NBER WORKING PAPER SERIES

SIZE DISCOVERY

Darrell Duffie
Haoxiang Zhu

Working Paper 21696
http://www.nber.org/papers/w21696

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
November 2015

For helpful discussions and comments, we thank Songzi Du, Eiichiro Kazumori, Pete Kyle, and Romans Pancs, as well as participants at the 2015 Annual Meetings of the American Economics Association and the University of Bonn. We are especially grateful for research assistance by Jun Yan, of the Stanford Statistics Department, and Hyungjune Kang, of the MIT Sloan School of Management. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Size Discovery
Darrell Duffie and Haoxiang Zhu
NBER Working Paper No. 21696
November 2015
JEL No. D47,D82,G14

ABSTRACT

Size discovery refers to the use of trade mechanisms by which large quantities of an asset can be exchanged at a price that does not respond to price pressure. Primary examples of size discovery include "workup," a trade protocol used in the markets for U.S. Treasuries and swaps, and block-trading "dark pools," used in equity markets. By freezing the execution price, a size-discovery mechanism does not clear the market, but overcomes large investors' concerns over their price impacts. Price-discovery mechanisms, which determine a market-clearing price by matching supply and demand, cause investors to internalize their price impacts, inducing costly delays in the reduction of position imbalances. We show that augmenting a price-discovery mechanism with a size-discovery mechanism such as workup or dark pools improves allocative efficiency. Because of adverse selection regarding the order imbalances of other investors, size discovery is used only by investors with large position imbalances.

Darrell Duffie
Graduate School of Business
Stanford University
Stanford, CA 94305-7298
and NBER
duffie@stanford.edu

Haoxiang Zhu
MIT Sloan School of Management
100 Main Street, E62-623
Cambridge, MA 02142
and NBER
zhuh@mit.edu

1 Introduction

This paper shows that size-discovery mechanisms, by which large transactions can be quickly arranged at fixed prices, improve allocative efficiency in markets with imperfect competition and private information over latent supply or demand imbalances.

An important aspect of market liquidity is the ability to quickly buy or sell large quantities of an asset with a small price impact. Price impact is primarily a concern of large strategic investors, and not of small or "price-taking" investors. Those particularly worried about price impact therefore include large institutional investors such as mutual funds, pension funds, and insurance companies. Price impact also concerns major financial intermediaries such as broker-dealers, who often absorb substantial inventory positions in primary issuance markets or from their client investors, and then seek to offload these positions in inter-dealer markets. For example, Duffie (2010) surveys widespread evidence of substantial price impact around large purchases and sales, even in settings with relatively symmetric and transparent information.

Under imperfect competition, the strategic avoidance of price impact is a major cause for allocative inefficiency in markets that offer price discovery, such as sequential double auctions or central limit order books. As we will explain, in order to reduce price impacts in price-discovery markets, investors "shade" their order sizes, meaning that they only partially express their true trading interest at any moment and at any given price. This is manifested, for example, by splitting large orders and executing them piecemeal over time, which can involve costly execution delays.

We argue that size discovery is an effective way to mitigate the allocative inefficiency caused by strategic avoidance of price impact. Examples of size-discovery mechanisms used in practice include:

- 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. Each participant in a workup has the option to drop out at any time. In the market for U.S. Treasuries, Fleming and Nguyen (2013) find that workup accounts for 43% to 56% of total trading volume on a typical day. Workup has been increasingly adopted for trading standardized over-the-counter derivatives.[1] Workup is the primary example of size discovery modeled in this paper.
- Block-crossing "dark pools," which are predominantly used in equity markets. In a typical "midpoint" dark pool, buyers and sellers match orders at the midpoint of the best bid price and best offer price shown on transparent exchanges. Dark-pool allocations are either by time priority or by pro-rata rationing on the heavy side of the market. Dark pools account for about 15% of trading volume in the U.S.

[1] For examples of workup protocols in "swap execution facilities" (SEFs), see Tradeweb (2014) and Tradition (2015).

equity markets. Certain dark pools offer limited price discovery. Others do not use price discovery at all. For an overview, see Zhu (2014). We examine dark pools as a secondary example of size discovery.

Despite some institutional differences discussed later in the paper, workups and mid-point dark pools share the key feature of size discovery: crossing orders at fixed prices without price impact. Although aware of the trade price, the participants in a workup or dark pool are uncertain of how much of the asset they will be able to trade at that price, which is not sensitive to their demands. One side of the market is eventually rationed, being willing to trade more at the given price. Thus, a size-discovery mechanism cannot clear the market, and is therefore inefficient on its own. Nevertheless, precisely by giving up on market clearing, a size-discovery mechanism reduces the adverse effect of investors' strategic incentives to dampen their immediate demands. We show, as a consequence, that a market design combining size discovery and price discovery offers substantial efficiency improvement over a market that relies only on price discovery. In particular, we demonstrate that in a market with imperfect competition and private information concerning order imbalances, allocative efficiency is improved by adding a size-discovery mechanism, modeled as a workup, before a price discovery mechanism, modeled as a sequential-double-auction market.

Our modeling approach and the intuition for our results can be roughly summarized as follows. An asset pays a liquidating dividend at a random, exponentially-distributed future time. Before this time, double auctions for the asset are held among n strategic traders at evenly spaced time intervals of some length Δ. Thus, the auctions are held at times 0, Δ, 2Δ, and so on. Before the first of these auctions, the inventory of the asset held by each trader has an undesired component, positive or negative, that is not observable to other traders. Each trader suffers a continuing cost that is increasing in his undesired inventory imbalance. In each of the successive double auctions, traders submit demand schedules. The market operator aggregates these demand schedules and calculates the market-clearing price, at which total demand and supply are matched.

Because there are only finitely many traders, each of them "shades" his demand schedule in order to account for his own impact on the market-clearing price. For example, each trader who wishes to sell submits a supply schedule that expresses, at each price, only a fraction of his actual trading interest in order to reduce his own downward pressure on the market-clearing price. The unique efficient allocation is that giving each trader the same magnitude of undesired inventory. At each successive double auction, however, traders' inventories adjust only partially toward this efficient allocation. This leaves substantial scope for inefficiency. That is, traders may bear significant costs of imbalanced inventories, relative to the efficient allocation. These excess costs are not reduced by holding more frequent auctions. As shown by Vayanos (1999) and Du and

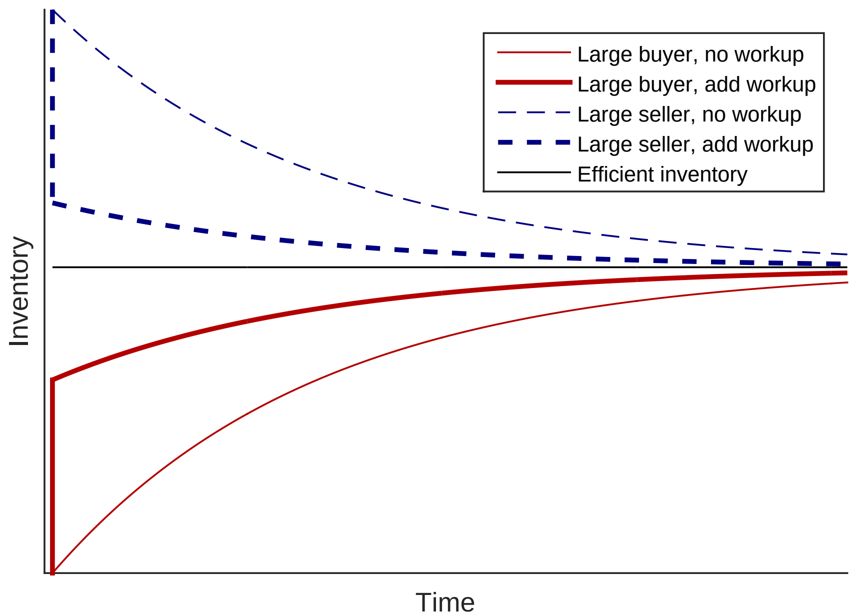

Figure 1: *Inventory paths with and without a workup.* The thin-line plots are the equilibrium inventory paths of a buyer and a seller in sequential-double-auction market. Plotted in bold are the equilibrium inventory paths of the same buyer and seller in a market with a workup followed by the same sequential-double-auction market.

Zhu (2014), even if trading is continuous, convergence to the efficient allocation is not instantaneous because traders' strategic incentives get stronger as the trading volume per round becomes smaller.

In summary, because of strategic bidding behavior and imperfect competition, the sequential-double-auction market is slow in reducing allocative inefficiencies. This point is well recognized in prior work, including the static models of Vives (2011), Rostek and Weretka (2012), and Ausubel, Cramton, Pycia, Rostek, and Weretka (2014), as well as the dynamic models of Vayanos (1999) and Du and Zhu (2014).

Figure 1 illustrates the time paths of inventories of a buyer and a seller for a parametric case of our sequential-double-auction market that we present later in the paper. The two thin-line plotted curves in Figure 1 illustrate the convergence over time of the inventories to those of the efficient allocation. (The inventory plot is for the corresponding continuous-time market, whose behavior we show to be equal to the limit behavior of the discrete-time model as the inter-auction time interval Δ shrinks to zero.)

Now, consider an alternative market design in which traders have the opportunity to conduct a workup before the first double auction. For simplicity of exposition, we first solve the equilibrium for bilateral workups, whose participants are chosen exogenously. Naturally, each pair comprises a trader with a negative inventory imbalance, the "buyer," and a trader with a positive inventory imbalance, the "seller." We then extend the equilibrium solution method to multilateral workups (with many buyers and sellers taking turns to trade at the same fixed price). Any trader who does not enter a workup participates only in the subsequent double auctions.

The fixed workup price is set exogenously. (We show that our efficiency results are robust to the choice of workup price.) As mentioned, the quantity to be exchanged between the buyer and seller in the workup is raised continually until the point at which one of the two traders drops out. That dropout quantity of the asset is then transferred from the seller to the buyer at the fixed workup price. Because the workup price is fixed, neither the buyer nor the seller is concerned about price impact. They are therefore able to exchange a potentially large block of the asset immediately, leading to a significant reduction in the total cost of maintaining undesired inventory over time.

Each workup participant recognizes that it is revealing information about his own inventory that is adverse to his interests. In equilibrium, we show that this effect inhibits traders with smaller inventory imbalances from entering workup. For the same reason, a "large trader" that enters a workup continues to work up the size of the trade only until his inventory imbalance falls to some interior endogenous threshold.

This endogenous dropout quantity is determined by two countervailing incentives. On one hand, each trader wishes to minimize the inventory size that is taken into the sequential-double-auction market, because these leftover inventories take time to optimally liquidate, involve price-impact costs, and in the meantime are accompanied by holding costs. On the other hand, each trader in a workup faces adverse selection regarding the inventory size of his workup counterparty. For example, conditional on the buyer being the first to drop out of the workup, the buyer infers that the seller's inventory must be larger than its unconditional mean. Given this, the buyer's conditional expectations of the subsequent double-auction prices are lower than the workup price. This information implies that the buyer should withhold some quantity from the workup, and reserve it for execution in the double-auction market at more favorable conditional expected prices. In equilibrium, these two effects determine a unique inventory threshold for dropping out of workup, which we calculate explicitly.

The two thick lines in Figure 1 illustrate the welfare-improving effect of augmenting the market design with an initial workup, which causes an instant reduction in inventory imbalances.

Comparative statics reveal that the workup accounts for a larger fraction of total trade volume if the double auctions are run more frequently, if the arrival of payoff-relevant information is less imminent, or if there are fewer traders in the market. Under any of these conditions, traders are more sensitive to price impact because they will liquidate their inventories more gradually, implying higher inventory holding costs. It is precisely under these conditions that adding workup to the market design offers the greatest improvement in allocative efficiency.

The same intuition applies in a multilateral workup. The price is fixed for the entire workup session, which begins (if at all) with a workup between the first buyer and seller in their respective queues. Eventually, either the currently active buyer or seller drops out. If, for example, the seller is the first to drop out, it is then revealed whether there is at least one more trader remaining in the seller's queue, and if so whether that seller wishes to continue selling the asset at the same price. Based on this information, the buyer may continue the workup or may choose to drop out and be replaced by another buyer, if there is one, and so on. This process continues until there are no more buyers or no more sellers, whichever happens first. The equilibrium is solved in terms of the dropout threshold for the remaining inventory of an active workup participant, which is updated as each successive counterparty drops out and is replaced with a new counterparty. For example, when a new seller arrives and begins to actively increase the workup quantity, the current buyer's conditional expectation of the total market-wide supply of the asset jumps up, and this causes the buyer's dropout threshold to jump up at the same time by an amount that we compute and that depends on the history of prior workup observations. That is, with the arrival of a new active replacement seller, the buyer infers that the conditional expected double-auction price has become more favorable, and holds back more inventory from the workup, reserving a greater fraction of its trading interest for the double-auction market.

Our main result is that adding a size-discovery mechanism like workup to a price-discovery mechanism like double auctions is a Pareto improvement over a market with only the price-discovery mechanism. Traders who execute a positive quantity in the workup strictly benefit from it, and traders who only participate in the double auctions are not harmed by the introduction of workup. Moreover, the efficiency improvement is greater to the degree that the initial allocation is worse. That is, size discovery is most helpful when order imbalances are highly concentrated among market participants, including both buyers and sellers.

As far as we are aware, our paper is the first to explicitly model how a size-discovery mechanism reduces allocative inefficiency caused by strategic demand reduction in price-discovery markets. The only prior theoretical treatment of workup, to our knowledge, is by Pancs (2013), who shows that a bilateral workup reduces front-running because the workup does not reveal the full magnitudes of the trading interests of individual market participants. We are also the first to solve for equilibrium behavior in multilateral workup markets. Empirical analyses of workup include those of Boni and Leach (2002, 2004), Dungey and McKenzie (2013), Fleming and Nguyen (2013), and Huang, Cai, and Wang (2002). Previous models of dark pools include those of Hendershott and Mendelson (2000), Degryse, Van Achter, and Wuyts (2009), Zhu (2014), and Buti, Rindi, and Werner (2015). These studies focus on the effect of dark trading on price discovery, liquidity, and certain metrics of welfare.[2] None of the studies mentioned above address the size-discovery benefit that we model here.

2 Dynamic Trading in Double Auctions

This section models dynamic trading in a flexible-price market consisting of a sequence of double auctions. The double-auction model is adapted from Du and Zhu (2014). Once we have solved this model, the associated indirect utilities for pre-auction inventory imbalances serve as the terminal utility functions for the prior size-discovery stage, which is modeled as a workup and introduced in the next section.

We fix a probability space and the time domain $[0, \infty)$. Time 0 may be interpreted as the beginning of a trading day. The market is populated by $n \geq 3$ risk-neutral agents trading a divisible asset. The payoff π of the asset, a random variable with some finite mean v, will be revealed publicly at a random time T that is exponential with parameter r. Thus $E(T) = 1/r$. There is no further incentive to trade once π is revealed at time T. Everyone has symmetric information about π.

The n traders' respective asset inventories at the beginning of the double-auction market are given by a vector $z_0 = (z_{10}, z_{20}, \ldots, z_{n0})$ of random variables that have non-zero finite variances. While the individual traders' inventories may be correlated with each other, there is independence among the asset payoff π, the revelation time T, and the vector z_0 of inventories.

At each nonnegative integer trading period $k \in \{0, 1, 2, \ldots\}$ a double auction is used to reallocate the asset. The trading periods are separated by some clock time $\Delta > 0$, so that the k-th auction is held at time $k\Delta$. As the first double auction begins, the information available to trader i includes the initial inventory z_{i0}, but does not include[3] the total inventory $Z = \sum_i z_{i0}$. This allows that some traders may be better informed about Z than others, and may have information about Z going beyond their own respective inventories. For simplicity, there are no inventory shocks after time 0. This assumption is not necessary in order to explicitly solve the double auctions, but improves the tractability of our later equilibrium analysis of the workup. In the next section we will make some additional assumptions about the joint probability distribution of initial trader inventories that provide tractability for the workup model.

Appendix B solves a continuous-time version of the double-auction model, showing that its equilibrium is equal to the limit of the equilibria in this discrete-time setting as

[2]Empirical studies of dark pools include Buti, Rindi, and Werner (2011), Ready (2014), and Menkveld, Yueshen, and Zhu (2015), among many others.

[3]Fixing the underlying probability space $(\Omega, \mathcal{F}, P)$, trader i is endowed with information given by a sub-σ-algebra $\mathcal{F}_{i0}$ of $\mathcal{F}$. The inventory z_{i0} is measurable with respect to $\mathcal{F}_{i0}$, whereas the total inventory Z has a non-zero variance conditional given $\mathcal{F}_{i0}$.

the time Δ between auctions goes to zero.

The inventory z_{ik} of trader i before the k-th auction is the sum of z_{i0} and the cumulative purchases and sales of the asset of that trader in prior auctions. The information available to trader i at period k consists[4] of the trader's initial information, the sequence $p_0, \ldots, p_{k-1}$ of prices observed in prior auctions, as well as the trader's current and lagged inventories, $z_{i0}, \ldots, z_{ik}$. At the k-th auction, trader i submits a continuous and strictly decreasing demand schedule, which could in principle depend on all of this available information. We focus on equilibria in which the demand schedule chosen by trader i optimally depends only on the trader's current inventory z_{ik}. That is, trader i submits a demand schedule of the form $x_{ik}(\,\cdot\,; z_{ik}) : \mathbb{R} \to \mathbb{R}$, which is an agreement to buy $x_{ik}(p_k; z_{ik})$ units of the asset at the unique market-clearing price p_k. Whenever it exists, this market clearing price p_k is defined by

$$\sum_i x_{ik}(p_k; z_{ik}) = 0. \tag{1}$$

The inventory of trader i thus satisfies the dynamic equation

$$z_{i,k+1} = z_{ik} + x_{ik}(p_k; z_{ik}). \tag{2}$$

This double-auction mechanism is typical of those used at the open and close of the day on equity exchanges.[5] The double-auction model captures the basic implications of a flexible-price market in which traders are rational and internalize the equilibrium price impacts of their own trades. In practice, participants in a multi-unit auction submit a package of limit orders rather than a demand function. A demand function can be approximated as the collection of a large number of limit orders at closely spaced limit prices.

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. The first contribution is trading profit, which is the final payoff of the position held when π is revealed 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. The cost per unit of time of holding q units of inventory is γq^2, for a coefficient $\gamma > 0$ that reflects the costs to the trader of holding risky inventory. In summary, for given demand schedules $x_{i1}(\,\cdot\,), x_{i2}(\,\cdot\,), \ldots$, the ultimate net payoff to be achieved by trader i, beginning at period k, is

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

[4]That is, the σ-algebra with respect to which the demand schedule of trader i in the k-th auction must be measurable is the join of the initial σ-algebra $\mathcal{F}_{i0}$ and the σ-algebra generated by $\{p_0, \ldots, p_{k-1}, z_{i1}, \ldots, z_{ik}\}$.

[5]See, for example, http://www.nasdaqtrader.com/content/ProductsServices/Trading/Crosses/fact_sheet.pdf.

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

$$V_{ik} = E\left(U_{ik} \;\middle|\; \mathcal{F}_{ik}\right), \tag{4}$$

where $\mathcal{F}_{ik}$ represents the information of trader i just before the k-th auction.

Even though they do not have direct aversion to risk, broker-dealers and asset-management firms have extra costs for holding inventory in illiquid risky assets. These costs may be related to regulatory capital requirements, collateral requirements, financing costs, agency costs related to the lack of transparency of the position to higher-level firm managers or clients regarding true asset quality, as well as the expected cost of being forced to raise liquidity by quickly disposing of remaining inventory into an illiquid market. Our quadratic holding cost assumption is common in models of divisible auctions, including those of Vives (2011), Rostek and Weretka (2012), and Du and Zhu (2014).

For simplicity, we have ignored time preferences. This is a reasonable approximation for trader inventory management in practice, at least if market interest rates are not extremely high, because traders lay off excess inventories over relatively short time periods, typically measured in hours or days.

Based on this calculation, the continuation utility of trader i given by (4) satisfies the recursion

$$V_{ik} = -x_{ik}(p_k; z_{ik})p_k - \gamma\eta z_{i,k+1}^2 + (1 - e^{-r\Delta})vz_{i,k+1} + e^{-r\Delta}E(V_{i,k+1} \,|\, \mathcal{F}_{ik}), \tag{5}$$

where η is the expected duration of time from a given auction (conditional on the event that the auction is before T) until the earlier of the next auction time and the payoff time T. We have the calculation

$$\eta = \int_0^\Delta rt\, e^{-rt}\, dt + e^{-r\Delta}\Delta = \frac{1 - e^{-r\Delta}}{r}. \tag{6}$$

The four terms on the right-hand side of (5) represent, respectively, the payment made in the k-th double auction, the expected inventory cost to be incurred in the subsequent period (or until the asset payoff is realized), the expectation of any asset payment to be made in the next period multiplied by the probability that T is before the next auction, and the conditional expected continuation utility in period $k + 1$ multiplied by the probability that T is after the next auction.

We will consider the equilibrium in which the first auction completely reveals the entire path of future equilibrium inverse-demand schedules, so that each trader's information is fully captured by that trader's inventory.

Each trader takes the total of the demand functions of the other traders as given. By the Bellman principle of dynamic optimality, in each period k, trader i is conjectured to submit a demand schedule $x_{ik}(\,\cdot\,; z_{ik})$ that maximizes the right-hand side of (5) subject to the dynamic equation (2), when taking $V_{i,k+1} = \mathcal{V}(z_{i,k+1})$, for some indirect inventory value function $\mathcal{V} : \mathbb{R} \to \mathbb{R}$. This indirect inventory value function $\mathcal{V}(\,\cdot\,)$ must satisfy (5) for this maximizing demand function, when taking $V_{ik} = \mathcal{V}(z_{ik})$. The equilibrium given by the following proposition is established by a standard Bellman verification argument, which implies the optimality of the corresponding demand functions. For this, we use the fact that, for the candidate optimal demand schedule, $e^{-rk\Delta}\mathcal{V}(z_{ik})$ converges to 0 as k goes to infinity, and the fact that $\limsup_{k\to\infty} e^{-rk\Delta}\mathcal{V}(z_{ik}) \leq 0$ for any feasible demand strategy. Details are in Appendix A.

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*

$$x_{ik}(p_k; z_{ik}) = a_\Delta \left(v - p_k - \frac{2\gamma}{r} z_{ik} \right), \tag{7}$$

where

$$a_\Delta = \frac{r}{2\gamma} \frac{2(n-2)}{(n-1) + \frac{2e^{-r\Delta}}{1-e^{-r\Delta}} + \sqrt{(n-1)^2 + \frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}}}. \tag{8}$$

The equilibrium price in any of the auctions is

$$p^* = v - \frac{2\gamma}{nr} Z. \tag{9}$$

The bidding strategies of this equilibrium are ex-post *optimal with respect to all realizations of inventory histories. That is, trader j would not strictly benefit by deviating from the equilibrium strategy even if he were able to observe the history of other traders' inventories, $\{z_{im} : i \neq j, m \leq k\}$.*

The ex-post optimality property of the equilibrium arises from the fact that each trader's marginal indirect value for additional units of the asset depends only on his own current inventory, and not on the inventories of other traders. This property will be useful in solving the workup equilibrium.

The slope a_Δ of the equilibrium supply schedule is increasing in Δ. That is, trading is more aggressive if double auctions are conducted at a lower frequency. We also have

$$\lim_{\Delta\to\infty} a_\Delta = \frac{r(n-2)}{2\gamma(n-1)} < \frac{r}{2\gamma}. \tag{10}$$

Moreover, as Δ goes to 0, a_Δ converges to 0.

The market-clearing price p^* reveals the total inventory Z in the first auction.[6] Although traders have symmetric information about the asset fundamental, uncertainty about the total inventory Z generates uncertainty about the market-clearing price. In this sense, this model also captures some degree of adverse selection. As we will see in the next section, uncertainty over Z is an important determinant of the optimal strategy in the workup stage of the model.

The full revelation of Z does not imply that the allocation is efficient. By symmetry and by the linearly decreasing nature of marginal values, the efficient allocation immediately assigns each trader the average inventory Z/n, whereas the double-auction market merely moves the allocation toward this equal distribution of the asset, as indicated in the following result.

Proposition 2. *In the equilibrium of Proposition 1, the inventory of each trader converges exponentially to the efficient level, which is the average inventory. That is, the excess inventory $z_{ik} - Z/n$ of trader i satisfies*

$$z_{ik} - \frac{Z}{n} = \left(1 - \frac{2a_\Delta \gamma}{r}\right)^k \left(z_{i0} - \frac{Z}{n}\right). \tag{11}$$

Moreover, the conditional expectation $V_{i,0+} = E(U_{i0} \,|\, z_{i0}, p_0)$ of the final net value of trader i, evaluated at time 0 after the market-clearing price p^* reveals the total inventory Z, is given by

$$V_{i,0+} = \left[v\frac{Z}{n} - \frac{\gamma}{r}\left(\frac{Z}{n}\right)^2\right] + \left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right)\left(z_{i0} - \frac{Z}{n}\right) - \frac{\gamma}{r}\frac{1 - 2a_\Delta \frac{\gamma}{r}}{n-1}\left(z_{i0} - \frac{Z}{n}\right)^2. \tag{12}$$

Proposition 2 shows the extent to which the double-auction allocations are inefficient. This motivates the introduction of workup, which we model in the next section. The excess inventory $z_{ik} - Z/n$ of trader i, relative to the efficient allocation, shrinks by a constant fraction $1 - 2a_\Delta \gamma/r$ per trading period. For a small inter-trade time Δ, this fractional reduction in the mis-allocation of inventory is correspondingly small.

The allocative inefficiency of the double-auction market is further illustrated by (12), whose first term is the total utility of trader i in the event that trader i already holds the efficient allocation Z/n. The second term of (12) is the amount that could be

[6] Because the total inventory Z does not change over time, after the completion of the first auction, there exist equilibria for every subsequent auction in which demands depend explicitly on Z, or equivalently on the clearing price of the first double auction. (Vayanos (1999) show that in a similar model with complete information and strategic trading, there exist a continuum of equilibria.) However, in a variation of our model in which the aggregate inventory is a non-trivial stochastic process $\{Z_k\}$, suffering noisy shocks in each period, a period-k bidding strategy that depends on past total inventories is generally not ex-post optimal, in that revealing other traders' inventories would provide an incentive to deviate from the equilibrium demand strategy. The stationary equilibrium that we have selected remains ex-post optimal even in the presence of revealed noisy changes in the total asset supply.

hypothetically received by trader i for immediately selling the entire excess inventory, $z_{i0} - Z/n$, at the market-clearing price, $v - 2\gamma Z/(rn)$. But this immediate beneficial reallocation of the asset does not actually occur because traders strategically shade their bids to account for the price impact of their orders. This price-impact-induced drag on each trader's expected ultimate net payoff, or "utility," is captured by the last term of (12), which is the utility loss caused by the fact that the demand schedule of trader i in each auction is increasing in the squared excess inventory $(z_{i0} - Z/n)^2$ and decreasing in a_Δ. Thus, as formalized in Appendix C, the sum across traders of their expected squared excess inventories is a natural welfare metric.

Because a_Δ is always smaller than $r/(2\gamma)$, full efficiency cannot be achieved. Moreover, because a_Δ is increasing in Δ, each trader's utility loss gets larger as Δ gets smaller. The basic intuition is as follows (see Du and Zhu (2014) for a detailed discussion). Although a smaller Δ gives traders more opportunities to trade, they are also strictly less aggressive in each trading round. A smaller Δ makes allocations less efficient in early rounds but more efficient in late rounds. Traders value early-rounds efficiency more because of the effective "time discounting" $e^{-r\Delta}$. The net effect is that allocative efficiency is worse if Δ is smaller.

Given that many actual functioning markets operate nearly continuously (during business hours), we also consider a continuous-time version of this model, analyzed in Appendix B, whose payoff relevant quantities are equal to the limits of their discrete-time analogues as Δ goes to zero. Here, we simply record that limiting behavior.

Corollary 1. *As $\Delta \to 0$, the equilibrium of Proposition 1 converges to the following continuous-time limit.*

1. *The limit demand schedule*[7] *of trader i at time t is*

$$x_{it}^{\infty}(p; z_{it}^{\infty}) = a^{\infty}\left(v - p - \frac{2\gamma}{r} z_{it}^{\infty}\right), \tag{13}$$

where

$$a^{\infty} = \frac{(n-2)r^2}{4\gamma} \tag{14}$$

[7]In a continuous-time setting, a demand schedule at time t can be expressed by a demand "rate function" $D_t(\cdot)$, which means that if the time path of prices is given by some function $\phi : [0, \infty) \to \mathbb{R}$, then the associated cumulative total quantity purchased by time t is $\int_0^t D_s(\phi(s))\, ds$, whenever the integral is well defined. In our case, the discrete-period demand schedule $x_{ik}(\cdot\,; z_{ik})$ has the indicated limit demand schedule, as a demand rate function, because $z_{i,K(t)} \to z_{it}^{\infty}$ and because, for any fixed price p and fixed inventory level z,

$$\lim_{\Delta \downarrow 0} \frac{a_\Delta}{\Delta}\left(v - p - \frac{2\gamma}{r} z\right) = a^{\infty}\left(v - p - \frac{2\gamma}{r} z\right).$$

and where the limiting inventory position of trader i at time t is

$$z_{it}^{\infty} = \frac{Z}{n} + e^{-(n-2)rt/2}\left(z_{i0} - \frac{Z}{n}\right). \tag{15}$$

The equilibrium price at time t is

$$p^* = v - \frac{2\gamma}{nr}Z. \tag{16}$$

2. *The limiting expected net payoff of trader i at time 0, conditional on z_{i0} and the initial auction price p^*, is*

$$V_{i,0+}^{\infty} = v\frac{Z}{n} - \frac{\gamma}{r}\left(\frac{Z}{n}\right)^2 + \left(v - \frac{2\gamma}{r}\frac{Z}{n}\right)\left(z_{i0} - \frac{Z}{n}\right) - \frac{\gamma}{r(n-1)}\left(z_{i0} - \frac{Z}{n}\right)^2. \tag{17}$$

From (15), we see that a continually operating market causes allocations to converge to their efficient levels exponentially at the rate $r(n-2)/2$, which is increasing in the number of traders and decreasing in the mean arrival rate of the asset payoff time. This leaves significant scope for improvements in allocative efficiency in settings, such as inter-dealer bond trading platforms, in which trading activity is highly concentrated among a small number of major dealers and large institutional investors.

3 Introducing Workup for Size Discovery

We saw in the previous section that successive rounds of double auctions move the inventories of the traders toward a common level. This reduction in inventory dispersion is only gradual, however, because at each round, each trader internalizes the price impact caused by his own quantity demands, and thus "shades" his demand schedule so as to trade off inventory holding costs against price impact.

We now examine the effect of introducing at time 0 a size-discovery mechanism, taken for concreteness to be a workup session, that gives traders the opportunity to reduce the magnitudes of their excess inventories without concern over price impact. It would be natural in practice to run a workup session whenever traders' inventories have been significantly disrupted. In the U.S. Treasury market, for example, primary dealers' positions can be suddenly pushed out of balance by unexpectedly large or small awards in a Treasury auction. Individual dealers' inventories could also be disrupted by large surges of demand or supply from their buy-side clients. We show that workup immediately reallocates a potentially large amount of inventory imbalances, which improves allocative efficiency relative to the double-auction market without a workup.

3.1 A model of bilateral workup

For expositional simplicity, we first 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, "the buyer," and one with positive inventory, "the seller." Any trader not participating in one of the bilateral workup sessions is active only in the subsequent double-auction market. Information held by a pair of workup participants regarding participation in other workup sessions plays no role in our model. That is, the equilibrium for the bilateral workup sessions and the subsequent double auctions is unaffected by information held by the participants in a given workup regarding how many other workup sessions are held and which traders are participating in them. For simplicity, we do not model the endogenous matching of workup partners.

In the next section, we generalize the model to cover a more realistic multilateral workup session.

In a given bilateral workup session, say that between traders 1 and 2, the seller has an initial inventory S^s that is exponentially distributed with parameter μ, thus having a mean of $1/\mu$. The initial inventory S^b of the buyer is negative, with a magnitude $|S^b|$ that independent of, and identically distributed with, the seller's initial inventory. The sum of the initial inventories $z_{30}, \ldots, z_{n0}$ of the other $n-2$ traders has mean zero and is independent of S^s and S^b. There is no need to further specify the joint probability distribution of other traders' inventories, because the equilibrium of the subsequent double auctions that is shown in Section 2 applies for any such joint distribution.

The workup price $\bar{p}$ is set without the use of information about traders' privately observed inventories, and therefore at some deterministic level $\bar{p}$. We will provide an interval of choices for $\bar{p}$ that is necessary and sufficient for interior equilibrium workup dropout policies. We will also show that the allocative efficiency improvement of workup is invariant to changes in the workup price $\bar{p}$ within this interval. A natural choice for $\bar{p}$ is the unconditional expectation of the asset payoff v, which can be interpreted as the expectation of the clearing price in the subsequent double-auction market, or as the price achieved in a previous round of auction-based trade, before new inventory shocks instigate a desire by traders to lay off their new unwanted inventories.

After each of a given pair of participants in a workup privately observes his own inventory, the workup proceeds in steps as follows:

1. The workup operator announces the workup price $\bar{p}$.
2. The workup operator provides a continual display, observable to buyer and seller, of the quantity $Q(t)$ of the asset that has been exchanged in the workup by time t on the workup "clock." The units of time on the workup clock are arbitrary, and the function $Q(\,\cdot\,)$ is any strictly increasing continuous function satisfying $Q(0) = 0$ and $\lim_{t\to\infty} Q(t) = \infty$. For example, we can take $Q(t) = t$. The workup clock can

run arbitrarily quickly, so workup can take essentially no time to complete.

3. At any finite time T_b on the workup clock, or equivalently at any quantity $\mathcal{Q}_b = Q(T_b)$, the buyer can drop out of the workup. Likewise, the seller can drop out at any time T_s or quantity $\mathcal{Q}_s = Q(T_s)$. The workup stops at time $T^* = \min(T_s, T_b)$, at which the quantity $Q^* = Q(T^*) = \min(\mathcal{Q}_b, \mathcal{Q}_s)$ is transferred from seller to buyer at the workup price $\bar{p}$, that is, for the total consideration $\bar{p}\,Q^*$.

After the bilateral workups terminate, all traders enter the sequence of double auctions described in Section 2.

As mentioned in the introduction, the workup procedure modeled here is similar to the matching mechanism used by certain dark pools, such as POSIT and Liquidnet, that specialize in executing large equity orders from institutional investors. In a dark-pool transaction with one buyer and one seller, each side privately submits a desired trade size to the dark pool, understanding that the dark pool would execute a trade for the minimum of the buyer's and seller's desired quantities. In a bilateral setting, workup and dark-pool matching are thus equivalent.

Among the institutional differences between these size-discovery trade mechanisms, workups and dark pools, two are worth mentioning here. First, workups are often integrated with a price-discovery mechanism on the same trading platform. Once the price is set, it is frozen until the quantity workup is completed. Dark pools are often operated by broker-dealers away from "lit" exchanges which determine the prices used by the dark pool. Second, workups are used primarily for U.S Treasuries and swaps, and thus for transactions of large (institutional) size. Dark pools are used primarily for equity markets having substantial retail participation. The average transaction size in dark pools, except for a few block-size dark pools, is similar to that of exchanges.

3.2 Characterizing the workup equilibrium

This section characterizes the equilibrium behavior of the two traders in a given bilateral workup session.

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 Q^* plays no role in the subsequent double-auction analysis.

We conjecture the following equilibrium workup strategies. The buyer and seller fix deterministic thresholds, M_b and M_s respectively, for the magnitudes of residual inventory at which they drop out of workup. That is, the buyer allows the workup transaction size to increase until the time T_b at which his residual inventory $|S^b + Q(T_b)|$ is equal to M_b. The seller likewise chooses a dropout time T_s at which his residual inventory $S^s - Q(T_s)$ reaches M_s. One trader's dropout is of course pre-empted by the other's.

The dropout strategies. For any $y > 0$, let F_y be the event that the buyer's candidate requested quantity $y > 0$ is filled. That is,

$$F_y = \left\{ 0 \le -(S^b + y) - M_b \le S^s - y - M_s \right\}. \tag{18}$$

The remaining inventory of the buyer, $-(S^b + y)$, is weakly larger than the dropout quantity M_b, for otherwise the buyer would have already dropped out. On the event F_y, the buyer's remaining inventory in excess of M_b is also weakly smaller than the remaining inventory of the seller in excess of M_s, for otherwise the seller would have already dropped out.

With the conjectured equilibrium dropout strategies, the memoryless property of the exponential distribution implies that, for the buyer, the seller's inventory in excess of the dropout quantity, which is $W \equiv S^s - y - M_s$, is F_y-conditionally exponential with the same parameter μ. Thus, recalling that Z is the aggregate inventory of the traders, we have

$$E(Z \mid F_y, S^b) = S^b + y + M_s + \frac{1}{\mu}, \tag{19}$$

using the fact that the expected total inventory of all traders not participating in this workup is zero.

By a similar calculation,

$$\begin{aligned} E(Z^2 \mid F_y, S^b) &= E\left[(S^b + y + M_s + W)^2 + \left(\sum_{i=3}^{n} z_{i0} \right)^2 \right] \\ &= (S^b + y + M_s)^2 + E(W^2) + 2(S^b + y + M_s)E(W) + \theta \\ &= (S^b + y + M_s)^2 + \frac{2}{\mu^2} + 2(S^b + y + M_s)\frac{1}{\mu} + \theta, \end{aligned} \tag{20}$$

where

$$\theta = E\left[\left(\sum_{i=3}^{n} z_{i0} \right)^2 \right].$$

On the other hand, given the initial inventory S^b and the candidate quantity $y \ge 0$ to be acquired in the workup, the buyer's conditional expected ultimate value, given $\{Z, S^b\}$, is

$$U^b = -\bar{p}y + \mathcal{V}(S^b + y), \tag{21}$$

where, based on Proposition 2,

$$\mathcal{V}(z) = v\frac{Z}{n} - \frac{\gamma}{r}\left(\frac{Z}{n}\right)^2 + \left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right)\left(z - \frac{Z}{n}\right) - \frac{\gamma}{r}\frac{1 - 2a_\Delta\gamma/r}{n-1}\left(z - \frac{Z}{n}\right)^2. \tag{22}$$

Organizing the terms, we get

$$\begin{aligned} E(U^b \mid F_y, S^b) = &-\bar{p}y + v(S^b + y) - \frac{\gamma}{r}C(S^b + y)^2 + 2\frac{\gamma}{r}(C-1)(S^b + y)\frac{E(Z \mid F_y, S^b)}{n} \\ &- \frac{\gamma}{r}(C-1)\frac{E(Z^2 \mid F_y, S^b)}{n^2}, \end{aligned} \tag{23}$$

where

$$C = \frac{1 - 2a_\Delta\gamma/r}{n-1}. \tag{24}$$

Substituting the expressions that we have shown above for $E(Z \mid F_y, S^b)$ and $E(Z^2 \mid F_y, S^b)$ into this expression for $E(U^b \mid F_y)$, we get

$$\begin{aligned} g(y) \equiv \frac{dE(U^b \mid F_y, S^b)}{dy} = &\; v - \bar{p} - 2\frac{\gamma}{r}C(S^b + y) + 2\frac{\gamma}{r}(C-1)\frac{1}{n}\left(2(S^b + y) + M_s + \frac{1}{\mu}\right) \\ &- \frac{\gamma}{r}(C-1)\frac{1}{n^2}\left(2(S^b + y + M_s) + \frac{2}{\mu}\right). \end{aligned} \tag{25}$$

The derivative $g(y)$ is everywhere strictly decreasing in y. Following the conjectured equilibrium, an optimal dropout quantity M_b for the buyer's residual inventory, if the optimum is interior (which we assume for now and then validate), is obtained at a level of y for which this derivative $g(y)$ is equal to zero, and by taking $S^b + y = -M_b$. That is,

$$\begin{aligned} 0 = &\; v - \bar{p} - 2\frac{\gamma}{r}C(-M_b) + 2\frac{\gamma}{r}(C-1)\frac{1}{n}\left(2(-M_b) + M_s + \frac{1}{\mu}\right) \\ &- \frac{\gamma}{r}(C-1)\frac{1}{n^2}\left(2(-M_b + M_s) + \frac{2}{\mu}\right). \end{aligned} \tag{26}$$

By completely analogous reasoning, the first-order conditions for the seller's optimal dropout threshold M_s is given by

$$\begin{aligned} 0 = &\; \bar{p} - v + 2\frac{\gamma}{r}CM_s + 2\frac{\gamma}{r}(C-1)\frac{1}{n}\left(-2M_s + M_b + \frac{1}{\mu}\right) \\ &- \frac{\gamma}{r}(C-1)\frac{1}{n^2}\left(-2(M_s - M_b) + \frac{2}{\mu}\right). \end{aligned} \tag{27}$$

The Equilibrium Dropout Thresholds. From the two equations (26) and (27), we can calculate that

$$M_b = \frac{n-1}{n + n^2 C/(1-C)} \frac{1}{\mu} + \delta = M + \delta, \tag{28}$$

$$M_s = \frac{n-1}{n + n^2 C/(1-C)} \frac{1}{\mu} - \delta = M - \delta, \tag{29}$$

where

$$M \equiv \frac{n-1}{n + n^2 C/(1-C)} \frac{1}{\mu}. \tag{30}$$

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

$$\delta = \frac{r}{2\gamma} \frac{\bar{p} - v}{C + (1-C)(3n-2)/n^2}. \tag{31}$$

As we see later, the symmetric solution M plays an important role in calculating the comparative statics. To be consistent with the initial conjecture that the equilibrium is interior, that is, $M_b > 0$ and $M_s > 0$, the workup price $\bar{p}$ must have the property that $|\delta| \leq M$, or equivalently,

$$|\bar{p} - v| \leq \frac{2\gamma M[C + (1-C)(3n-2)/n^2]}{r}. \tag{32}$$

We will only treat prices in satisfying this interior-dropout condition.

It is intuitive that a biased workup price causes asymmetric dropout behavior. If $\bar{p} > v$, the workup price is less favorable than the double auction price for the buyer, but more favorable for the seller. Thus, the buyer is more cautious than the seller in the workup, in that the buyer's dropout level is higher than the seller's. The opposite is true if $\bar{p} < v$.

This result is summarized by the following proposition.

Proposition 3. *Suppose that the workup price $\bar{p}$ satisfies (32). The workup session has a unique equilibrium in dropout inventory strategies. The buyer's and seller's dropout levels, M_b and M_s, for residual inventory are given by (28) and (29), respectively. 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.*

Figure 2 illustrates the impact of the workup on the undesired inventory levels of two traders. In this simple example, there are $n = 5$ traders and one bilateral workup session. The workup price is $\bar{p} = v$. The two workup participants have mean inventory size $1/\mu = 1$. We calculate the equilibrium outcome of the workup when the outcome of the workup buyer's pre-workup inventory S^b is -2, the outcome of the workup seller's pre-workup inventory S^s is 1.5, and the outcomes of all of the other traders' initial inventories are zero. The outcome for the efficient allocation of all traders is $Z/n = -0.1$. We focus on the continuous-time limit of the sequential-double-auction market. The equilibrium workup dropout quantity is in this case $M = 0.3$. Because we have the outcome that $|S^b| > |S^s|$, the seller exits the workup first, after executing the quantity $1.5 - 0.3 = 1.2$. The seller's inventory after the workup is 0.3, whereas the buyer's inventory after the workup is $-(2 - 1.2) = -0.8$.

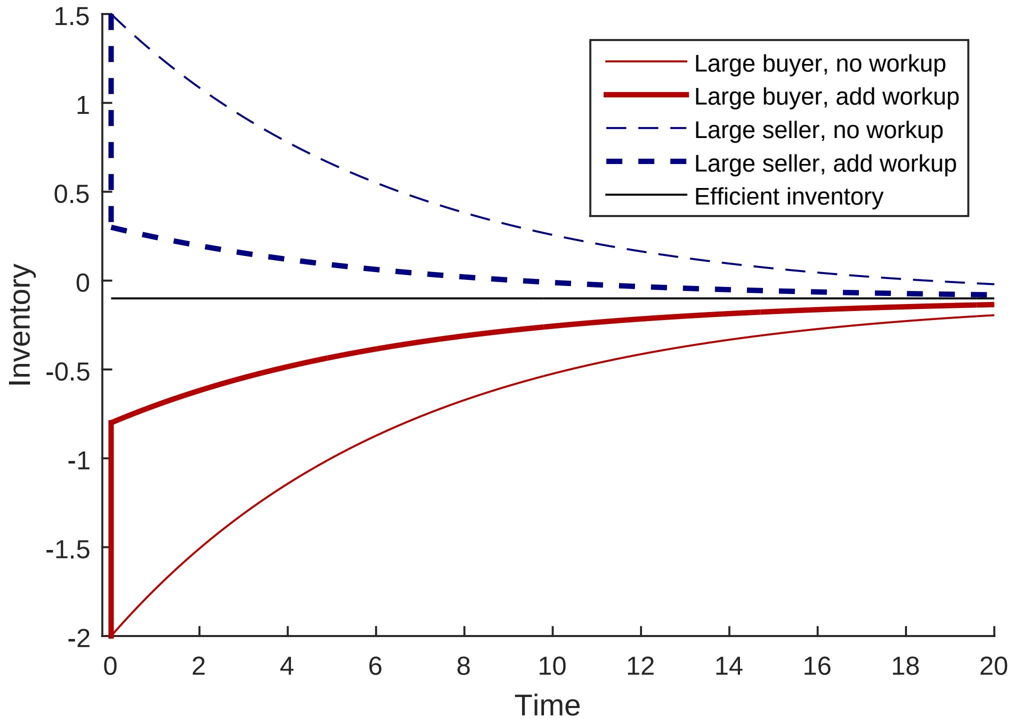

Figure 2: *Immediate inventory imbalance reduction by workup.* Parameters: $n = 5$, $\mu = 1$, $r = 0.1$, $\gamma = 0.05$, $\Delta = 0$, $S^b = -1.5$, $S^s = 2$. The outcomes of the inventories of traders not entering workup are zero.

Proposition 3 shows that as long as the workup price is not too biased, the two workup participants do not generally attempt to liquidate all of their inventories during the workup (in that $M_b > 0$ and $M_s > 0$). Their optimal target inventories are determined by two countervailing incentives. On one hand, 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. On the other hand, a trader faces adverse selection regarding the total inventory Z and the double-auction prices. For example, if the buyer expects that the auction price p^* will be lower than the workup price $\bar{p}$, the buyer would be better off buying some of the asset in the subsequent auction market, despite the associated price impact. This incentive encourages inefficient "self-rationing" in the workup. A symmetric argument holds for the seller. Depending on a trader's conditional expectation of the total market excess inventory Z, which changes as the workup progresses, the trader sets an endogenous dropout inventory threshold such that the two incentives are optimally balanced. In setting his optimal target inventory, a trader does not attempt to strategically manipulate other traders' inference of the total inventory Z, because optimal auction strategies do not depend on conditional beliefs about Z.

3.3 Properties of the workup equilibrium

We now describe additional properties of the workup equilibrium of Proposition 3.

Efficiency. Although adding workups does not lead to the efficient allocation, it causes a sudden and beneficial reduction in inventory imbalances. Because workup participation is voluntary, traders can only increase their utilities (at least weakly) by participating in workups. Traders suffer no loss in expected net benefit (relative to a market without workup) from not participating in workups, as can be checked from (17). Thus, adding workup sessions is a Pareto improvement among all traders, and offers a strict improvement to those who, in expectation, execute a strictly positive quantity in a workup. There is generally a non-trivial welfare improvement associated with adding workups, measured by the increase in the sum of the total of trades' ex-ante expected net payoff.

Proposition 4. *Adding workup sessions before the sequential-double-auction market is a Pareto improvement. That is, every trader's ex-post utility is increased, and the ex-post utility of participating traders who, in expectation, execute strictly positive workup quantities is strictly increased, over that associated with a sequential-double-auction market that is not preceded by workup.*

Robustness of market outcomes to the choice of workup prices. Provided that the workup price $\bar{p}$ is in the interval satisfying condition (32), the probability of active workup participation, the expected trading volume, and total welfare improvement are invariant to variation in $\bar{p}$.

For instance, the probability of triggering an active workup is

$$P(S^s > M_s, |S^b| > M_b) = e^{-\mu(M+\delta)} e^{-\mu(M-\delta)} = e^{-2\mu M}, \tag{33}$$

which is decreasing in M and does not depend on $\bar{p}$, within the range of interior solutions.

Moreover, by substituting in (30), we obtain:

$$P(S^s > M_s, |S^b| > M_b) = \exp\left(-\frac{2(n-1)}{n + n^2 C/(1-C)}\right). \tag{34}$$

That is, the probability of having an active workup does not depend on the average inventory sizes in the market. This probability of active workup depends instead on the competitiveness of the double-auction market (which is captured by the number n of traders), the mean arrival rate r of price-relevant information, and the auction-market frequency $(1/\Delta)$. We will discuss these comparative statics in detail shortly.

Similarly, the expected trading volume in the workup is

$$\begin{aligned} &E\left[\max\left(\min\left(|S^b| - M_b, S^s - M_s\right), 0\right)\right] \\ &\quad = \int_{x=M+\delta}^{\infty} \int_{y=M-\delta}^{\infty} \mu e^{-\mu x} \mu e^{-\mu y} \min(x - (M+\delta), y - (M-\delta))\, dx\, dy. \end{aligned} \tag{35}$$

By the change in variables $u = x - M - \delta$ and $w = y - M + \delta$, the integral can be re-expressed as

$$\begin{aligned} &\int_{u=0}^{\infty} \int_{w=0}^{\infty} \mu e^{-\mu(u+M+\delta)} \mu e^{-\mu(w+M-\delta)} \min(u, w)\, du\, dw \\ &\quad = e^{-2\mu M} \int_{u=0}^{\infty} \int_{w=0}^{\infty} \mu e^{-\mu u} \mu e^{-\mu w} \min(u, w)\, du\, dw, \end{aligned} \tag{36}$$

which is decreasing in M and is invariant to δ in the interval $[0, M]$. Thus, within the range of workup prices at which dropout thresholds are interior, the expected workup trade volume does not depend on the workup price $\bar{p}$.

The robustness of the total welfare improvement achieved by workup to variation with respect to the workup price $\bar{p}$ follows from a similar calculation shown in Appendix C.

Comparative statics and empirical predictions. For fixed inventory-size parameter μ, comparative statics reveal how the workup inventory dropout threshold M varies with respect to Δ, r, and n, and reveal how the attractiveness of the size-discovery workup mechanism varies with market conditions. We have just shown that the probability of triggering an active workup is decreasing in M.

We have

$$\zeta(\Delta) \equiv 1 - 2a_\Delta \frac{\gamma}{r} = \frac{\sqrt{(n-1)^2(1-e^{-r\Delta})^2 + 4e^{-r\Delta}} - (n-1)(1-e^{-r\Delta})}{2e^{-r\Delta}}. \tag{37}$$

By Proposition 2, $\zeta(\Delta)$ is the fraction of excess inventory that remains after each successive double auction. The smaller is this quantity, the more aggressive are traders' submitted demand schedules. The constant γ that scales the quadratic inventory cost does not in itself affect $\zeta(\Delta)$ or M. This is perhaps surprising, and applies in light of the fact that the aggressiveness of demand schedules fully offsets the effect of γ, causing $a_\Delta \gamma$ to be invariant to γ.

By calculation,

$$\zeta'(\Delta) = \frac{re^{r\Delta}(n-1)}{2}\left(\frac{\sqrt{(n-1)^2(e^{r\Delta}-1)^2+4e^{r\Delta}-4\left(1-\frac{1}{(n-1)^2}\right)}}{\sqrt{(n-1)^2(e^{r\Delta}-1)^2+4e^{r\Delta}}}-1\right) < 0. \quad (38)$$

Because M is decreasing in $C = \zeta(\Delta)/(n-1)$, M is therefore increasing in Δ. That is, the smaller is Δ (the more frequent the double auctions), the smaller is M, and the more active is an active workup. Intuitively, a small Δ discourages aggressive trading in the double auctions because traders have frequent opportunities to trade, which implies that the convergence to efficient inventory levels is slow. (This welfare cost of fast trading is also discussed by Vayanos (1999) and Du and Zhu (2014).) Workup is therefore more attractive to "large traders," that is, those tending to have large inventory imbalances, because workup can lead to more dramatic reductions in excess inventories, despite the adverse-selection effect of workup.

As the double auctions become more frequent, that is as Δ goes to zero, we know that $a_\Delta \to 0$ and thus $C \to 1/(n-1)$. In this case, M converges downward to the continual-auction limit

$$\frac{n-2}{2n}\frac{1}{\mu}, \quad (39)$$

and the probability of triggering a workup becomes maximal, at $\exp(-(n-2)/n)$. At this continuous-time limit, which is the same as the behavior of the corresponding continuous-auction model shown in Appendix B, the probability of triggering a workup decreases in n. Intuitively, the double auction market becomes more efficient as the number n of participants grows, getting closer and closer to price-taking competitive behavior. Hence, as n grows, there is less allocative benefit from size discovery. In fact, we can show that M increases with n regardless of the model parameters. (For details, see Appendix A.4.)

For example, in a market with $n = 20$ traders, if workup is preceded by a continuous-time auction market, the probability of active workup is $e^{-18/20} \approx 0.41$. With only $n = 5$ traders, this active-workup probability increases to $e^{-3/5} \approx 0.55$.

We also have

$$\frac{d}{dr}\left(1 - 2a_\Delta\frac{\gamma}{r}\right) < 0. \quad (40)$$

That is, the lower is the mean arrival rate of the asset payoff information, the smaller is M, and the more likely it is that an active workup is triggered. Intuitively, the more delayed is the final determination of asset payoffs, the less aggressive are traders in their double-auction demand schedules, which in turn increases the attractiveness of using workup to quickly reduce inventory imbalances.

These comparative statics are summarized by the following proposition.

Proposition 5. *An increase in the probability of triggering a workup and an increase in the expected workup trade volume (that is, a reduction in M) is caused by any of the following:*

1. *An increase in the frequency of subsequent double auctions (a reduction in Δ).*
2. *A reduction in the mean arrival rate of the asset payoff date (a reduction in r).*
3. *A reduction in the number n of traders.*

Menkveld, Yueshen, and Zhu (2015) find that the market share of dark pools in U.S. equity markets drops significantly before macroeconomic data releases, and that this drop is larger for midpoint dark pools, which offer the least price discovery. As we have discussed, midpoint dark pools are similar to workups in that they offer size discovery, but not price discovery. The evidence from this empirical work is consistent with the implication of Proposition 5 that more imminent news releases cause investors to be more eager to trade on exchanges, which offer price discovery and clear the market, and less anxious to shrink excess inventories in dark pools. We are not aware of empirical work that is directly related to Parts 1 and 3 of Proposition 5.

4 Multilateral Workups

In Section 3 we solved the equilibrium for bilateral workup sessions, and showed that workup provides size-discovery welfare benefits. This section extends our results to dynamic multilateral workups, which are more commonly used in practice, for example on electronic trading-platform markets. The intuition for the allocative efficiency benefits of size-discovery is similar to that for the simpler case of bilateral workup. There may be additional efficiency gains, however, from placing more participants into a single workup session. Moreover, additional insights are gained from the equilibrium dynamic dropout policies in multilateral workups.

Whenever a trader drops out of a workup, another trader can immediately step in on the same side and execute additional quantity at the same fixed price. This process continues until no one is willing to execute additional quantity at the same price. This dynamic workup model closely resembles the workup protocol used on the BrokerTec platform for U.S. Treasuries, which is reviewed by Fleming and Nguyen (2013). One may also view our dynamic workup model as an approximation of a block-trading dark pool that accepts orders sequentially and allocates matched trades using time priority.

The eventual numbers N_b of buyers and N_s of sellers to arrive for workup are independent and have the same geometric distribution. That is,

$$P(N_b = k) = P(N_s = k) = f(k) \equiv q^k(1-q), \quad k \in \{0, 1, 2, ...\}. \tag{41}$$

for some $q \in (0, 1)$. It is natural that the number of institutional investors and financial intermediaries seeking to trade large positions is unobservable and stochastic. We have $E(N_b) = E(N_s) = q/(1-q)$. The interpretation is that after each buyer exits the workup, there is a new buyer with probability q, and likewise for sellers.

Pre-workup inventories are positive for sellers and negative for buyers. For both buyers and sellers, the absolute magnitudes of pre-workup inventory sizes are *iid* exponentially distributed, with parameter μ, thus with mean $1/\mu$. The numbers of buyers and sellers and the pre-workup inventory sizes are independent. Before participating in workup, each trader observes only his own inventory.

It follows from the independence assumptions and the memoryless property of the geometric distribution that, conditioning on all information available to a trader during his turn at workup, the conditional distribution of the numbers of buyers and sellers that have not yet entered workup retain their original independent geometric distributions.

As in Section 3, the workup session takes place before the start of the double-auction market. 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 exit of a trader, whether a buyer or a seller, and the replacement of the trader is observable to everyone when it occurs. (The identities of the exiting traders are irrelevant, and not reported, beyond whether they are buyers or sellers.) The quantities executed by each departing trader are also observable. In particular, the event that a trader drops out of workup without executing any quantity is also observable. The workup ends when buyer number N_b exits or when seller number N_s exits, whichever is first.

Throughout this section, we assume for simplicity that the workup price $\bar{p}$ is equal to expected asset payoff v.

At any given point in the workup, we let:

- (i, j) be the current (buyer, seller) workup pair. That is, i is the queue position of the buyer and j is the queue position of the seller.
- $I(i, j)$ be the event that the (i, j) pair have just met and have not yet executed any quantity yet. (The "I" means "inactive.")
- $A(i, j, y)$ be the event that the (i, j) pair have already executed some quantity $y > 0$. (The "A" means "active.")
- X be the total conditional expected inventory of previously exited participants. Given our information structure, this conditional expectation is common to all workup participants.

n_b, n_s be the number of buyers and the number of sellers yet to enter the workup, respectively, excluding the current pair (i, j).

We let $\mathcal{M}_b(i, j, X) > 0$ and $\mathcal{M}_s(i, j, X) > 0$ denote the conjectured dropout thresholds of the i-th buyer and j-th seller in an *active* (i, j) workup state. That is, we conjecture that, when active, buyer i drops out once the absolute magnitude of his remaining inventory has been reduced to $\mathcal{M}_b(i, j, X)$. We conjecture and later verify an equilibrium in which these thresholds depend only on (i, j, X), and not on a trader's current inventory or on other aspects of the observable history of the game.

The distinction between an active workup pair (i, j) and a matched but currently inactive pair is important to the equilibrium policies. Suppose, for example, that we are in the $A(1, 1, y)$ state. That is, the first buyer and the first seller have executed a positive quantity y in the workup, and nothing else has yet happened. As we will show later, because $X = 0$, the buyer and the seller use a common dropout threshold, say $\mathcal{M}_0$. If, for example, the buyer exits, then every workup participant infers that the buyer's residual inventory level is $-\mathcal{M}_0$. By contrast, on the inactive event $I(1, 1)$, if the buyer immediately exits, then everyone else learns that the buyer's inventory size is *at most* $\mathcal{M}_0$, and in particular is distributed with a truncated exponential distribution, with the conditional expectation $\nu(\mathcal{M}_0)$, where, for any positive number y,

$$\nu(y) \equiv \frac{\int_{x=0}^{y} x\mu e^{-\mu x}\, dx}{1 - e^{y}} < y. \tag{42}$$

Thus, whether a trader exits without trading a strictly positive amount affects the inference of all traders.

The key step to solving the equilibrium is to calculate the dropout thresholds $\mathcal{M}_b(i, j, X)$ and $\mathcal{M}_s(i, j, X)$. We now compute these thresholds. For notational simplicity, we write

$$m \equiv i + j \tag{43}$$

for the total number of traders who have participated in workup, including the current pair.

First, we consider the problem of the current active buyer, whose initial inventory is S^b, on the event that the buyer has to this point executed some quantity $y > 0$ and the active seller has not yet exited. We will calculate the first-order optimality conditions by artificially including the residual queue sizes n_b and n_s in the buyer's conditioning information, and then later averaging with respect to the conditional distribution of (n_b, n_s). By the same logic used in Section 3, the buyer has conditional mean and variance[8] of aggregate market-wide inventory given by

$$E(Z \mid S^b, A(i,j,y), X, n_b, n_s) = S^b + y + \mathcal{M}_s(i,j,X) + \frac{1}{\mu} + X + (n_s - n_b)\frac{1}{\mu}, \tag{44}$$

$$\begin{aligned} E(Z^2 \mid S^b, A(i,j,y), X, n_b, n_s) &= (S^b + y + \mathcal{M}_s(i,j,X))^2 + 2(S^b + y + \mathcal{M}_s(i,j,X))\frac{1}{\mu} \\ &\quad + \Gamma_b(n_b, n_s), \end{aligned} \tag{45}$$

where $\Gamma_b(n_b, n_s)$ is a quantity that does not depend on y. Relative to the calculation (19) for the case of bilateral workup, the conditional mean $E(Z \mid A(i,j,y), X, n_b, n_s)$ includes the extra terms X and $(n_s - n_b)/\mu$. The exact level of the second moment $E(Z^2 \mid A(i,j,y), X, n_b, n_s)$ does not affect the equilibrium threshold, because it plays no role in the first-order optimality condition for the choice of y at which the buyer drops out.

Omitting the arguments of $\mathcal{M}_s$ and $\mathcal{M}_b$, we have

$$\begin{aligned} \frac{dE(U^b \mid S^b, A(i,j,y), X, n_b, n_s)}{dy} &= v - \bar{p} - 2\frac{\gamma}{r}C(n)(S^b + y) \\ &+ 2\frac{\gamma}{r}(C(n) - 1)\frac{1}{n}\left(2(S^b + y) + \mathcal{M}_s + \frac{1}{\mu} + X + (n_s - n_b)\frac{1}{\mu}\right) \\ &- \frac{\gamma}{r}(C(n) - 1)\frac{1}{n^2}\left(2(S^b + y + \mathcal{M}_s) + \frac{2}{\mu}\right), \end{aligned} \tag{46}$$

where $n = m + n_b + n_s$ and where

$$C(n) = \frac{1 - 2a_\Delta \gamma / r}{n - 1}. \tag{47}$$

By the law of iterated expectations, we can average with respect to the product distribution of (n_b, n_s), to obtain

$$\begin{aligned} \frac{dE(U^b \mid S^b, A(i,j,y), X)}{dy} &= v - \bar{p} - \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)2\frac{\gamma}{r}C(n)(S^b + y) \\ &+ \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)2\frac{\gamma}{r}(C(n) - 1)\frac{1}{n}\left(2(S^b + y) + \mathcal{M}_s + \frac{1}{\mu} + X + (\ell - k)\frac{1}{\mu}\right) \\ &- \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)\frac{\gamma}{r}(C(n) - 1)\frac{1}{n^2}\left(2(S^b + y + \mathcal{M}_s) + \frac{2}{\mu}\right), \end{aligned} \tag{48}$$

where $n = m + k + \ell$.

The first-order condition for optimal y should hold with equality if $S^b + y = -\mathcal{M}_b$,

[8]The event of executing y units has probability zero, but the stated conditional moments make sense when applying a regular version of the conditional distribution of Z given the executed quantity and given X.

that is,

$$
\begin{aligned}
0 =& \frac{dE(U^b \mid S^b, A(i,j,y), X)}{dy}\Big|_{S^b+y=-\mathcal{M}_b} \\
=& v - \bar{p} - \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)2\frac{\gamma}{r}C(n)(-\mathcal{M}_b) \\
&+ \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)2\frac{\gamma}{r}(C(n)-1)\frac{1}{n}\left(2(-\mathcal{M}_b)+\mathcal{M}_s+\frac{1}{\mu}+X+(\ell-k)\frac{1}{\mu}\right) \\
&- \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)\frac{\gamma}{r}(C(n)-1)\frac{1}{n^2}\left(2(-\mathcal{M}_b+\mathcal{M}_s)+\frac{2}{\mu}\right),
\end{aligned}
\tag{49}
$$

where $n = m + k + \ell$.

By a completely analogous calculation, the seller, whose initial inventory is S^s, stays in workup until the buyer has exited or the workup quantity has reached a level y satisfying the seller's first-order condition, whichever comes first. This occurs when the seller's remaining inventory reaches the threshold $S^s - y = \mathcal{M}_s$. Thus, the first-order condition for y takes the form

$$
\begin{aligned}
0 =& \frac{dE(U^s \mid S^s, A(i,j,y), X)}{dy}\Big|_{S^s-y=\mathcal{M}_s} \qquad (50)\\
=& \bar{p} - v + \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)2\frac{\gamma}{r}C(n)\mathcal{M}_s \\
&+ \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)2\frac{\gamma}{r}(C(n)-1)\frac{1}{n}\left(-2\mathcal{M}_s+\mathcal{M}_b+\frac{1}{\mu}-X-(\ell-k)\frac{1}{\mu}\right) \\
&- \sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)\frac{\gamma}{r}(C(n)-1)\frac{1}{n^2}\left(-2(\mathcal{M}_s-\mathcal{M}_b)+\frac{2}{\mu}\right),
\end{aligned}
$$

where $n = m + k + \ell$.

Substituting in $\bar{p} = v$ and solving the pair of linear equations, (49) and (50) for $\mathcal{M}_s$ and $\mathcal{M}_b$, we get

$$
\mathcal{M}_b(i,j,X) = M^*(i+j) + L(i+j)X_{i,j}, \tag{51}
$$

$$
\mathcal{M}_s(i,j,X) = M^*(i+j) - L(i+j)X_{i,j}, \tag{52}
$$

where, recalling that $m = i + j$ and letting $n = m + k + \ell$,

$$
M^*(m) = \frac{1}{\mu}\;\frac{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)(1-C(n))\frac{n-1}{n^2}}{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)\left(C(n)+\frac{1-C(n)}{n}\right)}
$$

and

$$L(m) = \frac{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)\frac{1-C(n)}{n}}{\sum_{k=0}^{\infty}\sum_{\ell=0}^{\infty} f(k)f(\ell)\left(C(n)+\frac{(1-C(n))(3n-2)}{n^2}\right)}.$$

The asymmetry between the buyer and the seller is intuitive. In a multilateral workup, the role of the conditional expected total inventory X of those traders who have already exited workup is similar to the role of the workup price "bias" $\bar{p}-v$ in the bilateral workup equilibrium described by Proposition 3. For example, an increase in X makes the current buyer more cautious, setting a higher dropout quantity, and makes the current seller more aggressive, setting a lower dropout quantity. This is so because as X rises, the conditional expected market-clearing price of the subsequent double auctions falls. This encourages the current buyer to wait for the double-auction market to cover his inventory shortfall, and encourages the seller to reserve less inventory for sale in the double-auction market. The opposite is true for a decrease in X.

In order for the above conjectured strategies to be consistent, we need to prove that the thresholds of incumbents are weakly increasing with each dropout, and that the thresholds are always nonnegative. That is, we want to show that

$$\mathcal{M}_b(i, j+1, X_{i,j+1}) \geq \mathcal{M}_b(i, j, X_{i,j}), \tag{53}$$

$$\mathcal{M}_s(i+1, j, X_{i+1,j}) \geq \mathcal{M}_s(i, j, X_{i,j}), \tag{54}$$

$$\mathcal{M}_b(i, j, X_{i,j}) \geq 0, \tag{55}$$

$$\mathcal{M}_s(i, j, X_{i,j}) \geq 0. \tag{56}$$

If these conditions fail, a trader's optimal dropout threshold may depend on his current inventory or the past threshold of his counterparty, among possibly other variables. These complications would render the problem intractable.

The monotonicity and positivity properties of (53)–(56) are satisfied if $e^{-r\Delta} > 1/2$, a relatively unrestrictive condition. For example, taking a day as the unit of time, if payoff-relevant information arrives once per day ($r = 1$) and the double auctions are held at least twice per day ($\Delta \leq 0.5$), we would have $e^{-r\Delta} \geq e^{-0.5} \approx 0.61 > 0.5$, and (53)–(56) are satisfied.[9]

Proposition 6. *The coefficients $M^*(m)$ and $M^*(m)/L(m)$ are always weakly increasing in m for $m \geq 2$. If $e^{-r\Delta} > 1/2$, $L(m)$ is also weakly increasing in m for $m \geq 2$. If $e^{-r\Delta} > 1/2$, the monotonicity and nonnegativity conditions of (53)–(56) are satisfied.*

The property that $M^*(m)$ and $L(m)$ are increasing in m is intuitive. As more traders drop out (that is, as m increases), the expected total number of traders in the subse-

[9]We have also checked that if Δ is large enough, then $L(m)$ is not monotone increasing in m. Although the non-monotonicity of $L(m)$ for large Δ blocks our particular proof method when Δ is sufficiently large, it does not necessarily rule out other approaches to demonstrating equilibria in threshold strategies for large Δ.

quent double-auction market goes up, by the memoryless property of the geometric distribution. Since the double-auction market becomes more competitive as more traders participate, and the associated inefficiency related to price impact thus becomes smaller, there is less advantage to using workup, so $M^*(m)$ goes up.

In addition, traders who have already exited the workup will enter the double-auction market with their residual inventories, causing a predictable shift in the double-auction price relative to the workup price. For example, if the conditional expected inventory X of past workup participants is positive, then the double-auction price is expected to be lower than v, a favorable condition for the workup buyer. Again, because a larger number m of past and current workup participants makes the double-auction market more competitive in expectation, those who have already exited the workup will be more aggressive in liquidating their residual inventories, thus front-loading their sales in the relatively early rounds. The workup buyer, therefore, expects to purchase the asset in the double-auction market sooner and at more favorable prices. Consequently, the buyer will set an even higher dropout threshold. By a symmetric argument, conditional on $X > 0$, a higher m means that the seller sets an even lower threshold. That is, a higher m means a higher sensitivity of the thresholds $\mathcal{M}_b$ and $\mathcal{M}_s$ to the total expected inventory X of those who have already exited workup.

The monotonicity of the thresholds, (53)–(54), means that after the exit of a trader, his counterparty's dropout threshold (weakly) increases. For example, if the current seller j exits before the current buyer i, then X goes up, so the new threshold $M^*(m) + L(m)X$ of buyer i goes up. Likewise, after each exit of a buyer, X goes down, and the dropout threshold of the seller who remains in the workup increases. Thus, after a counterparty exit, the incumbent either drops out immediately because of his increased threshold, or he stays in despite his new higher threshold. Conditional on the latter event, for other traders, the incumbent's remaining inventory in excess of his new, increased threshold is again an exponentially-distributed variable with mean $1/\mu$. The non-negativity of the thresholds, (55)–(56), implies that no trader wishes to "overshoot" across the zero inventory boundary. These properties ensure stationarity and are in fact needed for tractability of this general approach to solving for equilibria.

Summarizing, the following proposition provides a complete description of the workup equilibrium.[10] The equilibrium workup strategy of each player depends on that player's

[10] Because of the continuum of agent types and actions, we cannot formally apply the standard notion of perfect Bayesian equilibrium for dynamic games with incomplete information, because that would call for conditioning on events that have zero probability, such as a counterparty dropping out of workup after executing a trade of a specific size. In our setting, actions are commonly observable and there is no issue concerning off-equilibrium-path conjectures, so almost any natural extension of simple perfect Bayesian equilibrium to our continuum action and type spaces leads to our equilibrium. For example, we could apply the notion of open sequential equilibrium of Myerson and Reny (2015). For the present draft of the paper, we refer simply to an "equilibrium" in the sense that every agent applies Bayes' Rule based on a regular version of the conditional distribution of Z given the observed variables, in order to compute its optimal threshold strategy, given the

privately observed pre-workup asset inventory and on the publicly observable Markov state process[11] (B, X), where an outcome of B is of the form $I(i, j)$ or $A(i, j, y)$, and where X is the conditional expectation of the total inventory of all traders that have previously exited the workup.

Proposition 7. *Suppose that $e^{-r\Delta} > 1/2$. The multilateral dynamic workup game associated with workup price $\bar{p} = v$ has a unique equilibrium in threshold dropout strategies. This equilibrium is characterized by the following recursive determination of the workup state and of traders' equilibrium dropout strategies. Here, z_i^b and z_j^s denote the pre-workup inventories of the i-th buyer and the j-th seller, respectively. The initial workup state is $(I(1, 1), 0)$.*

1. *At any inactive workup state $(I(i, j), X)$:*
   - *(a) If $|z_i^b| \leq \mathcal{M}_b(i, j, X)$ and $z_j^s > \mathcal{M}_s(i, j, X)$, where $\mathcal{M}_b(i, j, X)$ and $\mathcal{M}_s(i, j, X)$ are given by (51) and (52) respectively, then the buyer, and only the buyer, exits immediately (that is, without trading any quantity). Unless $i + 1 > N_b$, the workup state then evolves to $(I(i + 1, j), X - \nu(\mathcal{M}_b(i, j, X)))$.*
   - *(b) If $|z_i^b| > \mathcal{M}_b(i, j, X)$ and $z_j^s \leq \mathcal{M}_s(i, j, X)$, then the seller, and only the seller, exits immediately. Unless $j + 1 > N_s$, the workup state evolves to $(I(i, j + 1), X + \nu(\mathcal{M}_s(i, j, X)))$.*
   - *(c) If $|z_i^b| \leq \mathcal{M}_b(i, j, X)$ and $z_j^s \leq \mathcal{M}_s(i, j, X)$, then both sides exit immediately, without trading any quantity. Unless $i + i > N_b$ or $j + 1 > N_s$, the workup state evolves to $(I(i + 1, j + 1), X - \nu(\mathcal{M}_b(i, j, X)) + \nu(\mathcal{M}_s(i, j, X)))$.*
   - *(d) If $|z_i^b| > \mathcal{M}_b(i, j, X)$ and $z_j^s > \mathcal{M}_s(i, j, X)$, then buyer i and seller j enter an active quantity workup. That is, the workup state evolves to $(A(i, j, 0), X)$.*
   - *(e) If, at any of the transitions above, $i + 1 > N_b$ or $j + 1 > N_s$, then the workup ends.*
2. *At any active workup state $(A(i, j, y), X)$, the buyer i and seller j remain in the workup as their traded quantity rises until the earlier of the following two following events (a) and (b).*

threshold strategies of other agents. As stated, there is a unique such equilibrium in threshold strategies because: *(i)* given the other traders' threshold strategies, a given trader's threshold strategy is uniquely determined by its first order necessary condition for optimality (which is sufficient because of concavity), and *(ii)* there is a unique solution for the pair of first-order conditions for the equations for the threshold strategies $\mathcal{M}_b$ and $\mathcal{M}_s$.

[11] Specifically, $(B, X) = (B_t, X_t)_{(t \geq 0)}$ is a continuous-time Markov process with state space $\mathcal{B} \times \mathbb{R}$, where $\mathcal{B}$ is the union of $\mathbb{N} \times \mathbb{N} \times \mathbb{R}$ and $\mathbb{N} \times \mathbb{N}$, where $\mathbb{N}$ is the space of natural numbers. That is, at each time t on the workup clock, the outcome of X_t is in $\mathbb{R}$ and the outcome of B_t is either of the form (i, j, y) (an active state outcome) or of the form (i, j) (a passive state outcome). One can add an artificial independent exponential "wait time" after each transition to an inactive state in order to simplify the form of the transition distribution function for (B, X). This ensures that the state cannot jump twice at the same time on the workup clock.

(a) *The remaining inventory of the buyer (which is negative) rises to the threshold* $-\mathcal{M}_b(i,j,X) = -(M^*(m) + L(m)X)$*. At this point, the buyer exits. Unless* $i+1 > N_b$*, the workup state evolves to* $(I(i+1,j), X - (M^*(m) + L(m)X))$.

(b) *The remaining inventory of the seller falls to the threshold* $\mathcal{M}_s(i,j,X) = M^*(m) - L(m)X$*. At this point, the seller exits. Unless* $j+1 > N_s$*, the state evolves to* $(I(i,j+1), X + M^*(m) - L(m)X)$.

(c) *On the zero-probability event that (a) and (b) occur simultaneously, the state evolves to* $(I(i+1,j+1), X - 2L(m)X)$ *unless* $i+1 > N_b$ *or* $j+1 > N_s$.

(d) *If, at either or both of (a) or (b), we have* $i+1 > N_b$ *or* $j+1 > N_s$*, then the workup ends.*

The property that a trader's dropout threshold increases after a counterparty exits leads to the following empirical prediction, which is not offered by the bilateral model. All else equal, the probability that, of a currently active workup pair, the buyer exits before the seller, conditional on the event that the last workup pair completed their trade with the exit of the buyer, is smaller than when conditioning on the event that the last workup pair completed their trade with the exit of the seller. That is,

$$\begin{aligned} &P(\text{buyer exits first in current workup} \mid \mathcal{G}, \text{buyer exits first in last workup}) \\ <\, &P(\text{buyer exits first in current workup} \mid \mathcal{G}, \text{seller exits first in last workup}), \end{aligned} \tag{57}$$

where $\mathcal{G}$ denotes the other available conditioning information.

5 Concluding Remarks

Price-discovery markets are efficient in an idealized price-taking competitive market, for example one in which traders are infinitesimally small, as in Aumann (1964). The First Welfare Theorem of Arrow (1951), by which market clearing allocations are efficient, is based on the price-taking assumption. In many functioning markets, however, price taking is a poor approximation of trading behavior because of the awareness of price impact, and efficiency is lost. For instance, in inter-dealer financial markets, there are often heavy concentrations of inventory imbalances among a relatively small set of market participants. These are large dealers, hedge funds, and other asset management firms that are extremely conscious of their potential to harm themselves by price impact. Taking the case of U.S. Treasury markets, government auctions often leave a small number of primary dealers with significant position imbalances. Some dealers are surprised by being awarded substantially more bonds in the auction than needed to meet their customer commitments and desired market-making inventories. Some receive significantly less than desired. Fleming and Nguyen (2013) explain how dealers exploit workups to lay off their imbalances.

We have shown that under imperfect competition, adding a size-discovery mechanism, such as a workup, improves allocative efficiency over a stand-alone price-discovery mechanism, such as sequential double auctions. Precisely because the workup freezes the transaction price, it avoids the efficiency losses in price-discovery mechanisms that are associated with strategic avoidance of price impact. Workup participants are therefore willing to trade large blocks of an asset almost instantly, leading to a quick reduction of inventory imbalances and improvement in allocative efficiency. (Conversely, adding a price-discovery market to a size-discovery market also improves allocative efficiency.)

We have also shown that equilibrium optimal workup strategies trade off the benefit of quickly eliminating large undesired positions against the cost of adverse selection associated with information about the total market inventory. As a result, only traders with large inventory imbalances actively participate in workups, and workup participants set an endogenous threshold for the level of remaining inventory at which they drop out.

Our results support a broader sense that in settings of imperfect competition, market designs involving a mixture of heterogeneous trade mechanisms are more effective from an allocative viewpoint than markets with a single form of trading mechanism.

Appendix

A Proofs

This appendix contains proofs of results stated in the main text.

A.1 Proof of Proposition 1

We simplify the notation by writing "$x_{i,k}$" in place of "$x_{ik}(p_k; z_{ik})$," and conjecture an equilibrium strategy of the form

$$x_{i,k} = av - bp_k + dz_{i,k}. \tag{58}$$

Trader i in round k effectively selects the optimal execution price p_k^*. Adapting the method of Du and Zhu (2014), we write the first-order optimality condition of trader i as

$$\begin{aligned}(n-1)b\Bigg[\left(1 - e^{-r\Delta}\right) \left(v - \frac{2\gamma}{r}(x_{i,k} + z_{i,k}) + \sum_{j=1}^{\infty} e^{-rj\Delta}(1+d)^j (v - \frac{2\gamma}{r} E_k(z_{i,k+j} + x_{i,k+j})) \right) \\ - p_k^* - \sum_{j=1}^{\infty} e^{-rj\Delta}(1+d)^{j-1} d\, p_k^* \Bigg] - x_{i,k} = 0,\end{aligned} \tag{59}$$

where $E_k(\,\cdot\,)$ denotes conditional expectation given $z_{i,k}$ and p_k^*.

By the evolution equation for the inventory $\{z_{i,k}\}$, we have, for all $j \geq 1$,

$$z_{i,k+j} + x_{i,k+j} = (1+d)^j (z_{i,k} + x_{i,k}) + \sum_{l=1}^{j-1} (av - bp_{k+l})(1+d)^{j-l} + (av - bp_{k+j}). \tag{60}$$

Because prices are conjectured to be constant,

$$E_k(z_{i,k+j} + x_{i,k+j}) = (1+d)^j (z_{i,k} + x_{i,k}) + (av - bp_k^*) \left(\frac{(1+d)^j}{d} - \frac{1}{d} \right). \tag{61}$$

The above equation is linear in $x_{i,k}$, v, p_k^*, and $z_{i,k}$. Matching the coefficients with those of the conjectured strategy $x_{i,k} = av - bp_k + dz_{i,k}$ and solving the three equations, we have

$$b = a, \tag{62}$$

$$d = -\frac{2\gamma}{r} a, \tag{63}$$

$$a = a_\Delta \equiv \frac{r}{2\gamma} \left(1 + \frac{(n-1)(1 - e^{-r\Delta}) - \sqrt{(n-1)^2 (1 - e^{-r\Delta})^2 + 4e^{-r\Delta}}}{2e^{-r\Delta}} \right). \tag{64}$$

A.2 Proof of Proposition 2

Given the equilibrium price p^*, we can write the law of motion of the inventory of trader i as

$$\begin{aligned} z_{i,k+1} &= z_{i,k} + a_\Delta \left(v - p^* - \frac{2\gamma}{r} z_{i,k} \right) \\ &= z_{i,k} - a_\Delta \frac{2\gamma}{r} \left(z_{i,k} - \frac{Z}{n} \right), \end{aligned} \tag{65}$$

which implies that

$$z_{i,k+1} - \frac{Z}{n} = \left(1 - a_\Delta \frac{2\gamma}{r} \right) \left(z_{i,k} - \frac{Z}{n} \right). \tag{66}$$

We let

$$V_{i,0+} = \sum_{k=0}^{\infty} e^{-r\Delta k} E \left[-x_{i,k} p^* + (1 - e^{-r\Delta}) \left(v(x_{i,k} + z_{i,k}) - \frac{\gamma}{r} (x_{i,k} + z_{i,k})^2 \right) \,\middle|\, z_{i0}, Z \right]. \tag{67}$$

The inventories evolve according to

$$z_{i,k+1} - \frac{Z}{n} = \left(1 - a_\Delta \frac{2\gamma}{r} \right) \left(z_{i,k} - \frac{Z}{n} \right) = \left(1 - a_\Delta \frac{2\gamma}{r} \right)^{k+1} \left(z_{i,0} - \frac{Z}{n} \right). \tag{68}$$

It follows that, in equilibrium,

$$x_{i,k} = a_\Delta \frac{2\gamma}{r} \left(\frac{Z}{n} - z_{i,k} \right) = a_\Delta \frac{2\gamma}{r} \left(1 - a_\Delta \frac{2\gamma}{r} \right)^k \left(\frac{Z}{n} - z_{i,0} \right). \tag{69}$$

The price-related term in (67) is

$$\begin{aligned} \sum_{k=0}^{\infty} e^{-r\Delta k} p^* x_{i,k} &= \sum_{k=0}^{\infty} e^{-r\Delta k} \left(v - \frac{2\gamma}{nr} Z \right) a_\Delta \frac{2\gamma}{r} \left(1 - a_\Delta \frac{2\gamma}{r} \right)^k \left(\frac{Z}{n} - z_{i,0} \right) \\ &= \left(v - \frac{2\gamma}{nr} Z \right) \left(\frac{Z}{n} - z_{i,0} \right) \frac{a_\Delta \frac{2\gamma}{r}}{1 - e^{-r\Delta}(1 - a_\Delta \frac{2\gamma}{r})}. \end{aligned} \tag{70}$$

In (67), the term that involves v is

$$\begin{aligned} & v \sum_{k=0}^{\infty} \left(1 - e^{-r\Delta} \right) e^{-r\Delta k} \left[\frac{Z}{n} + \left(1 - a_\Delta \frac{2\gamma}{r} \right)^{k+1} \left(z_{i,0} - \frac{Z}{n} \right) \right] \\ &= v \frac{Z}{n} + \frac{(1 - a_\Delta \frac{2\gamma}{r}) \left(1 - e^{-r\Delta} \right)}{1 - e^{-r\Delta}(1 - a_\Delta \frac{2\gamma}{r})} v \left(z_{i,0} - \frac{Z}{n} \right). \end{aligned} \tag{71}$$

In (67), the term that involves γ is

$$
\begin{aligned}
&-\frac{\gamma}{r}\sum_{k=0}^{\infty}\left(1-e^{-r\Delta}\right)e^{-r\Delta k}\left[\frac{Z}{n}+\left(1-a_\Delta\frac{2\gamma}{r}\right)^{k+1}\left(z_{i,0}-\frac{Z}{n}\right)\right]^2\\
&\quad=-\frac{\gamma}{r}\left(\frac{Z}{n}\right)^2-\frac{(1-a_\Delta\frac{2\gamma}{r})\left(1-e^{-r\Delta}\right)}{1-e^{-r\Delta}(1-a_\Delta\frac{2\gamma}{r})}\frac{2\gamma Z}{nr}\left(z_{i,0}-\frac{Z}{n}\right)-\frac{\gamma}{r}\frac{1-a_\Delta\frac{2\gamma}{r}}{n-1}\left(z_{i,0}-\frac{Z}{n}\right)^2.
\end{aligned}
\tag{72}
$$

Adding up the three terms, we get the claimed expression for $V_{i,0+}$.

A.3 Proof of Corollary 1

The only nontrivial part of the proof is the limit of the convergence rate. Because $1-a_\Delta 2\gamma/r$ is the convergence factor per auction period, the associated convergence factor per unit of time is

$$\left(1-a_\Delta\frac{2\gamma}{r}\right)^{1/\Delta}.$$

Here, we ignore the effect of partial integer periods per unit of time, which is irrelevant in the limit as Δ goes to zero. Finally, we have the limiting convergence rate

$$\lim_{\Delta\to 0}\frac{\log(1-a_\Delta\frac{2\gamma}{r})}{\Delta}=-\lim_{\Delta\to 0}\frac{a_\Delta\frac{2\gamma}{r}}{\Delta}=-\frac{(n-2)r}{2}. \tag{73}$$

A.4 Proof of Proposition 5

The comparative statics of M with respect to r and Δ are provided in the text. The only item left is to show that M increases in n.

Define

$$A\equiv\frac{2e^{-r\Delta}}{1-e^{-r\Delta}}, \tag{74}$$

$$B\equiv\frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}, \tag{75}$$

$$\alpha_n=(n-1)+A+\sqrt{(n-1)^2+B}. \tag{76}$$

Then, we can write

$$C(n)=\frac{1}{n-1}\left(1-\frac{2(n-2)}{\alpha_n}\right). \tag{77}$$

To show that M increases in n, it is equivalent to show that

$$\frac{n\left(1+\frac{nC(n)}{1-C(n)}\right)}{n-1}>\frac{(n+1)\left(1+\frac{(n+1)C(n+1)}{1-C(n+1)}\right)}{n}, \tag{78}$$

which, after simplification, is equivalent to

$$\frac{(n+1)^2}{n-1}\frac{1}{1+\frac{2}{\alpha_{n+1}}}-\frac{n^2}{n-2}\frac{1}{1+\frac{2}{\alpha_n}}<1. \tag{79}$$

Note that α_n is increasing n, fixing other parameters. So,

$$\alpha_{n+1} - \alpha_n = 1 + \sqrt{n^2 + B} - \sqrt{(n-1)^2 + B} < 1 + \frac{(2n-1)}{2\sqrt{(n-1)^2 + B}} < \frac{4n-3}{2(n-1)}. \quad (80)$$

Using the above inequality, we can show that

$$\frac{1}{1 + \frac{2}{\alpha_{n+1}}} - \frac{1}{1 + \frac{2}{\alpha_n}} < \frac{1}{(\alpha_n + 2)^2} \frac{4n-3}{n-1}. \quad (81)$$

Applying the above inequality, we can show that the left-hand side of (79) satisfies:

$$\begin{aligned} &\frac{(n+1)^2}{n-1} \frac{1}{1 + \frac{2}{\alpha_{n+1}}} - \frac{n^2}{n-2} \frac{1}{1 + \frac{2}{\alpha_n}} \\ &< \frac{n^2 - 3n - 2}{(n-1)(n-2)} \frac{\alpha_n}{\alpha_n + 2} + \frac{(n+1)^2(4n-3)}{(n-1)^2} \frac{1}{(\alpha_n + 2)^2} \equiv \psi(\alpha_n). \end{aligned} \quad (82)$$

We have

$$\psi'(\alpha_n) = \frac{2}{(\alpha_n + 2)^2} \underbrace{\left[\frac{n^2 - 3n - 2}{(n-1)(n-2)} - \frac{1}{\alpha_n + 2} \frac{(n+1)^2(4n-3)}{(n-1)^2} \right]}_{\lambda(\Delta)}. \quad (83)$$

We now fix n and r, and consider changes in α_n through the changes in $\Delta > 0$. For any fixed n and r, α_n and $\lambda(\Delta)$ decrease in Δ. In particular, as $\Delta \to 0$, we have $\alpha_n \to \infty$, and $\lambda(0) > 0$. But as $\Delta \to \infty$, we have $\alpha_n \downarrow 2(n-1)$, and $\lambda(\Delta)$ converges to

$$\frac{n^2 - 3n - 2}{(n-1)(n-2)} - \frac{1}{2n} \frac{(n+1)^2(4n-3)}{(n-1)^2} < 0. \quad (84)$$

Therefore, as a function of α_n and in the domain $[2(n-1), \infty)$, $\psi(\alpha_n)$ first decreases in α_n and then increases in α_n. To show that $\psi(\alpha_n) < 1$, it suffices to verify that $\lim_{\alpha_n \to \infty} \psi(\alpha_n) < 1$ and $\psi(2(n-1)) < 1$. As $\alpha_n \to \infty$, the second term of $\psi(\alpha_n)$ vanishes and the first term converges to $\frac{n^2-3n-2}{(n-1)(n-2)} < 1$. At $\alpha_n = 2(n-1)$,

$$\begin{aligned} \psi(2(n-1)) &= \frac{n^2 - 3n - 2}{(n-1)(n-2)} \frac{n-1}{n} + \frac{(n+1)^2(4n-3)}{(n-1)^2} \frac{1}{4n^2} \\ &= 1 + \frac{1}{n} \left(\frac{(n+1)^2(4n-3)}{4n(n-1)^2} - \frac{n+2}{n-2} \right) \\ &< 1 + \frac{1}{n} \left(\frac{(n+1)^2}{(n-1)^2} - \frac{n+2}{n-2} \right) = 1 + \frac{1}{n} \left(\frac{4n}{(n-1)^2} - \frac{4}{n-2} \right) < 1. \end{aligned}$$

This completes the proof.

A.5 Proof of Proposition 6

Monotonicity of $M^*(m)$, $L(m)$, and $M^*(m)/L(m)$

The given expressions for $M^*(m)$ and $L(m)$ can be simplified by noting that n_b and n_s always appear only in the form of their sum $n_b + n_s$, which has a probability mass function $g(\cdot)$ defined by $g(k) = P(n_b + n_s = k)$, which is the convolution $g = f * f$. One can show that

$$g(k) = (k+1)q^k(1-q)^2. \tag{85}$$

Thus, we can write

$$M^*(m) = \frac{1}{\mu} \frac{\sum_{k=0}^{\infty} g(k)(1-C(n))\frac{n-1}{n^2}}{\sum_{k=0}^{\infty} g(k)\left(C(n) + \frac{1-C(n)}{n}\right)} \tag{86}$$

and

$$L(m) = \frac{\sum_{k=0}^{\infty} g(k)\frac{1-C(n)}{n}}{\sum_{k=0}^{\infty} g(k)\left(C(n) + \frac{(1-C(n))(3n-2)}{n^2}\right)}, \tag{87}$$

where $n = m + k$.

We first prove the following lemma.

Lemma 1. *If for positive real numbers $\{\lambda_i, \alpha_i, \beta_i : i \geq 0\}$, we have $\frac{\lambda_i}{\lambda_{i+1}} < \frac{\lambda_{i+1}}{\lambda_{i+2}}$ and $\frac{\alpha_i}{\beta_i} < \frac{\alpha_{i+1}}{\beta_{i+1}}$, then for any positive integer k,*

$$\frac{\sum_{i=0}^{k} \lambda_i \alpha_i}{\sum_{i=0}^{k} \lambda_i \beta_i} < \frac{\sum_{i=0}^{k} \lambda_i \alpha_{i+1}}{\sum_{i=0}^{k} \lambda_i \beta_{i+1}}.$$

Proof. Because $\frac{\alpha_i}{\beta_i} < \frac{\alpha_{i+1}}{\beta_{i+1}}$ for $i \geq 0$, it is easy to see that

$$\frac{\sum_{i=0}^{k-1} \lambda_i \alpha_{i+1}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}} = \frac{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}\frac{\alpha_{i+1}}{\beta_{i+1}}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}} \leq \frac{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}\frac{\alpha_k}{\beta_k}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}} = \frac{\alpha_k}{\beta_k},$$

which implies

$$\frac{\alpha_{k+1}}{\beta_{k+1}} > \frac{\alpha_k}{\beta_k} \geq \frac{\sum_{i=0}^{k-1} \lambda_i \alpha_{i+1}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}}. \tag{88}$$

From (88) we have

$$\begin{aligned}
\frac{\sum_{i=0}^{k} \lambda_i \alpha_{i+1}}{\sum_{i=0}^{k} \lambda_i \beta_{i+1}} &= \frac{\sum_{i=0}^{k-1} \lambda_i \alpha_{i+1} + \lambda_k \beta_{k+1}\frac{\alpha_{k+1}}{\beta_{k+1}}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1} + \lambda_k \beta_{k+1}} \\
&\geq \frac{\sum_{i=0}^{k-1} \lambda_i \alpha_{i+1} + \lambda_k \beta_{k+1}\frac{\sum_{i=0}^{k-1} \lambda_i \alpha_{i+1}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1} + \lambda_k \beta_{k+1}} \\
&= \frac{\sum_{i=0}^{k-1} \lambda_i \alpha_{i+1}}{\sum_{i=0}^{k-1} \lambda_i \beta_{i+1}}.
\end{aligned} \tag{89}$$

Similarly, we can prove that

$$\frac{\sum_{i=0}^{k} \lambda_i \alpha_i}{\sum_{i=0}^{k} \lambda_i \beta_i} \leq \frac{\sum_{i=1}^{k} \lambda_i \alpha_i}{\sum_{i=1}^{k} \lambda_i \beta_i}. \tag{90}$$

Equations (89) and (90) imply that in order to prove Lemma 1 we only need to show that

$$\frac{\sum_{i=1}^{k}\lambda_i\alpha_i}{\sum_{i=1}^{k}\lambda_i\beta_i} < \frac{\sum_{i=0}^{k-1}\lambda_i\alpha_{i+1}}{\sum_{i=0}^{k-1}\lambda_i\beta_{i+1}},$$

which is equivalent to

$$\frac{\sum_{i=1}^{k}\lambda_i\alpha_i}{\sum_{i=1}^{k}\lambda_i\beta_i} < \frac{\sum_{i=1}^{k}\lambda_{i-1}\alpha_i}{\sum_{i=1}^{k}\lambda_{i-1}\beta_i}. \tag{91}$$

Notice that

$$\frac{\sum_{i=1}^{k}\lambda_{i-1}\alpha_i}{\sum_{i=1}^{k}\lambda_{i-1}\beta_i} - \frac{\sum_{i=1}^{k}\lambda_i\alpha_i}{\sum_{i=1}^{k}\lambda_i\beta_i} = \frac{\left(\sum_{i=1}^{k}\lambda_{i-1}\alpha_i\right)\left(\sum_{i=1}^{k}\lambda_i\beta_i\right) - \left(\sum_{i=1}^{k}\lambda_i\alpha_i\right)\left(\sum_{i=1}^{k}\lambda_{i-1}\beta_i\right)}{\left(\sum_{i=1}^{k}\lambda_{i-1}\beta_i\right)\left(\sum_{i=1}^{k}\lambda_i\beta_i\right)}. \tag{92}$$

So it suffices to prove the numerator of (92) is positive. By expansion we have

$$\begin{aligned}
&\left(\sum_{i=1}^{k}\lambda_{i-1}\alpha_i\right)\left(\sum_{i=1}^{k}\lambda_i\beta_i\right) - \left(\sum_{i=1}^{k}\lambda_i\alpha_i\right)\left(\sum_{i=1}^{k}\lambda_{i-1}\beta_i\right) \\
= &\sum_{1\le s<t\le k}\left(\lambda_{s-1}\alpha_s\lambda_t\beta_t + \lambda_{t-1}\alpha_t\lambda_s\beta_s\right) + \sum_{i=1}^{k}\lambda_{i-1}\alpha_i\lambda_i\beta_i \\
&- \sum_{1\le s<t\le k}\left(\lambda_s\alpha_s\lambda_{t-1}\beta_t + \lambda_t\alpha_t\lambda_{s-1}\beta_s\right) - \sum_{i=1}^{k}\lambda_{i-1}\beta_i\lambda_i\alpha_i \\
= &\sum_{1\le s<t\le k}\left(\lambda_{s-1}\alpha_s\lambda_t\beta_t + \lambda_{t-1}\alpha_t\lambda_s\beta_s - \lambda_s\alpha_s\lambda_{t-1}\beta_t - \lambda_t\alpha_t\lambda_{s-1}\beta_s\right).
\end{aligned} \tag{93}$$

Because, for all $s < t$,

$$\begin{aligned}
&\lambda_{s-1}\alpha_s\lambda_t\beta_t + \lambda_{t-1}\alpha_t\lambda_s\beta_s - \lambda_s\alpha_s\lambda_{t-1}\beta_t - \lambda_t\alpha_t\lambda_{s-1}\beta_s \\
= &\lambda_s\lambda_t\beta_s\beta_t\left(\frac{\lambda_{s-1}}{\lambda_s}\frac{\alpha_s}{\beta_s} + \frac{\lambda_{t-1}}{\lambda_t}\frac{\alpha_t}{\beta_t} - \frac{\lambda_{t-1}}{\lambda_t}\frac{\alpha_s}{\beta_s} - \frac{\lambda_{s-1}}{\lambda_s}\frac{\alpha_t}{\beta_t}\right) \\
= &\lambda_s\lambda_t\beta_s\beta_t\left(\frac{\lambda_{s-1}}{\lambda_s} - \frac{\lambda_{t-1}}{\lambda_t}\right)\left(\frac{\alpha_s}{\beta_s} - \frac{\alpha_t}{\beta_t}\right) \\
> &0,
\end{aligned} \tag{94}$$

the right side of (93) is positive, and the proof of the Lemma is complete. □

Letting $k \to \infty$ in Lemma 1, we get

$$\frac{\sum_{i=0}^{\infty}\lambda_i\alpha_i}{\sum_{i=0}^{\infty}\lambda_i\beta_i} \le \frac{\sum_{i=0}^{\infty}\lambda_i\alpha_{i+1}}{\sum_{i=0}^{\infty}\lambda_i\beta_{i+1}}. \tag{95}$$

Letting $\lambda_i = g(i) = (i+1)q^i(1-q)^2$, we have

$$\frac{\lambda_i}{\lambda_{i+1}} = \frac{(i+1)}{(i+2)q} < \frac{(i+2)}{(i+3)q} = \frac{\lambda_{i+1}}{\lambda_{i+2}}. \tag{96}$$

Monotonicity of $M^*(m)$. Given Lemma 1, to show that $M^*(m) \le M^*(m+1)$, it suffices to show that

$$\frac{(1-C(n))\frac{n-1}{n^2}}{C(n)+\frac{1-C(n)}{n}} \tag{97}$$

is increasing in n for $n \ge 2$.

In the continuous-time double-auction market of Appendix B, we have $\Delta = 0$ and $C(n) = 1/(n-1)$, so the ratio (97) simplifies to

$$\frac{n-2}{2n},$$

which is increasing in n.

For $\Delta > 0$, we denote

$$D(n) = 2 - \frac{2(n-2)}{(n-1)+\frac{2e^{-r\Delta}}{1-e^{-r\Delta}}+\sqrt{(n-1)^2+\frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}}}. \tag{98}$$

It is easy to see that $D(n)$ is decreasing in n. Using $C(n) = (1-2a_\Delta\gamma/r)/(n-1)$, we have

$$\frac{(1-C(n))\frac{n-1}{n^2}}{C(n)+\frac{1-C(n)}{n}} = \frac{1-D(n)/n}{D(n)} = \frac{1}{D(n)} - \frac{1}{n}. \tag{99}$$

Since $D(n)$ is decreasing in n, the right-hand side of the above expression is increasing in n, and the proof for the monotonicity of $M^*(m)$ is complete.

Monotonicity of $L(m)$ if $e^{-r\Delta} > 1/2$. Given Lemma 1, to show that $L(m) \le L(m+1)$, it suffices to show that

$$\frac{\frac{1-C(n)}{n}}{C(n)+\frac{(1-C(n))(3n-2)}{n^2}} \tag{100}$$

is increasing in n for $n \ge 2$.

In the continuous-time double-auction market, with $\Delta = 0$ and $C(n) = 1/(n-1)$, the ratio (100) simplifies to

$$\frac{(n-2)n}{4(n-1)^2},$$

which is indeed increasing in n.

For $\Delta > 0$, we define

$$t = \frac{2e^{-r\Delta}}{1-e^{-r\Delta}}, \tag{101}$$

and

$$R(n) = \sqrt{(n-1)^2+\frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}} - (n-1). \tag{102}$$

Now we can write

$$C(n) = \frac{1}{n-1}\left(1 - \frac{2(n-2)}{n-1+t+n-1+R(n)}\right). \tag{103}$$

We claim that

$$0 \leq R(n) \leq t, \text{ and } R(n) \text{ decreases in } n. \tag{104}$$

It is obvious that $R(n)$ is non-negative and decreases in n. To see that $R(n) \leq t$, we can directly calculate

$$\frac{\frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}}{\sqrt{(n-1)^2 + \frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}} + n - 1} \leq \frac{\frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}}{\sqrt{1 + \frac{4e^{-r\Delta}}{(1-e^{-r\Delta})^2}} + 1} = t.$$

Using (101) and (102), we can write

$$\frac{\frac{(1-C(n))}{(n)}}{\left(C(n) + \frac{(1-C(n))(3n-2)}{(n)^2}\right)} = \frac{(n-2)n(2n+t+R(n))}{2(n-1)(3n^2+2n(t-2)-2t+2(n-1)R(n))}. \tag{105}$$

We denote the numerator and denominator of the right-hand side of (105) by $Y_1(n)$ and $Y_0(n)$, respectively. To show monotonicity, it is enough to prove that

$$Y_1(n+1)Y_0(n) - Y_1(n)Y_0(n+1) > 0.$$

After expansion, we get

$$\begin{aligned} & Y_1(n+1)Y_0(n) - Y_1(n)Y_0(n+1) \\ &= 2(R(n) - R(n+1))n^5 + 2(t-2+R(n+1))n^4 \\ &\quad + (24 - 4t - 6R(n) + 2R(n+1))\, n^3 + 2(6 + 13t + 6R(n) + 7R(n+1))n^2 \\ &\quad + 8(-2 + t^2 + (t-1)R(n+1) + R(n)(t+1+R(n+1)))n \\ &\quad - 4(t+R(n))(t+2+R(n+1)). \end{aligned} \tag{106}$$

Using (104), lower bounds on the coefficients of each term in the polynomial on the right-hand side are as follows:

$$n^5: \quad 2(R(n) - R(n+1)) \geq 0. \tag{107}$$

$$n^4: \quad 2(t-2+R(n+1)) \geq 2(t-2). \tag{108}$$

$$n^3: \quad 24 - 4t - 6R(n) + 2R(n+1) \geq 24 - 10t. \tag{109}$$

$$n^2: \quad 2(6 + 13t + 6R(n) + 7R(n+1)) \geq 2(6+13t). \tag{110}$$

$$n: \quad 8(-2 + t^2 + (t-1)R(n+1) + R(n)(t+1+R(n+1))) \geq 8(t^2-2). \tag{111}$$

$$\text{Constant}: \quad -4(t+R(n))(t+2+R(n+1)) \geq -8t(2t+2). \tag{112}$$

With the above inequalities, we get

$$\begin{aligned} & Y_1(n+1)Y_0(n) - Y_1(n)Y_0(n+1) \\ &\geq 2(t-2)n^4 + (24-10t)\, n^3 + 2(6+13t)n^2 + 8(t^2-2)n - 8t(2t+2) \\ &= 2(t-2)n^2(n^2 - 5n + 6.5) + 4n^3 + (38+13t)n^2 + 8(t^2-2)n - 8t(2t+2) \\ &= 2(t-2)n^2\left((n-2.5)^2 + 0.25\right) + 4n^3 + 38n^2 - 16n + t^2(8n-16) + t(13n^2 - 16). \end{aligned} \tag{113}$$

Under the condition $e^{-r\Delta} > \frac{1}{2}$, we have $t > 2$. It is easy to see the right-hand side of (113) is positive for $n \geq 2$.

Monotonicity of $M^*(m)/L(m)$. We can write

$$\frac{M^*(m)}{L(m)} = \frac{1}{\mu}\frac{\sum_{k=0}^{\infty} g(k)(1-C(n))\frac{n-1}{n^2}}{\sum_{k=0}^{\infty} g(k)\frac{1-C(n)}{n}} \cdot \frac{\sum_{k=0}^{\infty}\left(C(n)+\frac{(1-C(n))(3n-2)}{n^2}\right)}{\sum_{k=0}^{\infty} g(k)\left(C(n)+\frac{1-C(n)}{n}\right)}. \tag{114}$$

Given Lemma 1, to show that $M^*(m)/L(m)$ increases in m, it suffices to show that

$$\frac{(1-C(n))\frac{n-1}{n^2}}{\frac{1-C(n)}{n}} \quad \text{and} \quad \frac{C(n)+\frac{(1-C(n))(3n-2)}{n^2}}{C(n)+\frac{1-C(n)}{n}}$$

are both increasing in n.

Monotonicity in m of the first expression is obvious, for

$$\frac{(1-C(n))\frac{n-1}{n^2}}{\frac{1-C(n)}{n}} = 1-\frac{1}{n}.$$

The second expression can be expressed as

$$\frac{C(n)+\frac{(1-C(n))(3n-2)}{n^2}}{C(n)+\frac{1-C(n)}{n}} = 1+2\frac{(1-C(n))\frac{n-1}{n^2}}{C(n)+\frac{1-C(n)}{n}}. \tag{115}$$

The last term in the above expression is increasing in n, as shown in the proof of monotonicity of $M^*(m)$.

Proof of (53)–(56)

Suppose that $e^{-r\Delta} > 1/2$. We have shown that in this case $M^*(m)$, $L(m)$ and $M^*(m)/L(m)$ are all increasing in m for $m \geq 2$. We now prove (53)–(56) by induction.

At $i = j = 1$ and $X_{1,1} = 0$. Clearly, both thresholds are equal to $M^*(2)$ and are positive at this initial state. Moreover, if the buyer exits, then $X_{2,1} < 0$, and

$$\mathcal{M}_s(2,1,X_{2,1}) = M^*(3) - L(3)X_{2,1} > M^*(2) = \mathcal{M}_s(1,1,X_{1,1}). \tag{116}$$

If the seller exits, then $X_{1,2} > 0$, and

$$\mathcal{M}_b(1,2,X_{1,2}) = M^*(3) + L(3)X_{1,2} > M^*(2) = \mathcal{M}_b(1,1,X_{1,1}). \tag{117}$$

At generic (i,j) and $X_{i,j}$. By symmetry, it suffices to prove these inequalities for the exit of seller j. By the conjectured update rule,

$$X_{i,j+1} - X_{i,j} = \begin{cases} M^*(i+j) - L(i+j)X_{i,j}, & \text{if seller } j \text{ traded positive quantity} \\ \nu(M^*(i+j) - L(i+j)X_{i,j}), & \text{if seller } j \text{ traded zero quantity.} \end{cases} \tag{118}$$

where the last line follows from the induction step

$$\mathcal{M}_s(i,j,X_{i,j}) = M^*(i+j) - L(i+j)X_{i,j} \geq 0.$$

In this case, we want to show that

$$M^*(i+j+1) + L(i+j+1)X_{i,j+1} \geq M^*(i+j) + L(i+j)X_{i,j}, \tag{119}$$

$$M^*(i+j+1) - L(i+j+1)X_{i,j+1} \geq 0. \tag{120}$$

If established, the first inequality (119) would imply that the incumbent buyer's new threshold remains positive if the old threshold is positive. Since it is the seller who exited, the inequality for the "incumbent seller" is irrelevant.

To show (119), we calculate

$$\begin{aligned}
&M^*(i+j+1) + L(i+j+1)X_{i,j+1} - M^*(i+j) - L(i+j)X_{i,j} \\
&\quad \geq M^*(i+j+1) - M^*(i+j) + (L(i+j+1) - L(i+j))X_{i,j} \\
&\quad \geq M^*(i+j)\frac{L(i+j+1)}{L(i+j)} - M^*(i+j) + (L(i+j+1) - L(i+j))X_{i,j} \\
&\quad = \frac{L(i+j+1) - L(i+j)}{L(i+j)}(M^*(i+j) + L(i+j)X_{i,j}) \geq 0,
\end{aligned} \tag{121}$$

where the last inequality follows from the induction step that $\mathcal{M}_b(i,j,X_{i,j}) \geq 0$ and the monotonicity of $L(m)$, and the penultimate inequality follows from the monotonicity of $M^*(m)/L(m)$.

To show (120), we calculate

$$\begin{aligned}
&M^*(i+j+1) - L(i+j+1)X_{i,j+1} \\
&\quad \geq M^*(i+j+1) - L(i+j+1)\left(X_{i,j} + M^*(i+j) - L(i+j)X_{i,j}\right) \\
&\quad \geq M^*(i+j)\frac{L(i+j+1)}{L(i+j)} - L(i+j+1)M^*(i+j) - L(i+j+1)(1 - L(i+j))X_{i,j} \\
&\quad = \frac{L(i+j+1)(1 - L(i+j))}{L(i+j)}\left(M^*(i+j) - L(i+j)X_{i,j}\right) \geq 0,
\end{aligned} \tag{122}$$

where the last inequality follows from the induction step that $\mathcal{M}_s(i,j,X_{i,j}) \geq 0$ and the penultimate inequality follows from the monotonicity of $M^*(m)/L(m)$.

B Continuous-Time Double-Auction Market

This appendix models equilibrium for dynamic trading in a continuous double-auction market that is a precise analogue of the discrete-time auction market of Section 2, and whose equilibrium behavior is equal to the limit behavior of the discrete-time market as the time Δ between discretely-spaced auctions goes to zero.

We fix a probability space and the time domain $[0, \infty)$. The setup for the joint distribution of the exponential payoff time T, the payoff π of the asset, and the initial inventories $z_0 = (z_{10}, z_{20}, \ldots, z_{n0})$ of the $n \geq 3$ of traders is precisely the same as that for the discrete-time auction model of Section 2. The initial information structure is also as in Section 2. In our application of this model in Section 4, the number n of traders is an outcome of the random workup population size $N_b + N_s$. The outcome of $N_b + N_s$ is publicly known when workup is complete. So, it is enough to solve the continuous-time auction model for any fixed trader population size n.

In our new continuous-time setting, a demand schedule at time can be expressed by a demand "rate function" $D : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, representing the rate of demand $D_t(p)$ of asset per unit of time at time t and at price p. This means that if the time path of prices is given in some state of the world by some function $\phi : [0, \infty) \to \mathbb{R}$, then by time t the cumulative quantity purchased is $\int_0^t D_s(\phi(s))\, ds$ and the total price paid is $\int_0^t \phi(s) D_s(\phi(s))\, ds$, whenever these integrals are well defined.

We will consider an equilibrium in which demand $D_{it}(p)$ of trader i at time t and price p is continuous in both t and p and strictly decreasing in p, and such that the market clearing price $\phi(t)$ at time t, when well defined, is the solution in p of the market-clearing equation:

$$\sum_i D_{it}(p) = 0. \tag{123}$$

This market clearing price, when well defined, is denoted $\Phi(\sum_i D_{it})$.

An equilibrium is a collection $(D_1, \ldots, D_n)$ of demand functions such that, for each time t the market-clearing price $\Phi(\sum_i D_{it})$ is well defined and such that, for agent i, the demand function D_i solves the problem, taking $D_{-i} = \sum_{j \neq i} D_j$ as given,

$$\sup_D E \left[z_i^D(T)\pi - \int_0^T \left[\gamma z_i^D(t)^2 + D_t \left[\Phi \left(D_t + D_{-i,t} \right) \right] \Phi \left(D_t + D_{-it} \right) \right] \, dt \right], \tag{124}$$

where $\gamma > 0$ is a holding-cost parameter and

$$z_i^D(t) = z_{i0} + \int_0^t D_s \left[\Phi \left(D_s + D_{-i,s} \right) \right] \, ds$$

is the inventory of agent i at time t.

We will look for an equilibrium in which the initial price $p(0)$ instantly reveals the total market supply Z and in which the demand function of trader i depends only his current inventory z_{it}.

We will conjecture and verify the equilibrium given by

$$D_{it}(p) = a \left(v - p - \frac{2\gamma}{r} z_{it} \right), \tag{125}$$

where

$$a = \frac{(n-2)r^2}{4\gamma}. \tag{126}$$

The unique associated market-clearing price at any time t is

$$p^* = v - \frac{2\gamma}{nr} Z. \tag{127}$$

From this, the inventory position of trader i at time t can be calculated as

$$z_{it} = \frac{Z}{n} + e^{-(n-2)rt/2} \left(z_{i0} - \frac{Z}{n} \right). \tag{128}$$

Given this conjectured equilibrium, for any agent i, the sum $D_{-i,t}$ of the demand functions of the other agents at time t is

$$D_{-i,t}(p) = \mathcal{D}_{-i}(p; z_{it}, Z) \equiv \sum_{j \neq i} a\left(v - p - \frac{2\gamma}{r} z_{jt}\right) = (n-1)a(v-p) - \frac{2a\gamma}{r}(Z - z_{it}).$$

Based on this calculation, the continuation utility $V(z)$ for the inventory level of any trader at any time $t < T$ who has the inventory level z satisfies the Hamilton-Jacobi-Bellman (HJB) equation

$$\begin{aligned} 0 = \sup_D & \left[-D(\Phi(D + \mathcal{D}_{-i}(\,\cdot\,; z, Z))\Phi(D + \mathcal{D}_{-i}(\,\cdot\,; z, Z)) + V'(z)D(\Phi(D + \mathcal{D}_{-i}(\,\cdot\,; z, Z))\right] \\ & - \gamma z^2 + r(vz - V(z)). \end{aligned} \tag{129}$$

The first term on the right-hand side of (129) is the rate of cost of acquiring inventory in auctions, that is, the quantity rate $D(\Phi(D + \mathcal{D}_{-i}(\,\cdot\,; z, Z))$ multiplied by the price $\Phi(D+\mathcal{D}_{-i}(\,\cdot\,; z, Z))$. The second term is the marginal value $V'(z)$ of inventory multiplied by the rate $D(\Phi(D + \mathcal{D}_{-i}(\,\cdot\,; z, Z))$ of inventory accumulation. The sum of these first two terms is optimized by choosing some demand function D. The next term accounts for the rate of inventory holding cost, γz^2. The final term is the product of the mean rate r of arrival of the time of the asset payoff and the expected change $vz - V(z)$ in the trader's indirect utility if that payoff were to occur immediately.

Because Z is constant and observable after time 0, the HJB equation does not pin down a unique optimizing demand function $D(\,\cdot\,)$. Instead, the HJB equation makes the demand problem for agent i equivalent to picking the quantity x the agent wishes to buy, and then submitting any demand function $D(\,\cdot\,)$ with the property that $D(p) = x$, where p solves $x + \mathcal{D}_{-i}(p; z, Z) = 0$. In order to avoid degenerate behavior of this type, we require that the submitted demand function $D_i(\,\cdot\,)$ must depend only on the inventory z_{it} trader i and of course the price p. That is, we require that $D_{it}(p) = f_t(p, z_{it})$ for some function $f_t : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$. Nevertheless, in equilibrium, the resulting demand will turn out to be optimal even if the class of demand functions is expanded to allow dependence on Z.

We will conjecture and verify that, in equilibrium,

$$V(z) = v\frac{Z}{n} - \frac{\gamma}{r}\left(\frac{Z}{n}\right)^2 + \left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right)\left(z - \frac{Z}{n}\right) - \frac{\gamma}{r}\frac{1}{n-1}\left(z - \frac{Z}{n}\right)^2. \tag{130}$$

We use the fact that V is quadratic and concave, thus bounded above.

Proposition 8. *Suppose, for a given trader i, that the demand function D_j for any trader $j \neq i$ is given by* (125). *The function V given by* (130) *satisfies the HJB equation* (129). *Given this choice for V, the optimization problem posed within the HJB equation is satisfied by the demand function D_{it} of* (125). *The optimal demand problem* (124) *for agent i is also solved by* (125).

This result is shown as follows. With (130), the HJB equation, applied to agent i at time t, is equivalent to solving, for each outcome of Z, the optimal demand

$$\sup_x \left[-x\mathcal{D}_{-i}^{-1}(-x; z_{it}, Z) + V'(z_{it})x\right], \tag{131}$$

where $\mathcal{D}_{-i}^{-1}(q; z_{it}, Z)$ is the inverse total demand of the other agents at any quantity q, meaning that price p for which

$$q = (n-1)a(v-p) - \frac{2a\gamma}{r}(Z - z_{it}).$$

Solving,

$$\mathcal{D}_{-i}^{-1}(q; z_{it}, Z) = v - \frac{1}{a(n-1)}\left[q + \frac{2a\gamma}{r}(Z - z_{it})\right].$$

Thus, the demand problem of agent i is

$$\sup_x \left[-x\left(v - \frac{1}{a(n-1)}\left[-x + \frac{2a\gamma}{r}(Z - z_{it})\right]\right) + V'(z_{it})x\right]. \tag{132}$$

The first-order necessary condition for optimality of x^* is

$$-v + \frac{2\gamma}{r(n-1)}(Z - z_{it}) + V'(z_{it}) - 2x^* \frac{1}{a(n-1)} = 0,$$

where

$$V'(z_{it}) = v - 2\frac{\gamma}{r}\frac{Z}{n} - 2\frac{\gamma}{r}\frac{1}{n-1}\left(z_{it} - \frac{Z}{n}\right).$$

The unique solution x^* of this first-order condition also satisfies the second-order sufficiency condition, and is given by

$$x^* = \frac{2a\gamma}{r}\left(\frac{Z}{n} - z_{it}\right).$$

The associated market clearing price is

$$p^* = \mathcal{D}_{-i}^{-1}(-x^*; z_{it}, Z) = v - \frac{1}{a(n-1)}\left[-x^* + \frac{2a\gamma}{r}(Z - z_{it})\right] = v - 2\frac{\gamma}{r}\frac{Z}{n}. \tag{133}$$

We now verify that the postulated demand function D_{it} for agent i achieves the above demand x^*, regardless of the outcome of Z. We have

$$D_{it}(p^*) = a\left(v - p^* - \frac{2\gamma}{r}z_{it}\right) = a\left(v - \left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right) - \frac{2\gamma}{r}z_{it}\right) = \frac{2a\gamma}{r}\left(\frac{Z}{n} - z_{it}\right),$$

which is indeed equal to the optimal demand x^*.

In order to prove that the proposed indirect utility function V satisfies the HJB equation, we substitute our expressions for $V(z)$, p^*, and $D_{it}(p^*)$ into the right-hand-side of the HJB equation (129). To confirm that (129) is satisfied, we must show that for all real z and Z,

$$0 = -\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right)\left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right) + V'(z)\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) + r(vz - V(z)) - \gamma z^2. \tag{134}$$

To see that (134) holds, note that

$$V'(z)\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) = v\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) - 2\frac{\gamma}{r}\frac{Z}{n}\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) + 2\frac{\gamma}{r}\frac{1}{n-1}\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right)^2$$

and that

$$r(vz - V(z)) = rvz - rv\frac{Z}{n} + r\frac{\gamma}{r}\left(\frac{Z}{n}\right)^2 - r\left(v - \frac{2\gamma}{r}\frac{Z}{n}\right)\left(z - \frac{Z}{n}\right) + r\frac{\gamma}{r(n-1)}\left(z - \frac{Z}{n}\right)^2$$
$$= r\frac{\gamma}{r}\left(\frac{Z}{n}\right)^2 + r\frac{2\gamma}{r}\frac{Z}{n}\left(z - \frac{Z}{n}\right) + r\frac{\gamma}{r(n-1)}\left(z - \frac{Z}{n}\right)^2. \tag{135}$$

The right-hand side of (134) is thus computed as

$$-\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right)\left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right) + V'(z)\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) + r(vz - V(z)) - \gamma z^2$$
$$= \frac{4a\gamma}{r}\left(\frac{Z}{n} - z\right)\frac{\gamma}{r}\frac{Z}{n} + -2\frac{\gamma}{r}\frac{Z}{n}\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) + 2\frac{\gamma}{r}\frac{1}{n-1}\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right)^2$$
$$+\gamma\left(\frac{Z}{n}\right)^2 + 2\gamma\frac{Z}{n}\left(z - \frac{Z}{n}\right) + \frac{\gamma}{n-1}\left(z - \frac{Z}{n}\right)^2 - \gamma z^2.$$

Substituting $a = (n-2)r^2/4\gamma$, we have

$$-\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right)\left(v - 2\frac{\gamma}{r}\frac{Z}{n}\right) + V'(z)\frac{2a\gamma}{r}\left(\frac{Z}{n} - z\right) + r(vz - V(z)) - \gamma z^2$$
$$= (n-2)\gamma\left(\frac{Z}{n} - z\right)\frac{Z}{n} + -(n-2)\gamma\frac{Z}{n}\left(\frac{Z}{n} - z\right) + \frac{(n-2)\gamma}{n-1}\left(\frac{Z}{n} - z\right)^2$$
$$+\gamma\left(\frac{Z}{n}\right)^2 + 2\gamma\frac{Z}{n}\left(z - \frac{Z}{n}\right) + \frac{\gamma}{n-1}\left(z - \frac{Z}{n}\right)^2 - \gamma z^2.$$

So, V satisfies the HJB equation because

$$\frac{(n-2)}{n-1}\left(\frac{Z}{n} - z\right)^2 + \left(\frac{Z}{n}\right)^2 + 2\frac{Z}{n}\left(z - \frac{Z}{n}\right) + \frac{1}{n-1}\left(z - \frac{Z}{n}\right)^2 - z^2 = 0.$$

Thus, using the fact that the demand function D_{it} solves the maximization problem of the HJB equation, and using the fact that V solves the HJB equation, an application of Ito's formula to the process J defined by $J(t) = V(z_{it})$ for $t < T$, and by $J(t) = \pi z_{iT}$ for $t \geq T$ implies that

$$V(z_{i0}) = E\left[z_{iT}(T)\pi - \int_0^T \left[\gamma z_{it}^2 + D_{it}\left[\Phi\left(D_{it} + D_{-i,t}\right)\right]\Phi\left(D_{it} + D_{-it}\right)\right]\,dt\right].$$

For any other demand function D for agent i, the HJB equation and Ito's formula implies that implies that

$$V(z_{i0}) \geq E\left[z_i^D(T)\pi - \int_0^T \left[\gamma z_i^D(t)^2 + D_t\left[\Phi\left(D_t + D_{-i,t}\right)\right]\Phi\left(D_t + D_{-it}\right)\right]\,dt\right].$$

Thus D_i is indeed optimal for trader i given D_{-i}, and $V(z)$ is indeed the indirect utility of any agent with inventory z. This proves Proposition 8.

We can now recapitulate the main results of this section:

Proposition 9. *An equilibrium of the continuous-time double-auction market is as follows.*

1. *The demand function D_{it} of trader i at time t is given by:*

$$D_{it} = \frac{(n-2)r^2}{4\gamma}\left(v - p - \frac{2\gamma}{r}z_{it}\right), \tag{136}$$

where the equilibrium inventory of trader i at time t is

$$z_{it} = \frac{Z}{n} + e^{-(n-2)rt/2}\left(z_{i0} - \frac{Z}{n}\right). \tag{137}$$

The equilibrium price at time t is constant at

$$p^* = v - \frac{2\gamma}{nr}Z. \tag{138}$$

2. *The indirect utility $V(z)$ of any agent i for inventory z at any time $t > 0$ that is before the asset payoff time T is given by*

$$V(z) = v\frac{Z}{n} - \frac{\gamma}{r}\left(\frac{Z}{n}\right)^2 + \left(v - \frac{2\gamma}{r}\frac{Z}{n}\right)\left(z - \frac{Z}{n}\right) - \frac{\gamma}{r(n-1)}\left(z - \frac{Z}{n}\right)^2. \tag{139}$$

C Welfare and Squared Asset Dispersion

A re-allocation of the inventory vector $(z_{10}, \ldots, z_{in})$ is an allocation $z' = (z'_1, \ldots, z'_n)$ with the same total Z. A re-allocation z' is a Pareto improvement if, when replacing z_{i0} with z'_i, the equilibrium utility $E(V_{i,0+})$ before entering the sequential-double-auction market is weakly increased for every i and strictly increased for some i. We have the following corollary of Proposition 2.

Corollary 2. *The total expected ex-ante utility $W(z_0) = \sum_{i=1}^n E(V_{i,0+})$ is one-to-one and strictly monotone decreasing (in fact linear) in the sum of mean squared excess asset positions,*

$$D(z_0) = E\left(\sum_{i=1}^n \left(z_{i0} - \frac{Z}{n}\right)^2\right).$$

Thus, if a re-allocation $z' = (z'_1, \ldots, z'_n)$ is a Pareto improvement, then $D(z') < D(z_0)$.

This result follows from the fact that $W(z_0)$ is a constant plus the product of $D(z_0)$ and a negative constant.

Because traders' preferences are linear with respect to total net pecuniary benefits, $W(\cdot)$ is a reasonable social welfare function. This follows from the fact that for any allocations z' and z with $W(z') > W(z)$, the allocation z' is Pareto preferred to z after allowing for transfer payments.

The magnitude of welfare improvement offered by the bilateral workup can be calculated explicitly. We focus on the welfare of the two workup participants under consideration. Starting from any pre-workup inventory levels x and y consistent with active workup, that is satisfying $-x < -(M+\delta)$ (for the buyer) and $y > M - \delta$ (for the seller), the post-workup inventories are $-x+V$ and $y-V$, where $V \equiv \min(x-(M+\delta), y-(M-\delta))$

is the workup trading volume. The welfare improvement induced by workup is therefore

$$
\begin{aligned}
&E\left[(-x - Z/n)^2 + (y - Z/n)^2 - (-x + V - Z/n)^2 - (y - V - Z/n)^2\right] \\
&\quad = E\left[-2V^2 + 2(x+y)V\right] \\
&\quad = \int_{x=M+\delta}^{\infty} \int_{y=M-\delta}^{\infty} \mu e^{-\mu x} \mu e^{-\mu y} (-2V^2 + 2(x+y)V)\, dx\, dy. \qquad (140)
\end{aligned}
$$

A change of variables, taking $u = x - M - \delta$ and $w = y - M + \delta$, allows one to re-express the integral as

$$
\int_{u=0}^{\infty} \int_{w=0}^{\infty} \mu^2 e^{-2\mu M} e^{-\mu(u+w)} \left[-2\min(u,w)^2 + 2(u + w + 2M)\min(u,w)\right] du\, dw, \quad (141)
$$

which is invariant to δ in the interval $[0, M]$.

This calculation can obviously be done on any bilateral workup.

References

Arrow, K. (1951): "An Extension of the Basic Theorems of Classical Welfare Economics," in *Second Berkeley Symposium on Mathematical Statistics and Probability*, ed. by J. Neyman, Berkeley: University of California Press, 507–532.

Aumann, R. (1964): "Markets with a Continuum of Traders," *Econometrica*, 32, 39–50.

Ausubel, L. M., P. Cramton, M. Pycia, M. Rostek, and M. Weretka (2014): "Demand Reduction, Inefficiency and Revenues in Multi-Unit Auctions," *Review of Economic Studies*, 81, 1366–1400.

Boni, L. and C. Leach (2004): "Expandable Limit Order Markets," *Journal of Financial Markets*, 7, 145–185.

Boni, L. and J. C. Leach (2002): "Supply Contraction and Trading Protocol: An Examination of Recent Changes in the U.S. Treasury Market," *Journal of Money, Credit, and Banking*, 34, 740–762.

Buti, S., B. Rindi, and I. M. Werner (2011): "Diving into Dark Pools," Working paper, Fisher College of Business, Ohio State University.

——— (2015): "Dark Pool Trading Strategies, Market Quality and Welfare," Working paper, University of Toronto.

Degryse, H., M. Van Achter, and G. Wuyts (2009): "Dynamic Order Submission Strategies with Competition between a Dealer Market and a Crossing Network," *Journal of Financial Economics*, 91, 319–338.

Du, S. and H. Zhu (2014): "Welfare and Optimal Trading Frequency in Dynamic Double Auctions," Working paper, M.I.T.

Duffie, D. (2010): "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," *Journal of Finance*, 65, 1237–1267.

Dungey, Mardi, Ólan. H. and M. McKenzie (2013): "Modelling Trade Duration in U.S. Treasury Markets," *Quantitative Finance*, 13, 1431–1442.

Fleming, M. and G. Nguyen (2013): "Order Flow Segmentation and the Role of Dark Trading in the Price Discovery of U.S. Treasury Securities," Working paper, Federal Reserve Bank of New York.

Hendershott, T. and H. Mendelson (2000): "Crossing Networks and Dealer Markets: Competition and Performance," *Journal of Finance*, 55, 2071–2115.

Huang, R. D., J. Cai, and X. Wang (2002): "Information-Based Trading in the Treasury Note Interdealer Broker Market," *Journal of Financial Intermediation*, 11, 269–296.

Menkveld, A. J., B. Z. Yueshen, and H. Zhu (2015): "Shades of Darkness: A Pecking Order of Trading Venues," Working paper, M.I.T.

Myerson, R. and P. Reny (2015): "Sequential Equilibria of Multi-Stage Games with Infinite Sets of Types and Actions," Working paper, University of Chicago.

Pancs, R. (2013): "Workup," *Review of Economic Design*, 18, 37–71.

Ready, M. J. (2014): "Determinants of Volume in Dark Pool Crossing Networks," Working paper, University of Wisconsin-Madison.

Rostek, M. and M. Weretka (2012): "Price Inference in Small Markets," *Econometrica*, 80, 687–711.

Tradeweb (2014): "Market Regulation Advisory Notice – Work-Up Protocol," Tech. rep.

Tradition (2015): "Tradition SEF Platform Supplement," Tech. rep.

Vayanos, D. (1999): "Strategic Trading and Welfare in a Dynamic Market," *The Review of Economic Studies*, 66, 219–254.

Vives, X. (2011): "Strategic Supply Function Competition with Private Information," *Econometrica*, 79, 1919–1966.

Zhu, H. (2014): "Do Dark Pools Harm Price Discovery?" *Review of Financial Studies*, 27, 747–789.