 + M\mu}{M\mu}. \]  

(25)

(Here again, \( R \) is derived assuming zero inventory shocks after time 0.) Because 

\[ e^{-\frac{2}{n}}\mu\frac{1 + M\mu}{M\mu} \]

is decreasing in \( M \), which in turn is increasing in \( n \), this proportional cost reduction \( R \) decreases with the number \( n \) of market participants. That is, in terms of its relative effectiveness in eliminating inventory-cost inefficiencies caused by imperfect competition in price-discovery markets, workup is more valuable for markets with fewer participants.

For the continuous-time version of the double-auction market (or in the limit as \( \Delta \) goes to zero), we have simply 

\[ R = \frac{3n-2}{4(n-1)}e^{-\frac{n-2}{n}}/n. \]  

(26)

For \( n = 3 \), this cost-reduction ratio is 

\[ R = 0.627. \]  

As \( n \) gets large, \( R \to 0 \).

Figure 3: The proportional welfare improvement of traders participating in workup. The plot shows the fraction \( R \) of the total inefficiency cost of the buyer and the seller that is eliminated by their participation in bilateral workup.
Proposition 4

All else equal, for a given buyer-seller pair, the probability of having a positive-volume workup and the expected workup volume are higher if:

1. The frequency of subsequent double auctions is higher ($\Delta$ is smaller).
2. The number $n$ of traders is lower.
3. The mean arrival rate of asset payoff news $r$ is lower.
Proposition 4 - empirical predictions

1. A proxy for frequency of double auctions is inter-trade time. A speed “upgrade” corresponds to a smaller $\Delta$.
2. The number $n$ of traders could be estimated by the number of active (or sufficiently active) participants on a particular electronic trading platform, or some measure of concentration.
3. Finally, the mean rate $r$ of payoff arrival information may be proxied by the arrival rate of important news, even scheduled news such as a scheduled press release of the Federal Open Market Committee (FOMC), a macroeconomic data release, or an earnings announcement.

Comment: proxy #3 seems especially problematic.
After each buyer exits the workup, there is a new buyer with probability $q$, and likewise for sellers. Specifically, for any non-negative integer $k$, $N_b$ number of buyers, and $N_s$ number of sellers:

$$P(N_b = k) = P(N_s = k) = f(k) = q^k (1 - q)$$

for some $q \in (0, 1)$. We have $E(N_b) = E(N_s) = \frac{q}{1 - q}$.

The number of market participants is then known to all in the subsequent double auction, for tractability.
The workup begins by pairing the first buyer and first seller.

During the workup, the exit from workup of the i-th buyer causes the (i + 1)-st buyer to begin workup, provided \( N_b > i \). The (i + 1)-st buyer can then choose whether to begin actively buying or to immediately drop out without trading.

Similarly, when seller \( j \) exits, he is replaced with another seller if \( N_s > j \).

The workup ends when buyer number \( N_b \) exits or when seller number \( N_s \) exits, whichever is first.
At any given point during the workup, the state vector on which the equilibrium strategies depend is of the form \((m, X, y)\), where:

- \(m\) is the total number of buyers and sellers that have already entered workup, including the current buyer and seller.
- \(X\) is the total expected inventory held by previously exited participants.
- \(y\) is the quantity that the current workup pair has already executed.

We let \(M_b(m, X) > 0\) and \(M_s(m, X) > 0\) be the conjectured dropout thresholds of the current buyer and seller, respectively, in a workup state \((m, X, y)\) that is active, meaning \(y > 0\).
Proposition 5

Suppose that \( \bar{\rho} = \nu \). A necessary condition for a Markov equilibrium is that the inventory dropout thresholds of the buyer and the seller in the current active workup are, respectively:

\[
M_b(m, X) = M^*(m) + L(m)X
\]  
(17)
\[
M_s(m, X) = M^*(m) - L(m)X
\]  
(18)

where, letting \( g(k) = (k + 1)q^k(1 - q)^2 \) and \( n = m + k \).

\[
M^*(m) = \frac{1}{\mu} \frac{\sum_{k=0}^{\infty} g(k)(1 - C(n))^{n-1}}{\sum_{k=0}^{\infty} g(k) \left( C(n) + \frac{1-C(n)}{n} \right)}
\]  
(19)
\[
L^*(m) = \frac{\sum_{k=0}^{\infty} g(k) \frac{1-C(n)}{n}}{\sum_{k=0}^{\infty} g(k) \left( C(n) + \frac{(1-C(n))(3n-2)}{n^2} \right)}
\]  
(20)
Proposition 6

Define monotonicity and positivity conditions as:

\[ M_b(m + 1, X') \geq M_b(m, X) \] (21)
\[ M_s(m + 1, X') \geq M_s(m, X) \] (22)
\[ M_b(m, X) \geq 0 \] (23)
\[ M_s(m, X) \geq 0 \] (24)

If \( e^{-r\Delta} > 1/2 \), the monotonicity and nonnegativity conditions are satisfied, and the strategies given in Proposition 5 constitute the unique Markov workup equilibrium.

Comment: Here, more participants in workup signals more traders (more competition/better efficiency) in double-auction round, by memoryless property. Is this realistic?
Conclusion

- Comparing the possible trading mechanisms, we can write:
  \[ \text{workup} + \text{double auctions} \succeq \text{double auctions only} \succeq \text{workup only} \]

- What is the “optimal” mechanism?