NBER WORKING PAPER SERIES

AUGMENTING MARKETS WITH MECHANISMS

Samuel Antill
Darrell Duffie

Working Paper 24146
http://www.nber.org/papers/w24146

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2017

The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research. This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-114747.

Augmenting Markets with Mechanisms
Samuel Antill and Darrell Duffie
NBER Working Paper No. 24146
December 2017
JEL No. D47,D82,G14

ABSTRACT

We compute optimal mechanism designs for each of a sequence of size-discovery sessions, at which traders submit reports of their excess inventories of an asset to a session operator, which allocates transfers of cash and the asset. The mechanism design induces truthful reports of desired trades and perfectly reallocates the asset across traders. Between sessions, in a dynamic auction market, traders strategically lower their price impacts by shading their bids, causing socially costly delays in rebalancing the asset across traders. As the expected frequency of size-discovery sessions is increased, market depth is further lowered, offsetting the efficiency gains of the size-discovery sessions. Adding size-discovery sessions to a double-auction market has no social value, beyond that of an initializing session. If the mechanism design relies on the double-auction market for information from prices, bidding incentives are further weakened, strictly reducing overall market efficiency.

Samuel Antill
Graduate School of Business
Stanford University
Stanford, CA 94305-7298
Santill@stanford.edu

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

An appendix is available at http://www.nber.org/data-appendix/w24146

1 Introduction

In financial markets, investors with large trading interests are concerned about their price-impact costs. Because of this, they execute large orders slowly. This reallocates the asset across traders more gradually than is socially optimal. This concern is exacerbated, under post-crisis regulations, by higher shadow costs of intermediary dealer banks for absorbing large customer orders onto their own balance sheets. Market participants have attempted to lower their price impacts with size-discovery trading protocols, such as workups and dark pools. We show that, at least in our model setting, allocative efficiency cannot be improved by augmenting price-discovery markets with size-discovery sessions, except perhaps for an initializing session. This conclusion applies whether or not size-discovery sessions have an optimal mechanism design.

In each size-discovery session, traders are induced by the mechanism design to truthfully report their excess inventories of an asset to a platform operator, which then allocates transfers of cash and the asset. In equilibrium, each session is ex-post individually rational and incentive compatible, budget balanced, and reallocates the asset perfectly efficiently among traders. Between size-discovery sessions, traders exchange the asset in a sequential double-auction market,[1] modeled on the lines of Du and Zhu (2017).

It is already well understood from the work of Vayanos (1999), Rostek and Weretka (2015), and Du and Zhu (2017) that traders bid less aggressively in a financial market in order to strategically lower their price impacts, causing socially costly delays in rebalancing positions across traders.[2] Duffie and Zhu (2017) showed that a significant fraction of the efficiency loss caused by rebalancing delays in the double-auction market can be avoided by introducing a single, initializing, size-discovery session, before the sequential-double-auction market opens. For this purpose, they analyzed workup, a form of size discovery that is heavily used in dealer-dominated markets, such as those for treasuries and swaps. Duffie and Zhu (2017) also showed that workup is not a fully efficient form of size discovery because traders under-report the sizes of their positions (or equivalently, under-submit trade requests), relative to socially optimal order submissions, due to a winner's-curse effect.

As a mechanism design, the workup protocol places strong restrictions on the allowable forms of messages and transfers. We calculate the optimal mechanism design for size-discovery sessions. In equilibrium, under natural conditions, the optimal mechanism is a new form of size discovery, a direct-revelation scheme that perfectly reallocates the asset among traders. After each size-discovery session, traders' asset inventories are hit by new supply and demand

[1] Each auction is a demand-function submission game, in the sense of Wilson (1979) and Klemperer and Meyer (1989).

[2] Sannikov and Skrzypacz (2016) study a similar setting with heterogeneous traders. They also consider mechanism design, but solely as an analytical device to solve for the equilibrium of a double-auction model.

shocks over time that cause a desire for further rebalancing, which is partially achieved in the double-auction market that runs continually until the next size-discovery session, and so on. For modeling simplicity, the size-discovery sessions are held at Poisson arrival times.

If the mechanism design must rely in part on prior double-auction price information to set the cash-compensation terms, then traders respond strategically in their preceding double-auction order submissions, reducing market depth and strictly reducing overall market efficiency relative to a sequential-double-auction market with no size-discovery sessions (with the possible exception of an initializing size-discovery session).

Even if the mechanism designer has enough information to avoid reliance on preceding double-auction prices, welfare cannot be improved by adding size-discovery sessions. As the expected frequency of size-discovery sessions is increased, the aggressiveness of double-auction market bidding is lowered, precisely offsetting the expected efficiency gains associated with future size-discovery sessions. Traders anticipate the opportunity to lay off excess positions at low cost in the next size-discovery session, and correspondingly lower the aggressiveness of their double-auction bidding.

In summary, adding size-discovery mechanisms to a double-auction market has no social value, with the possible exception of an initializing session, because any allocative benefits of size-discovery sessions are offset, or even dominated, by a corresponding reduction in the depth of price-discovery markets. While one might imagine that this relatively discouraging result is caused by a size-discovery mechanism design that is "too efficient," we show that overall allocative efficiency is not helped by impairing the efficiency of the size-discovery protocol in order to better support market depth and trade volumes in the price-discovery market.

We also discuss some potential implications for the competition for order flow between price-discovery and size-discovery venues, and for potential harm to the price-formation process when size-discovery venues draw sufficiently large volumes of trade away from price-discovery venues, a common point of debate among practitioners and policy makers, and also a point of contention in academic research.[3]

In prior work on mechanism design in dynamic settings, Bergemann and Välimäki (2010) show that a generalization of the Vickrey-Clarke-Groves pivot mechanism can implement efficient allocations in dynamic settings with independent private values.[4] Similarly, Athey and Segal (2013) and Pavan, Segal, and Toikka (2014) study optimal mechanism designs in dynamic settings with independent types. As opposed to this prior research, we focus on a market setting in which agents cannot be contractually obligated[5] to participate in mechanisms or to abstain

[3]See, for example, CFA Institute (2012) and the discussions of Zhu (2014) and Ye (2016).

[4]In unreported results, and prompted by correspondence with Romans Pancs, we find that such a mechanism also implements an efficient allocation in the primitive stochastic setting of our model.

[5]Specifically, we always impose an ex-post participation condition that, at every mechanism session, all

from trading in alternative venues.

Dworczak (2017) precedes this paper in considering a mechanism design problem in which the designer cannot prevent agents from participating in a separate market.[6] Beyond that likeness of perspective, the problems addressed by our respective models are quite different. Ollár, Rostek, and Yoon (2017) address a design problem associated with double-auction markets, but focus instead on information revelation within the market, rather than an augmentation of the double-auction market with mechanism-based sessions. Du and Zhu (2017) considered the optimal frequency of double-auctions, as an alternative design approach to reducing allocative inefficiencies associated with the strategic avoidance of price impact. Pancs (2014) analyzed the implications of workup for its ability to mitigate front-running.[7]

2 Static Mechanism Design

This section models a static mechanism-design problem in which a designer, say a trade platform operator, elicits reports from each of $n \geq 3$ traders about their asset positions, and based on those reports makes cash and asset transfers.

For trader i, the initial quantity z_0^i of assets is a finite-variance random variable[8] that is privately observable, meaning that z_0^i is measurable with respect to the information set $\mathcal{F}^i$ of trader i. The aggregate inventory $Z \equiv \sum_{i=1}^n z_0^i$ of assets is also observable to all traders and to the platform operator. For example, Z could be deterministic. We relax the observability of Z in Section 5.

A report from trader i is a random variable $\hat{z}^i$ that is measurable with respect to the information set of trader i. Given a list $\hat{z} = (\hat{z}^1, \ldots, \hat{z}^n)$ of trader reports, a reallocation is a list $y = (y^1, \ldots, y^n)$ of finite-variance random variables that is measurable with respect to[9] $\{Z, \hat{z}\}$ and satisfies $\sum_i y^i = 0$.

traders prefer participation to the outside option of not entering this mechanism and trading in a double-auction market until the next mechanism. In contrast, Pavan, Segal, and Toikka (2014) force agents to commit at time zero to participate in all future mechanisms (or post an arbitrarily large bond to be forfeited in the event of exit), and Bergemann and Välimäki (2010) force agents to forgo all future mechanism participation in order to sit out one mechanism event. Athey and Segal (2013) provide conditions under which efficient allocations can be reached without participation constraints, but only if agents are arbitrarily patient relative to the most extreme (finite) realization of uncertainty.

[6] In a macroeconomic setting, Di Tella (2017) and Di Tella, Sannikov et al. (2016) consider mechanism design problems in which principals cannot stop intermediaries from stealing their funds through "hidden trade." We focus on market inefficiencies rather than agency problems between households and intermediaries.

[7] The seller in Panc's model has private information about the size of his or her desired trade. The buyer is either a "front-runner" or a dealer. If the seller cannot sell the entire large position in workup, he would need to liquidate the remainder by relying on an exogenously given outside demand curve.

[8] Fixing a probability space $(\Omega, \mathcal{F}, P)$, trader i has information represented by a sub-σ-algebra $\mathcal{F}^i$ of $\mathcal{F}$. That is, trader i is initially informed of any random variable that is measurable with respect to $\mathcal{F}^i$.

[9] That is, z is measurable with respect to the sub-σ-algebra of $\mathcal{F}$ generated $\{\hat{z}, Z\}$.

Anticipating the form of post-mechanism indirect utility for the equilibrium of our eventual model of a dynamic market, we assume that the value to trader i of a given reallocation y is $\mathbb{E}[V^i(z_0^i + y^i, Z) \mid \mathcal{F}^i]$, where

$$V^i(z^i, Z) = u^i(Z) + \left(\beta_0 + \beta_1 \overline{Z}\right)\left(z^i - \overline{Z}\right) - K\left(z^i - \overline{Z}\right)^2, \tag{1}$$

where $u^i : \mathbb{R} \to \mathbb{R}$ is a real-valued measurable function to be specified such that $u^i(Z)$ has a finite expectation, $\overline{Z} \equiv Z/n$, and β_0, β_1, and K are real numbers, with $K > 0$, that do not depend on i.

A reallocation is welfare maximizing given a list $\hat{z}$ of reports if it solves

$$\sup_{y \,\in\, \mathcal{Y}(\hat{z}, Z)} \mathbb{E}\left[\sum_{i=1}^{n} V^i(z_0^i + y^i, Z)\right],$$

where $\mathcal{Y}(\hat{z}, Z)$ is the set of reallocations. A reallocation is said to be perfect if it is optimal for the case in which the reports are perfectly revealing,[10] for example when $\hat{z}^i = z_0^i$. From the quadratic costs of asset dispersion across traders reflected in the last term of $V^i(z^i, Z)$, it is immediate that a reallocation y is perfect if and only if $z_0^i + y^i = \overline{Z}$ for all i.

We will now calculate a mechanism design that achieves a perfect reallocation. Specifically, a mechanism is a function that maps Z and a list $\hat{z}$ of reports to a reallocation denoted $Y(\hat{z}) = (Y^1(\hat{z}), \ldots, Y^n(\hat{z}))$ and a list $T(\hat{z}, Z) = (T^1(\hat{z}, Z), T^2(\hat{z}, Z), \ldots, T^n(\hat{z}, Z))$ of real-valued "cash" transfers with finite expectations. In the game induced by a mechanism (Y, T), $\hat{z}$ is an equilibrium if, for each trader i, the report $\hat{z}^i$ solves

$$\sup_{\tilde{z}} \; U^i((\tilde{z}, \hat{z}^{-i})),$$

where, for any list $\hat{z}$ of reports,

$$U^i(\hat{z}) = \mathbb{E}\left[V^i(z_0^i + Y^i(\hat{z}), Z) + T^i(\hat{z}, Z) \mid \mathcal{F}^i\right], \tag{2}$$

and where we adopt the standard notation by which for any $x \in \mathbb{R}^n$ and $w \in \mathbb{R}$,

$$(w, x^{-i}) \equiv (x^1, x^2, \ldots, x^{i-1}, w, x^{i+1}, \ldots, x^n).$$

In words, each trader i takes the strategies of the other traders as given and chooses a report $\hat{z}^i$ depending only on the information available to trader i that maximizes the conditional expected sum of the reallocated asset valuation and the cash transfer.

[10] A report $\hat{z}^i$ from trader i is perfectly revealing if z_0^i is measurable with respect to $\{Z, \hat{z}^i\}$.

For any constant $\kappa_0 < 0$ and any Lipschitz-continuous functions $\kappa_1 : \mathbb{R} \to \mathbb{R}$ and $\kappa_2 : \mathbb{R} \to \mathbb{R}$ of the commonly observed aggregate inventory Z, we will consider the properties of the mechanism $\mathcal{M}^\kappa$ defined by the asset reallocation

$$Y^i(\hat{z}) = \frac{\sum_{j=1}^n \hat{z}^j}{n} - \hat{z}^i \tag{3}$$

and the cash transfer

$$T^i_\kappa(\hat{z}, Z) = \kappa_0 \left(n\, \kappa_2(Z) + \sum_{j=1}^n \hat{z}^j \right)^2 + \kappa_1(Z)(\hat{z}^i + \kappa_2(Z)) + \frac{\kappa_1^2(Z)}{4\kappa_0 n^2}. \tag{4}$$

The second term of (4) is analogous to compensation at a fixed marginal price of $\kappa_1(Z)$. This is the essential feature of size-discovery mechanisms, such as a dark pools, workups, and matching sessions, which is to freeze the price and thus eliminate the adverse effect of price-impact.[11] Going beyond typical versions of size discovery that have been used in practice, however, the first term of (4) forces trader i to internalize some of quadratic cost of an uneven cross-sectional distribution of the asset. The final term in (4) can be viewed as a fixed participation fee, which ensures that the platform operator does not lose money. That is, for any list $\hat{z}$ of reports, the mechanism $\mathcal{M}^\kappa$ always leaves a weakly positive profit for the platform operator because $\sum_i T^i_\kappa(\hat{z}, Z) \leq 0$.

The following proposition, proven in the appendix, provides an equilibrium of the mechanism report game. The proposition also shows that for a carefully chosen κ_0, each trader can actually ignore the reports of other traders.

Proposition 1. *Consider a mechanism of the form $\mathcal{M}^\kappa$, defined by any $\kappa_0 < 0$, and any Lipschitz-continuous $\kappa_1(\,\cdot\,)$ and $\kappa_2(\,\cdot\,)$.*

1. *Suppose trader i anticipates that, for each $j \neq i$, trader j will submit the report $\hat{z}^j = z_0^j$. There is a unique solution to the optimal report problem for trader i induced by the mechanism $\mathcal{M}^\kappa$. This solution is $\hat{z}^i = z_0^i$ almost surely, if and only if*

$$\kappa_2(Z) = -\overline{Z} + \frac{-\kappa_1(Z) + \left(\frac{n-1}{n}\right)\left(\beta_0 + \beta_1 \overline{Z}\right)}{2\kappa_0 n}. \tag{5}$$

[11] Not all dark pools are designed primarily for the purpose of mitigating price impacts for large orders. Drawing from an industry report by Rosenblatt Securities, Ye (2016) notes that "In May 2015, among the 40 active dark pools operating in the US, there are 5 dark pools in which over 50% of their Average Daily Volumes are block volume (larger than 10k per trade). Those pools can be regarded as "Institutional dark pools," and they include Liquidnet Negotiated, Barclays Directx, Citi Liquifi, Liquidnet H20, Instinet VWAP Cross, and BIDS Trading." Other objectives of dark pool users include a reduction in the leakage of private information motivating trade, and the avoidance of bid-ask spread costs. Some broker-dealers use their own dark pools to internalize order executions among their clients.

That is, $\mathcal{M}^\kappa$ is a direct revelation mechanism if and only if $\kappa_2(Z)$ is given by (5).

2. *Suppose $\kappa_2(Z)$ is given by (5). If trader i anticipates the report $\hat{z}^j = z_0^j$ for each $j \neq i$, then the truthful report $z^{*i} = z_0^i$ is ex-post optimal, that is, optimal whether or not we take the special case in which trader i observes[12] z_0^{-i}.*

3. *For the list $z^* = (z^{*1}, \ldots, z^{*n})$ of such truthful reports, the reallocation $Y(z^*)$ of (3) is perfect. That is, $z_0^i + Y^i(z^*) = \overline{Z}$ for all i.*

4. *For any $\kappa_1(\,\cdot\,)$, for $\kappa_2(Z)$ given by (5), and for $\kappa_0 = -K(n-1)/n^2$, the mechanism $\mathcal{M}^\kappa$ is strategy proof. That is, the truthful report $z^{*i} = z_0^i$ is a dominant strategy, being an optimal report for trader i regardless of the conjecture by trader i of the reports $\hat{z}^{-i}$ of the other traders.*

The ex-post optimality property stated in the proposition is in the spirit of Du and Zhu (2017), although for a much different market game. In particular, it is a Nash equilibrium[13] of the complete information game (in which all traders know z_0) for traders to submit the list z^* of reports. For the special case $\kappa_0 = -K(n-1)/n^2$, this is the unique Nash equilibrium because, for any trader i, the report z^{*i} is a dominant strategy and because of the strict concavity of $U^i((\tilde{z}, \hat{z}^{-i}))$ with respect to $\tilde{z}$.

We have not yet considered whether trader i could do better by not entering the mechanism at all. From this point, we always fix κ_2 as specified by (5).[14] For arbitrary κ_0 and $\kappa_1(\,\cdot\,)$, the mechanism $\mathcal{M}^\kappa$ need not be ex-post individually rational. That is, there could be realizations of (z_0^i, Z) at which trader i would strictly prefer $V^i(z_0^i, Z)$ over the expected equilibrium value to trader i. However, because the platform operator observes Z, he or she can choose $\kappa_1(Z)$ so as to ensure that all traders strictly prefer to participate in the mechanism, except in the trivial case in which the initial allocation is already perfect.

Proposition 2. *Fix κ_2 as in (5), let $\kappa_1(Z) = \beta_0 + \beta_1 \overline{Z}$, and let κ_0 be arbitrary. For the equilibrium reports z^* of the mechanism $\mathcal{M}^\kappa$, we have*

$$U^i(z^*) = V^i(z_0^i, Z) + K\left(z_0^i - \overline{Z}\right)^2. \tag{6}$$

With probability one, trader i weakly prefers this equilibrium value to the value $V(z_0^i, Z)$ of the

[12] To be able to observe z_0^{-i} means that z_0^{-i} is measurable with respect to $\mathcal{F}^i$.

[13] Likewise, this is also a Bayesian Nash equilibrium of the incomplete information game, after specifying beliefs about other traders' inventories.

[14] By the Revelation Principle (Myerson (1981)), it is natural to focus on direct-revelation mechanisms.

initial inventory z_0^i. *That is,*

$$U^i(z^*) = V^i(z_0^i + Y^i(z^*), Z) + T_\kappa^i(z^*, Z) \geq V^i(z_0^i, Z).$$

The inequality is strict unless $z_0^i = \overline{Z}$. *Provided that the probability distribution of* z_0 *has full support, this inequality holds with probability one if and only if* $\kappa_1(Z) = \beta_0 + \beta_1 \overline{Z}$.

A proof is found in the appendix. In summary, if the aggregate inventory Z is known to all traders and to the size-discovery platform operator, then the budget-balanced mechanism $\mathcal{M}^\kappa$ can implement a perfect reallocation in an ex-post individually rational equilibrium.[15] Proposition 2 also implies that the equilibrium payoffs do not depend upon the choice of κ_0. For $\kappa_1(\,\cdot\,)$ and $\kappa_2(\,\cdot\,)$ as specified in Proposition 2, some algebra shows that the equilibrium cash transfer to trader i is

$$\kappa_1(Z)\left(z_0^i - \overline{Z}\right) = \left(\beta_0 + \beta_1 \overline{Z}\right)\left(z_0^i - \overline{Z}\right). \tag{7}$$

The mechanism designer is thus free to choose any $\kappa_0 < 0$, because the choice of κ_0 has no impact on equilibrium transfers or allocations. Result 4 of Proposition 1 nevertheless indicates the strategy-proofness advantage of the particular choice $\kappa_0 = -K(n-1)/n^2$.

Figure 1 illustrates the cash and asset transfers that are obtainable by trader i for the mechanism of Proposition 2, when other traders follow the equilibrium report z^{*j}. The asset transfer schedule $\hat{z}^i \mapsto Y(\hat{z})$ is linear. The cash transfer schedule $\hat{z}^i \mapsto T_\kappa^i(\hat{z}, Z)$ can be close to linear, similar to the case of size-discovery mechanisms such as workups and dark pools. However, a report by trader i that is large in magnitude induces a significant cash penalty associated with the quadratic component of the cash transfer schedule. From a welfare viewpoint, this penalty appropriately disciplines trader i from over-exploiting the mechanism by trying to completely eliminate his or her excess inventory. A workup or dark pool handles this problem of disciplining demand and supply by rationing whichever side of the market has a greater absolute magnitude of excess inventory. Workup rations by time prioritization of orders (first come, first served). A typical dark pool rations the heavier side of the market pro rata to requested trade sizes. These rationing schemes, however, are only rules of thumb, and are strictly suboptimal. The mechanism $\mathcal{M}^\kappa$ of Proposition 2, on the other hand, achieves the first best.

As mentioned previously, a linear-quadratic utility of the form $V^i(z, Z)$ emerges in the next section as the equilibrium continuation value in the sequential double-auction market, even if

[15]As noted to one of us by Romans Pancs, a Vickrey-Clarke-Groves (VCG) pivot mechanism can also implement a perfect reallocation in an ex-post equilibrium in this setting. However, the standard pivot mechanism cannot be both budget balanced and ex-post individually rational. The AGV mechanism of Arrow (1979), d'Aspremont and Gérard-Varet (1979) does not apply to this setting because the private information of traders is correlated.

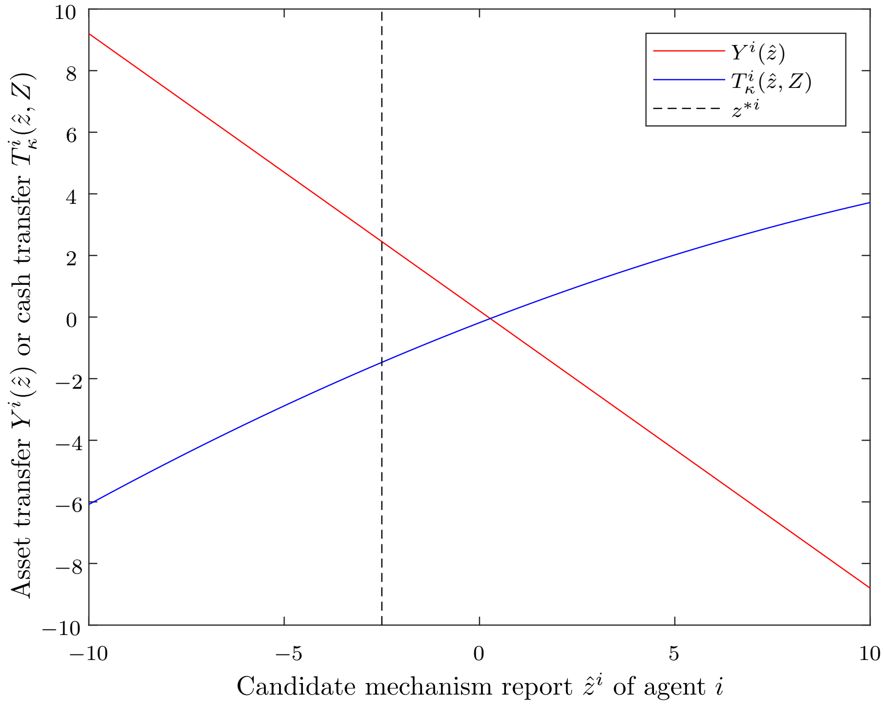

Figure 1: Mechanism Transfers and Reallocations. This figure plots the possible transfers and reallocations available in the mechanism for a trader, in an equilibrium. The parameters are $v = 0.5$, $r = 0.1$, $n = 10$, $\gamma = 0.1$, $Z = -0.5$, and $z^i_0 = -2.5$. The value function $V(\cdot)$ corresponds to the continuation value for the subsequent double-auction market equilibrium, so that $\beta_0 = v$, $\beta_1 = -2\gamma/r$, and $K = \gamma/[r(n-1)]$. We take $\kappa_0 = -K(n-1)/n^2$, $\kappa_1(Z) = \beta_0 + \beta_1\overline{Z}$, and $\kappa_2(Z)$ defined as in Proposition 1. The report of each of the other nine traders is fixed at the equilibrium level z^{*j}.

the market is augmented with future reallocation sessions. Proposition 1 therefore implies that if our mechanism is run at time 0, before the market opens, then all traders will instantly move to the socially efficient allocation. However, as traders receive subsequent inventory shocks over time, their allocation becomes inefficient, leaving some scope for later improvements in the allocations. This is the central issue addressed by this paper.

3 The Welfare Cost of Price-Impact Avoidance

In this section, we model a sequential double-auction market in which traders strategically avoid price impact, causing a socially inefficient delay in the re-balancing of asset positions across agents. This issue is well covered by the results of Vayanos (1999), Rostek and Weretka (2015), Du and Zhu (2017), and Duffie and Zhu (2017). However, for our later purpose of exploring the augmentation of a sequential double-auction market with a sequence of size-discovery sessions, we develop in this section a suitable generalization of the continuous-time double-auction model of Duffie and Zhu (2017).

The continuous-time presentation of our results is chosen for its expositional simplicity. A discrete-time analogue of our model is found in the appendix. While the discrete-time setting leads to messier looking results, it allows us to demonstrate a standard equilibrium robustness property, Perfect Bayes. The equilibrium behavior of the discrete-model converges to that of the continuous-time model as the length of a time period shrinks to zero.

We fix a probability space, the time domain $[0, \infty)$, and an information filtration $\mathbb{F} = \{\mathcal{F}_t : t \geq 0\}$ satisfying the usual conditions.[16] The market is populated by $n \geq 3$ risk-neutral agents trading a divisible asset. The payoff π of the asset is a bounded random variable with mean v. The payoff π is revealed publicly and paid to traders at a random time $\mathcal{T}$ that is exponentially distributed with parameter r. Thus $\mathbb{E}(\mathcal{T}) = 1/r$. There is no further incentive to trade once π is revealed at time $\mathcal{T}$, which is therefore the ending time of the model.

Trader i has information given by a sub-filtration $\mathbb{F}^i = \{\mathcal{F}_t^i : t \geq 0\}$ of $\mathbb{F}$. The traders have symmetric information about the asset payoff. Specifically, we suppose that the conditional distribution of π given $\mathcal{F}_t$ is constant until the payoff time $\mathcal{T}$, so that no trader ever learns anything about π until the market ends. The traders may, however, have asymmetric information about their respective asset positions at each time. Price fluctuations are thus driven only by allocative concerns, and not by learning about ultimate asset payoffs. This informational setting is more relevant for markets such as those for stock index products, major currencies, and fixed income products such as swaps and government bonds. For example, there is always symmetric information about the payoff of a treasury bill, but the price of a treasury bill fluctuates randomly over time, partly caused by shocks to the allocation of the T-bills across market participants.

The initial inventories of the asset for the n traders are specified as in Section 2 by a list $z_0 = (z_0^1, z_0^2, \ldots, z_0^n)$ of finite-variance random variables, with z_0^i measurable with respect to $\mathcal{F}_0^i$.

In a continually operating double-auction market, at each time t, trader i submits an $\mathcal{F}_t^i$-measurable demand function $\mathcal{D}_t^i : \Omega \times \mathbb{R} \to \mathbb{R}$. Thus, in state ω at time t, the trader would buy the asset at the quantity "flow" rate $\mathcal{D}_t^i(\omega, p)$ if the auction price p is chosen. Given a double-auction price process ϕ, trader i would thus purchase the total quantity $\int_s^u \mathcal{D}_t^i(\omega, \phi_t(\omega))\, dt$ of the asset over some time interval $[s, u]$ (assuming the integral exists). We only consider equilibria in which demand functions are of the affine form

$$\mathcal{D}_t^i(\omega, p) = a + bp + cz_t^i(\omega), \tag{8}$$

[16]For the "usual conditions" on a filtration see, for example, Protter (2005).

for constants a, $b < 0$, and c that do not depend on i or t, and where z_t^i is the quantity of the asset held by trader i at time t. To be clear, the traders are not restricted to affine demand functions, but in equilibrium we will show that each trader optimally chooses a demand function that is affine if he or she assumes that the other traders do so.

At time t, given the demand-function coefficients (a, b, c) and the current list $z_t = (z_t^1, \ldots, z_t^n)$ of trader inventories, a price ϕ_t is chosen by a trade platform operator to clear the market. A complete equilibrium model of the demand coefficients (a, b, c) and of the evolution of the inventory processes $(z^1, \ldots, z^n)$ will be provided shortly.

Lemma 1. *Fix any demand-function coefficients (a, b, c) with $b < 0$, some time t, and some trader i. For any candidate demand $d \in \mathbb{R}$ by trader i, there is a unique price p with $d + \sum_{j \neq i}(a + bp + cz_t^j) = 0$. This clearing price is calculated as*

$$p = \Phi_{(a,b,c)}(d; Z_t^{-i}) \equiv \frac{-1}{b(n-1)} \left(d + (n-1)a + cZ_t^{-i}\right), \tag{9}$$

where $Z_t^{-i} = \sum_{j \neq i} z_t^j$.

Thus, for any non-degenerate affine demand function used by $n - 1$ of the traders, there is a unique market clearing price for each quantity chosen by the remaining trader.

The asset inventory of trader i is randomly shocked over time with additional units of the asset. The cumulative shock to the inventory of trader i by time t is H_t^i, for some finite-variance Lévy process H^i that is a martingale with respect to $\mathbb{F}$ and thus with respect to the information filtration $\mathbb{F}^i$ of trader i. A simple example of H^i is an $\mathbb{F}$-Brownian motion with zero drift. The defining property of a Lévy process is that it has independent increments and identically distributed increments over any equally long time intervals. Without loss of generality, we take $H_0^i = 0$. The inventory shock processes $H = (H^1, \ldots, H^n)$ need not be independent across traders, but we assume that H is independent of $\{\mathcal{T}, \pi, z_0\}$ and that $\sum_i H^i$ is also a Lévy process.

Letting $\sigma_i^2 \equiv \text{var}(H_1^i)$, the Lévy property[17] implies that for any time t we have $\text{var}(H_t^i) = \sigma_i^2 t$. Likewise, letting $\sigma_Z^2 = \text{var}\left(\sum_i H_1^i\right)$ and $\rho^i = \text{cov}(Z_1, H_1^i)$, the Lévy property implies that $\text{var}(Z_t) = \text{var}(Z_0) + \sigma_Z^2 t$ and that $\text{cov}(Z_t, H_t^i) = \rho^i t$ for some constant ρ^i.

Traders suffer costs associated with unwanted levels of inventory, whether too large or too small. One may think in terms of a market maker that is attempting to run a matched book of positions, but which may accept customer positions over time that shock its inventory. The market maker may then trade so as to lay off excess inventories with other market makers in an inter-dealer double-auction market.

[17] Because H^i is a finite-variance process, its characteristic exponent $\psi_i(\,\cdot\,)$ has two continuous derivatives, and $\sigma_i^2 = \psi_i''(0)$. As an example, if H^i is a Brownian motion with variance parameter φ, then $\sigma_i^2 = \varphi$.

The market practitioners Almgren and Chriss (2001) proposed a simple model of inventory costs for financial firms that is now popular among other practitioners and also in the related academic research literature, by which the rate of inventory cost to trader i at time t is $\gamma(z_t^i)^2$, for some coefficient $\gamma > 0$. Here, we have normalized so that the inventory level z_t^i is measured net of the desired inventory level. With this model, trader i perceives, at any time t, an expected total cost of future undesired inventory of

$$\mathbb{E}\left[\int_t^{\mathcal{T}} -\gamma(z_s^i)^2\,ds \;\middle|\; \mathcal{F}_t^i\right].$$

Although financial firms do not have direct aversion to risk, broker-dealers and asset-management firms do have extra costs for holding inventory in illiquid or risky assets. These costs can be related to regulatory capital requirements, collateral requirements, financing costs, agency costs associated with a lack of transparency of the position to higher-level firm managers or clients regarding the true asset quality, as well as the expected cost of being forced to suddenly raise liquidity by quickly disposing of remaining inventory into an illiquid market. Although it has not been given a structural foundation, the quadratic holding-cost assumption is common in dynamic market-design models, including those of Vives (2011), Rostek and Weretka (2012), Du and Zhu (2017), and Sannikov and Skrzypacz (2016).

Lemma 1 allows any given trader i to simplify his or her strategic bidding problem to the selection of a real-valued demand process D^i, which then determines the market clearing price process $\Phi_{(a,b,c)}(D_t^i; Z_t - z_t^i)$. A demand process D^i is optimal for trader i given the demand coefficients (a, b, c) of the other traders if D^i solves the stochastic control problem of optimizing expected net profits, defined by

$$V^i(z_0^i, Z) \equiv \sup_{D\,\in\,\mathcal{A}^i}\; \mathbb{E}\left[z_{\mathcal{T}}^D\pi - \int_0^{\mathcal{T}} \gamma\left(z_s^D\right)^2 + \Phi_{(a,b,c)}\left(D_s; Z_s - z_s^D\right) D_s\,ds \;\middle|\; \mathcal{F}_0^i\right], \tag{10}$$

where $\mathcal{A}^i$ is the space of integrable $\mathbb{F}^i$-adapted processes such that the expectation in (10) exists, and where

$$z_t^D = z_0^i + \int_0^t D_s\,ds + H_t^i. \tag{11}$$

The total expected profit (10) is finite or negative infinity for any demand process D, and is finite at any optimum demand process, given that $D = 0$ is a candidate demand process.

Demand coefficients (a, b, c) with $b < 0$ are said to constitute a symmetric affine equilibrium if, for any trader i, given (a, b, c), the demand process $D_t^i = a + b\phi_t + cz_t^i$ is optimal, where ϕ_t is the market clearing price process

$$\phi_t = \frac{a + c\overline{Z}_t}{-b},$$

where $\bar{Z}_t = Z_t/n$ and z^i solves the stochastic differential equation

$$z_t^i = z_0^i + \int_0^t (a + b\phi_s + cz_s^i)\, ds + H_t^i.$$

This definition of equilibrium implies market clearing, individual trader optimality given the assumed demand functions of other traders, and consistent conjectures about the demand functions used by other traders. This notion of equilibrium was developed by Du and Zhu (2017), who emphasized that the equilibrium demands are ex-post optimal. That is, no trader would bid differently even if he or she were able to observe the inventories of all other traders.

Although we are working here for expositional simplicity in a continuous-time setting, the equilibria that we propose may safely be considered to be Perfect Bayesian Equilibrium. That is, in light of the ex-post optimality property, beliefs about other traders' inventories are irrelevant. This is tied down rigorously in a discrete-time analogue of our model found in the appendix. In discrete time, the ex-post optimality property implies subgame perfection for the complete information game. Moreover, the primitive parameters of the discrete-time model and the associated discrete-time equilibrium bidding behavior converge to those for the continuous-time model as the length of a time interval shrinks to zero. This convergence was shown by Duffie and Zhu (2017) for a simpler version of this model, and applies also in the current setting.

A proof of the following proposition appears in the appendix.

Proposition 3. *There is a unique symmetric affine equilibrium. The equilibrium market-clearing price process is*

$$\phi_t = v - \frac{2\gamma}{r}\overline{Z}_t. \tag{12}$$

In this equilibrium, for any trader i and any time t, the indirect utility of trader i defined by (10) is

$$V^i(z_t^i, Z_t) = \theta_i + v\overline{Z}_t - \frac{\gamma}{r}\overline{Z}_t^2 + \phi_t\left(z_t^i - \overline{Z}_t\right) - \frac{\gamma}{r}\frac{1}{n-1}\left(z_t^i - \overline{Z}_t\right)^2, \tag{13}$$

where

$$\theta_i = \frac{\gamma\sigma_Z^2}{r^2n^2} - \frac{\gamma}{r^2(n-1)}\left(\frac{\sigma_Z^2}{n^2} + \sigma_i^2 - 2\frac{\rho^i}{n}\right) - \frac{2\gamma\rho^i}{r^2n}.$$

The equilibrium demand function of any trader i evaluated at an arbitrary price p, state ω, and time t is

$$\mathcal{D}_t^i(\omega, p) = \frac{(n-2)r^2}{4\gamma}\left(v - p - \frac{2\gamma}{r}z_t^i(\omega)\right). \tag{14}$$

That is, the equilibrium demand function is affine with coefficients

$$a = \frac{(n-2)r^2 v}{4\gamma}, \qquad b = \frac{-(n-2)r^2}{4\gamma}, \qquad c = \frac{-(n-2)r}{2}. \tag{15}$$

We can now define the equilibrium welfare, given the initial list z_0 of positions, as

$$W(z_0) \equiv \sum_{i=1}^{n} V^i(z_0^i, Z_0) = \sum_i \theta_i + vZ_0 - \frac{\gamma}{r}\frac{Z_0^2}{n} - \frac{\gamma}{r(n-1)}\sum_{i=1}^{n}\left(z_0^i - \overline{Z}_0\right)^2. \tag{16}$$

An additive welfare function is appropriate for market efficiency considerations because our traders are maximizing total expected profits net of costs, measured in "dollar" values.

A social planner who is free to reallocate inventories among the n traders can obviously improve on this welfare $W(z_0)$, except in the unique trivial case in which the initial total inventory is equally split across traders (that is, $z_0^i = \overline{Z}_0$ for all i) and in which there are symmetric future inventory shocks ($H^i = H^j$ for all i, j, almost surely). By constantly reallocating inventories so as to keep $z_t^i = \overline{Z}_t$, a social planner can achieve the first-best welfare of

$$W_{fb}(Z_0) = -\frac{\gamma}{r^2}\frac{\sigma_Z^2}{n} + vZ_0 - \frac{\gamma}{r}\frac{Z_0^2}{n}. \tag{17}$$

Relative to first best, the equilibrium behavior of Proposition 3 is inefficient because each trader strategically bids so as to reduce the price impact associated with the dependence of the clearing price $\Phi_{(a,b,c)}(D_t; Z_t - z_t^i)$ on his or her demand D_t. In order to reduce a costly inventory imbalance more rapidly, the trader would suffer a bigger price impact. In light of this, the trader reduces the sizes of orders, trading off price impact against inventory costs. But price impacts are mere wealth transfers, and have no direct social costs. It is not socially efficient for traders to internalize their price-impact costs. In this paper, we are mainly interested in how this loss of welfare might be mitigated with size-discovery sessions, such as workup or the optimal reallocation sessions described in the previous section, at which there are no price impacts. In our setting, social welfare is determined entirely by total expected inventory costs. To repeat, the welfare inefficiency of strategic avoidance of price impact is well covered by the prior results of Vayanos (1999), Rostek and Weretka (2015), and Du and Zhu (2017).

4 Augmenting Price Discovery with Size Discovery

An obvious improvement in welfare is obtained by an initializing size-discovery session. For example, Duffie and Zhu (2017) showed a significant improvement in welfare associated with running a workup session at time zero, before the sequential double-auction market opens.

Workup does not optimally reallocate initial inventory. We showed in Section 2 that running an optimal mechanism at time zero achieves a perfect initial allocation, after which all traders have the same inventory $\overline{Z}_0$. If no further size-discovery reallocation sessions are run, so that after the market opens traders rely entirely on the sequential double-auction market, then the corresponding welfare is

$$W^*(Z_0) \equiv W_{fb}(Z_0) + \frac{\gamma\sigma_Z^2}{r^2 n} + \sum_i \theta_i. \tag{18}$$

A direct calculation[18] then shows that

$$W^*(Z_0) \leq W_{fb}(Z_0), \tag{19}$$

with strict inequality unless $H^i = H^j$ for all i, j. The negative constant $\sum_i \theta_i$ reflects the aggregate costs to all traders of future random inventory shocks that are only slowly rebalanced in the subsequent sequential double-auction market.

Somewhat surprisingly, we are about to show that welfare is not improved by adding optimal-mechanism reallocation sessions after time zero, even though the traders' inventories are perfectly reallocated at each of these sessions. In the following section, we will show that augmenting the market with perfect reallocation sessions *strictly lowers* welfare if the size-discovery platform operator cannot directly observe the evolution of the aggregate inventory. This welfare loss is caused by bidding behavior that attempts to strategically distort the platform operator's inference of the current inventory Z_t from observing prior double-auction prices.

In this section, the aggregate inventory Z_t is assumed to be observable by the size-discovery mechanism operator. Later, we relax the assumption of observable aggregate inventory in order to analyze the adverse welfare impact of bidding that is designed to strategically influence the inference of the size-discovery platform operator, who will rely on double-auction prices for inference regarding the aggregate inventory.

We maintain the model setup of the previous section, with one exception. We now add a sequence of size-discovery sessions, each of which uses the perfect-reallocation mechanism developed in Section 2. These sessions occur at the event times $\tau_1, \tau_2, \ldots$ of a commonly observable Poisson process N with mean arrival rate $\lambda > 0$. The session-timing process N is

[18] Rearranging terms, we have

$$\theta_i = \frac{\gamma(n-2)}{r^2(n-1)}\mathrm{var}(\bar{Z}_1 - H_1^i | Z_0) - \frac{\gamma}{r^2}\sigma_i^2.$$

We note that $\sum_i \mathrm{var}(\bar{Z}_1 - H_1^i|Z_0) = -n\mathrm{var}(\bar{Z}_1|Z_0) + \sum_i \mathrm{var}(H_1^i)$. The inequality follows from the fact that $n\sum_i \mathrm{var}(H_1^i) \geq \mathrm{var}(\sum_i H_1^i)$, with equality if and only if $H^i = H^j$ for all i, j .

independent of the other primitive processes and random variables, $\{H, \mathcal{T}, \pi, z_0\}$.

In practice, the mean frequency of size-discovery sessions varies significantly across markets. For example, workup sessions in BrokerTec's market for treasury securities occur at an average frequency of about 600 times a day for the 2-year note, and about 1400 times a day for the 5-year note, according to statistics provided by Fleming and Nguyen (2015). These size-discovery sessions account for approximately half of all trade volume in treasury securities on BrokerTec, which is by far the largest trade platform for U.S. treasuries, accounting for an average of over \$30 billion in daily transactions for each of the 2-year, 5-year, and 10-year on-the-run treasury notes. Consistent with our model, BrokerTec workup sessions are held at randomly spaced times. As opposed to our model, however, the times of BrokerTec workup sessions are chosen directly by market participants, rather than at exogenous random times. In the corporate bond market, "matching sessions," another form of size-discovery, occur with much lower frequency, such as once per week for some bonds. The matching sessions on Electronifie, a corporate bond trade platform, are triggered automatically by an algorithm that depends on the current limit order book and the unfilled portion of the last trade on the central limit order book. Again, this differs from our simplifying assumption that size-discovery reallocation sessions occur at independent exogenously chosen times.

In many designs for size-discovery sessions, and in the setting of the next section of our paper, the platform operator exploits prior market prices as a guide to (or automatic determinant of) the "frozen price" used in the size-discovery session. This introduces additional incentive effects that we consider in the next section. In this section, because the aggregate inventory Z is observable, the size-discovery platform operator does not need to rely on prior double-auction market prices to set the mechanism's cash compensation rates.

In addition to choosing a double-auction market demand process D^i, as modeled in the previous section, trader i also chooses an $\mathbb{F}^i$-adapted and jointly measurable[19] process $\hat{z}^i$ for mechanism reports.

Our size-discovery sessions will use the mechanism design (Y, T_κ) of Section 2, restricting attention to the affine functions $\kappa_1(\,\cdot\,)$ and $\kappa_2(\,\cdot\,)$ of Z_t that exploit the properties of Propositions 1 and 2. We will calculate intercept and slope coefficients of both κ_1 and κ_2 that are consistent with the resulting endogenous continuation value functions.

We will show that the double-auction equilibrium demand behavior in this new setting is of the same affine form that we found in the market without reallocation sessions, however with different demand coefficients. The traders' demands are altered by the prospect of getting a perfectly re-balanced allocation at the next size-discovery session.

In equilibrium, the demand process D^i of trader i and the vector $\hat{z}$ of report processes of all

[19]For the formal definition of adapted, please refer to Protter (2005).

traders imply that the inventory process of trader i is

$$z_t^i = z_0^i + \int_0^t D_s^i \, ds + H_t^i + \int_0^t \left(\frac{\sum_{j=1}^n \hat{z}_s^j}{n} - \hat{z}_s^i \right) dN_s. \tag{20}$$

Given the direct-revelation mechanism design (Y, T_κ) for the size-discovery sessions, an equilibrium of the associated dynamic demand and reporting game (involving symmetric affine demand functions) consists of demand coefficients (a, b, c), with the properties:

A. If each trader i assumes that each other trader j uses these demand coefficients and truthfully report the position $\hat{z}_t^j = z_t^j$ for the purposes of size-discovery sessions, then trader i optimally uses the same affine demand function coefficients (a, b, c) and also reports truthfully.

B. Participation in the size-discovery sessions is individually rational. Specifically, given the equilibrium strategies, at every time τ_j that a mechanism occurs, each trader i prefers, at least weakly, to participate in the session and obtain the resulting conditional expected cash and asset transfers, over the alternative of not participating.

It turns out that, in equilibrium, the continuation value of trader i at time t depends only on z_t^i and Z_t. So, it does not matter to trader i whether or not the other $n-1$ traders participate, in the off-equilibrium event that trader i opts out of the mechanism.

Our notion of equilibrium implies market clearing, rational conjectures of other traders' strategies, and individual trader optimality, including the incentive compatibility of truth-telling and individual rationality of participation in all reallocation sessions. The appendix analyzes the discrete-time version of this model, showing that the analogous equilibrium is Perfect Bayes.

The definition of individual trader optimality in this dynamic game is relatively obvious from the previous sections, but is now stated for completeness. Taking as given the demand coefficients (a, b, c) used by other traders and the mechanism design (Y, T_κ) for size-discovery sessions, trader i faces the problem of choosing a demand process D^i and report process $\hat{z}^i$ that solve the Markov stochastic control problem

$$V_A^i(z_0^i, Z) = \sup_{D, \tilde{z}} \mathbb{E}^i \left[z_{\mathcal{T}}^{D,\tilde{z}} \pi - \int_0^{\mathcal{T}} \gamma (z_t^{D,\tilde{z}})^2 + \Phi_{(a,b,c)}(D_t; Z_t - z_t^{D,\tilde{z}}) D_t \, dt + \int_0^{\mathcal{T}} T_\kappa^i((\tilde{z}_t, \hat{z}_t^{-i}), Z_t) \, dN_t \right],$$

where $\mathbb{E}^i$ denotes expectation conditional on $\mathcal{F}_0^i$ and

$$z_t^j = z_0^j + \int_0^t \hat{D}_s^j \, ds + H_t^j + \int_0^t \left(\frac{\tilde{z}_s + \sum_{j \neq i}^n \hat{z}_s^j}{n} - \hat{z}_s^j \right) dN_s \tag{21}$$

$$z_t^{D,\tilde{z}} = z_0^i + \int_0^t D_s\, ds + H_t^i + \int_0^t \left(\frac{\tilde{z}_s + \sum_{j \neq i}^n \hat{z}_s^j}{n} - \tilde{z}_s \right) dN_s, \tag{22}$$

taking $\hat{D}_t^j = a + b\Phi_{(a,b,c)}\left(D_t; Z_t - z_t^{D,\tilde{z}}\right) + c z_t^j$.

The definition of incentive compatibility for the equilibrium is that the report process $\hat{z}^i = z^i$ must be optimal for each trader. The equilibrium ex-post individual rationality condition for agent i is that, for all t,

$$V_A(z_t^i, Z_t) \leq V_A^i\left(z_t^i + \frac{\sum_j \hat{z}_t^j}{n} - \hat{z}_t^i, Z_t \right) + T_\kappa^i(\hat{z}_t, Z_t). \tag{23}$$

Proposition 4. *Suppose that $\lambda < r(n-2)$. Let $\kappa_0 < 0$ be arbitrary, and fix the mechanism design (Y, T_κ) specified by (3) and (4), where*

$$\kappa_1(Z_t) = v - \frac{2\gamma \bar{Z}_t}{r}, \qquad \kappa_2(Z) = -\overline{Z}_t - \frac{\kappa_1(Z_t)}{2\kappa_0 n^2}.$$

1. *Among equilibria in the dynamic game associated with the sequential double-auction market augmented with size-discovery sessions, there is a unique equilibrium with symmetric affine double-auction demand functions. In this equilibrium, the double-auction demand function $\mathcal{D}_t^i$ of trader i in state ω at time t is given by*

$$\mathcal{D}_t^i(\omega, p) = \frac{-r\lambda + r^2(n-2)}{4\gamma}\left(v - p - \frac{2\gamma}{r} z_t^i(\omega) \right). \tag{24}$$

   That is, the coefficients (a, b, c) of the demand function are

$$a = \frac{[-r\lambda + r^2(n-2)]v}{4\gamma}, \qquad b = \frac{r\lambda - r^2(n-2)}{4\gamma}, \qquad c = \frac{\lambda - r(n-2)}{2}.$$

2. *The market-clearing double-auction price process ϕ is given by $\phi_t = \kappa_1(Z_t)$.*

3. *The mechanism design (Y, T_κ) achieves the perfect post-session allocation $z^i(\tau_k) = \overline{Z}(\tau_k)$ for each trader i at each session time τ_k.*

4. *For each trader i, the equilibrium indirect utility $V_A^i(z_t^i, Z_t)$ at time t is identical to the indirect utility $V^i(z_t^i, Z_t)$ given by (13) for the model without size-discovery sessions. Thus, welfare is invariant to this augmentation of the double-auction market with size-discovery mechanisms.*

The equilibrium strategies are ex-post optimal in the same sense described in earlier sections.

That is, even if traders were to observe each others' current and past asset inventories, their equilibrium strategies would remain optimal.

From a comparison of the equilibrium demand schedules (14) and (24) that apply before and after augmenting the double-auction market with size-discovery mechanism sessions, we see that the introduction of size-discovery sessions reduces the magnitude of the slope of the demand functions by $r\lambda/(4\gamma)$. With size-discovery sessions, traders shade their demands in the double auction to mitigate price impact even more than they would in a market without size-discovery sessions. The next size-discovery session is expected by each trader to be so effective at reducing the magnitude of that trader's excess inventory, with a low price impact, that it is individually optimal for traders to reduce the speed with which they rebalance their inventories in the double-auction market. Of course, this is not socially efficient. The welfare cost of this relaxation of order submission in the double-auction market exactly offsets the welfare improvement directly associated with the size-discovery sessions. The two market designs are not only equivalent in terms of total welfare, they are also equally desirable from the viewpoint of each individual trader. In particular, there is no incentive for any subset of traders to set up a size-discovery platform.

Figure 2 illustrates the implications of augmenting a price-discovery market with size-discovery sessions. This figure shows simulated sample paths for the excess inventories of two of the $n = 10$ traders, with and without size-discovery mechanisms, based on the equilibria characterized by Propositions 4 and 3, respectively. For each of the two traders whose inventories z^i are pictured, the inventory shock process H^i is an independent Brownian motion with standard deviation ("volatility") parameter $\sigma_i = 0.05$. The aggregate inventory Z_t is a Brownian motion that is independent of $\{H^1, H^2\}$, with standard deviation parameter $\sigma_Z = 0.15$. The mean frequency of size-discovery sessions is $\lambda = 0.12$. The other parameters are shown in the caption of the figure. The graphs of the asset positions are shown in heavy line weights for the market with optimal size-discovery mechanisms, and in light line weights for the market with no size-discovery sessions. In the market that is augmented with size discovery, the first such mechanism session is held at about time $t = 10$, and causes a dramatic reduction in inventory imbalances, bringing the excess inventories of all traders to the perfectly efficient level, the cross-sectional average inventory $\overline{Z}(\tau_1) = -0.05$. In the illustrated scenario, although there are no more size-discovery sessions until time 680, traders in the market that includes size discovery anticipate that they will be able to shed excess inventories at the next such session, whenever it will occur, so they allow their excess inventories to wander relatively far from the efficient level $\bar{Z}_t$, avoiding price impact in the meantime by bidding relatively inaggressively in the double-auction market. For each trader i, because the anticipation of size-discovery sessions causes other traders to bid less aggressively, market depth is lowered, so that trader i has this additional incentive to bid less aggressively, relative to the market without size-discovery sessions. Indeed, as one can see, during the period that roughly spans from time 110 until time 680, the market without size discovery performs somewhat better, ex post, than the market with size discovery. However, ex ante, or looking forward from any point in time, the two market designs have the same allocative efficiency, as stated by Result 4 of Proposition 4.

5 Unobservable Aggregate Market Inventory

We now remove the assumption that the aggregate inventory $Z_t \equiv \sum_i z_0^i + H_t^i$ is observable. If Z_t is not directly observable by the size-discovery platform operator, then the size-discovery mechanism designer cannot use the cash-transfer function T_κ, because the κ_1 and κ_2 coefficients of T_κ depend on Z_t. As a consequence, the mechanism design and equilibrium behavior change significantly.

Even though the mechanism designer cannot directly observe Z_t, it turns out that the perfect reallocation $z_t^i = \overline{Z}_t$ can be achieved at each session time because the mechanism designer can infer the aggregate inventory Z_t precisely[20] from the "immediately preceding" double-auction market price $\phi_{t-} = \lim_{s\uparrow t}$. However, traders now understand that they can strategically influence their cash compensation in the next size-discovery session by influencing the double-auction price in advance of that. For example, a buyer now has an additional incentive to lower the market clearing price, and will demand less in the double-auction market. Likewise, a seller will supply less. This delays the rebalancing of positions across traders, strictly lowering welfare relative to a market with no size discovery.

In the double-auction market, we will limit attention to equilibria involving symmetric affine demand strategies, as in the model of the previous section, although with potentially different demand coefficients (a, b, c). We will restrict attention to a direct revelation mechanism $(Y, \hat{T})$ that exploits the perfect-reallocation scheme $Y(\,\cdot\,)$ of (3). Thus, the inventory processes are again defined by (20).

We will apply the mechanism cash transfers $\hat{T}(\hat{z}_t; \phi_{t-})$ associated with the function $\hat{T}$: $\mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$ defined, for an arbitrary constant $\kappa_0 < 0$, by

$$\hat{T}^i(\hat{z}; p) = \kappa_0 \left(-n\delta(p) + \sum_{j=1}^{n} \hat{z}^j \right)^2 + p\left(\hat{z}^i - \delta(p)\right) + \frac{p^2}{4\kappa_0 n^2}, \tag{25}$$

[20] This applies except in the zero-probability event that a mechanism session happens to be held precisely at a jump time of Z. Because this event has zero probability, it can without loss of generality be ignored in our calculations.

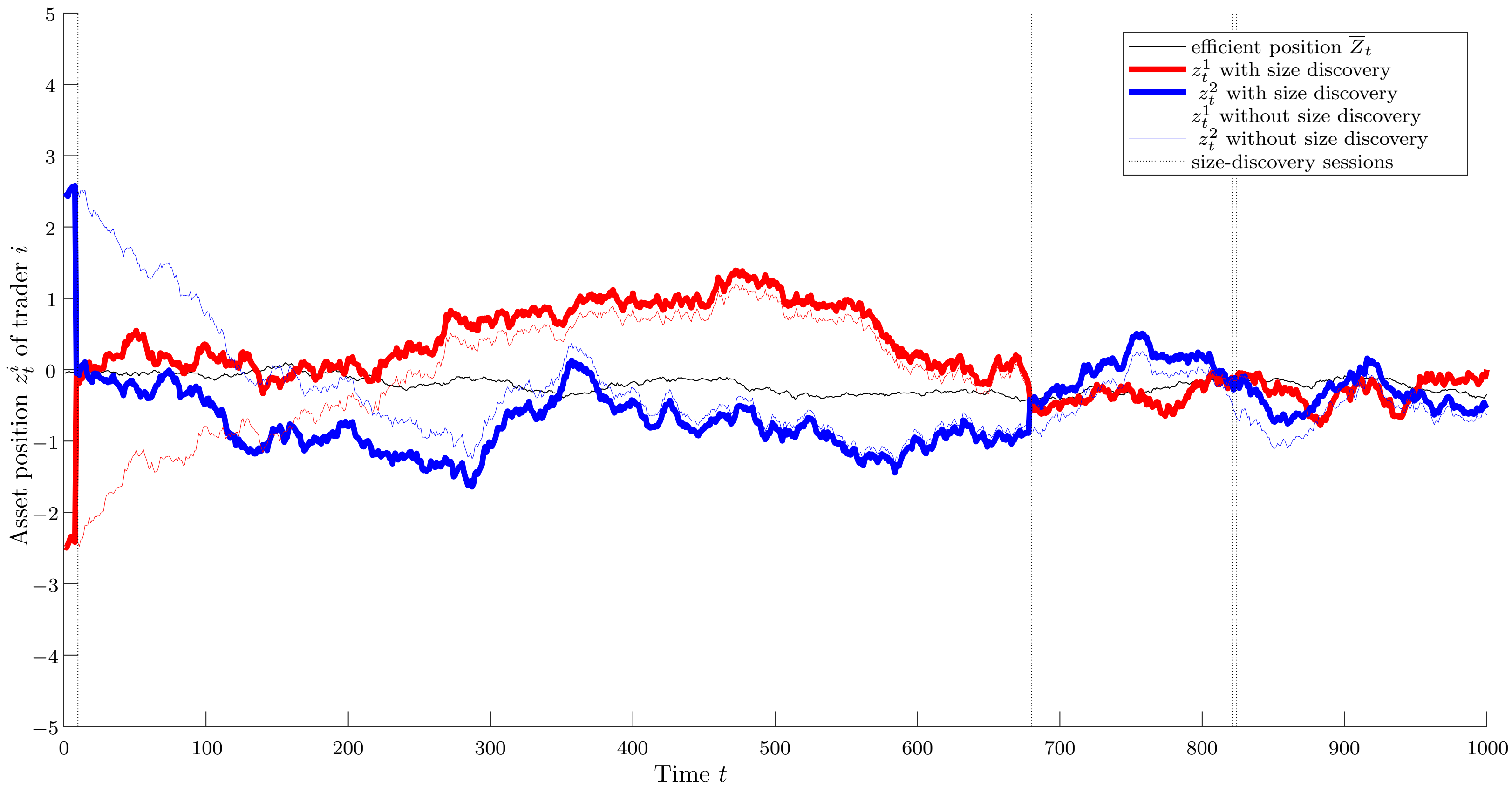

Figure 2: Inventory sample paths with and without size-discovery. This figure plots the inventory sample paths of 2 of the $n = 10$ traders, with and without size-discovery mechanisms, based on the equilibria characterized by Propositions 4 and 3, respectively. For each agent, the inventory shock process is an independent Brownian motion with standard-deviation parameter $\sigma_i = 0.05$. The aggregate inventory is an independent Brownian motion with standard-deviation parameter $\sigma_Z = 0.15$. The other parameters are mean asset payoff $v = 0.5$, mean rate of arrival of asset payoff $r = 0.1$, inventory cost coefficient $\gamma = 0.1$, initial aggregate market inventory $Z_0 = -0.5$, an initial asset position of trader 1 of $z_0^1 = -2.5$, an initial asset position of trader 2 of $z_0^2 = 2.5$, and a mean frequency $\lambda = 0.1167 = 0.99\bar{\lambda}$ of size-discovery sessions. The graphs of the asset positions are shown in heavy line weights for the market with optimal size-discovery mechanisms, and in light line weights for the market with no size-discovery sessions.

where

$$\delta(p) = \frac{-rv}{2\gamma} + p\left(\frac{r}{2\gamma} - \frac{1}{2n^2\kappa_0}\right). \tag{26}$$

The role of the prior price ϕ_{t-} is analogous to that applied in conventional forms of size-discovery used in practice, such as workup and dark pools. In a dark pool, as explained by Zhu (2014), the per-unit price is set by rule to the immediately preceding mid-price in a designated limit-order-book market. In BrokerTec's Treasury-market workup sessions, as explained by Fleming and Nguyen (2015), the frozen price used for workup compensation is fixed at the last trade price in the immediately preceding order-book market operated by the same platform provider. Thus, in dark pools, workup, and other forms of size-discovery used in practice, and also in this setting for our model, there is an incentive for traders to bid strategically in the double-auction market so as to avoid worsening their cash compensation terms in the next size-discovery session, through their impact on the market price ϕ_{t-}.

As in the previous section, given the mechanism $(Y, \hat{T})$, a symmetric equilibrium for the associated dynamic game is defined by a collection (a, b, c) of demand coefficients with the same properties described in the previous section of (A) individual optimality for each trader at all times, including optimal truthtelling, given rational conjectures of other trader's strategies, and (B) rationality of individual participation.

In particular, the problem faced by trader i is the choice of a double-auction-market demand process D^i and a report process $\hat{z}^i$ solving

$$V_S^i(z_0^i, Z_0) = \sup_{D,\tilde{z}} \mathbb{E}^i\left[z_{\mathcal{T}}^{D,\tilde{z}}\pi - \int_0^{\mathcal{T}}\left[\gamma\left(z_t^{D,\tilde{z}}\right)^2 + \Phi_{(a,b,c)}(D_t; Z_t - z_t^{D,\tilde{z}})D_t\right]dt\right] \tag{27}$$

$$+\ \mathbb{E}^i\left[\int_0^{\mathcal{T}} \hat{T}^i((\tilde{z}_t, \hat{z}_t^{-i}); \Phi_{(a,b,c)}(D_{t-}; Z_{t-} - z_{t-}^{D,\tilde{z}}))\, dN_t\right],$$

subject to Equations (21) and (22).

In contrast to the previous setting, for any fixed κ_0, there are exactly two such symmetric equilibria. The demand function of one of these equilibria has a bigger slope than that of the other. One equilibrium therefore has low order flow and high price impact. The other equilibrium has higher order flow and lower price impact. The following proposition characterizes the equilibria, and calculates the equilibrium associated with higher order flow, which is the more efficient of the two equilibria.

For this purpose, let $\bar{\lambda}$ be the unique positive solution of the equation

$$3\bar{\lambda} + \sqrt{8\bar{\lambda}(r + \bar{\lambda})} = (n-2)r. \tag{28}$$

Proposition 5. *Suppose $\lambda \leq \bar{\lambda}$. Fix any $\kappa_0 < 0$. Given the mechanism $(Y, \hat{T})$ defined by (3) and (25), there exist equilibria with symmetric affine double-auction demand functions for the dynamic game associated with the sequential auction markets augmented with size-discovery sessions. Each such equilibrium has the following properties.*

1. *The market-clearing double-auction price process ϕ is given by*

$$\phi_t = v - \frac{2\gamma}{r}\overline{Z}_t. \tag{29}$$

2. *The double-auction market demand of trader i at time t is $a+b\phi_t+cz_t^i$, for some coefficients (a,b,c) with $b<0$.*

3. *The post-session allocation at each size-discovery session time at each session time τ_k is the perfect allocation $z^i(\tau_k) = \overline{Z}(\tau_k)$, almost surely.*

4. *For each trader i, the equilibrium indirect utility at time t is*

$$V_S^i(z_t^i, Z_t) = \theta_i' + v\overline{Z}_t - \frac{\gamma}{r}\overline{Z}_t^2 + \phi_t\left(z_t^i - \overline{Z}_t\right) - K\left(z_t^i - \overline{Z}_t\right)^2, \tag{30}$$

where

$$K = \frac{\gamma}{r(n-1)} - \frac{\lambda}{2b(n-1)} \tag{31}$$

and

$$\theta_i' = \frac{1}{r}\left(\frac{\gamma}{r}\frac{\sigma_Z^2}{n^2} - K\left(\frac{\sigma_Z^2}{n^2} + \sigma_i^2 - 2\frac{\rho^i}{n}\right) - \frac{2\gamma}{r}\frac{\rho^i}{n}\right). \tag{32}$$

5. *In the more efficient equilibrium, the double-auction demand function coefficients are given by*

$$a = -vb \tag{33}$$

$$b = \frac{-r^2}{8\gamma}\left(-\frac{3\lambda}{r} + (n-2) + \sqrt{\left(\frac{\lambda}{r} - (n-2)\right)^2 - \frac{4\lambda n}{r}}\right) < 0 \tag{34}$$

$$c = \frac{2\gamma}{r}b. \tag{35}$$

6. *In this particular equilibrium (33)-(35), the slope b of the demand function is monotonic increasing*[21] *with respect to the mean frequency λ of size-discovery sessions. (The magni-*

[21] That is, for each $\lambda_0 < \bar{\lambda}$ and each associated equilibrium demand function coefficients (a_0, b_0, c_0), there is a mapping $\lambda \mapsto (a_\lambda, b_\lambda, c_\lambda)$ on a neighborhood of λ_0 to a neighborhood of (a_0, b_0, c_0), specifying the unique equilibrium demand coefficients $(a_\lambda, b_\lambda, c_\lambda)$ for each λ in the neighborhood of λ_0. The coefficient b_λ is increasing in λ.

tude of b is therefore decreasing in λ.)

In an equilibrium postulated by the proposition, traders are free to deviate from their affine strategies, and could consider manipulating the double-auction price so as to influence the size-discovery session operator's inference of the aggregate inventory Z_t from the market clearing price ϕ_{t-}. For example, if their inventory z_t^i is large, then trader i, absent any motive to affect inference by the session platform operator, would naturally submit large orders to sell. By instead submitting a small buy order, the resulting (off-equilibrium-path) price ϕ_t would be higher, suggesting to the platform operator a smaller aggregate inventory. If a size-discovery session were to occur immediately afterward, the designer would then implement cash transfers based on this "distorted" price. The cash transfers would more generously compensate traders who have (and report) larger inventories, given the rebalancing objective of the platform operator. If the mechanisms are run too frequently, however, this incentive to distort the price through order submission becomes so great that the double-auction market breaks down, in that linear equilibrium demand functions cease to exist.

We now focus on the particular equilibrium defined by (33)-(35). As λ increases from zero to the solution $\bar{\lambda}$ of (28), the expected total volume of trade in the double-auction market declines. Once λ exceeds $\bar{\lambda}$, if an equilibrium were to exist there would be so little order flow that it becomes sufficiently cheap for traders to manipulate the price, in order to benefit from the next size-discovery session, that markets could not clear. That is, the double-auction market would break down, and there is in fact no equilibrium with $\lambda > \bar{\lambda}$.

Given that the equilibrium double-auction demand functions have slope $b < 0$, the second term in the definition (31) of the quadratic coefficient K is positive, provided there is a non-zero mean arrival rate λ for size-discovery sessions. This implies that the inability of the platform operator to directly observe the aggregate inventory balance Z_t causes an additional reduction in allocative efficiency. In fact, in this setting, adding size-discovery sessions to the price-discovery double-auction market causes a strict reduction in welfare! The welfare at any time t in this setting is

$$\hat{W}(z_t) \equiv \sum_{i=1}^{n} V_S^i(z_t^i, Z_t) = \sum_{i=1}^{n} \theta_i' + vZ_t - \frac{n\gamma}{r}\overline{Z}_t^2 - K\sum_{i=1}^{n}\left(z_t^i - \overline{Z}_t\right)^2, \tag{36}$$

which is strictly lower than the welfare for the same market without size-discovery.[22] With stochastic and unobservable total inventory, each trader shades his or her orders in the double-auction market because of the adverse expected impact of aggressive order submissions on the terms of cash compensation that will be received in the next reallocation session.

[22]The exception is of course the degenerate case of $\lambda = 0$, for which $K = -\gamma/(r(n-1))$ and the two welfare functions coincide.

We see from (36) that equilibrium welfare is strictly decreasing in K and strictly increasing in $\sum_{i=1}^{n} \theta_i'$. In the equilibrium of Proposition 5, K is monotonically increasing in λ,[23] while each θ_i' is monotonically decreasing in λ. That is, equilibrium welfare only gets worse as the frequency of size-discovery sessions is increased, until size-discovery sessions are so frequent that the price-discovery market breaks down.

Moreover, each trader individually strictly prefers the market design without size discovery. That is, if size discovery exists, it is individually rational for traders to participate in each size-discovery session, but all traders would prefer to commit to a market design in which size discovery does not exist.

Figure 3 illustrates the implications of augmenting a price-discovery market with price-based size-discovery sessions. As in Figure 2, this figure shows simulated inventory sample paths of two of the $n = 10$ traders, with and without size-discovery mechanisms, now based on the equilibria characterized by Propositions 5 and 3, respectively. Figures 2 and 3 are based on the same model parameters and the same simulated scenarios for the inventory shock process $H = (H^1, \ldots, H^n)$ and size-discovery session times $\tau_1, \tau_2, \ldots$. The graphs of the asset positions shown in heavy line weights are for the market with optimal size-discovery mechanisms. Those paths shown in light line weights correspond to the market with no size-discovery sessions. In the market that is augmented with size-discovery, the first such session is held at about time $t = 10$, and causes a dramatic reduction in inventory imbalances, bringing the excess inventories of all traders to the perfectly efficient level, the cross-sectional average inventory $\overline{Z}(\tau_1) = -0.05$. However, because traders shade their bids even more than in the equilibrium of Proposition 4, from roughly time 110 until time 680 for these inventory sample paths, the market without size discovery performed dramatically better, ex post, than the market with size discovery. This is consistent with the result that, looking forward from any point in time, the market design of Proposition 5 has strictly worse allocative efficiency than that of Proposition 3. A comparison with Figure 2 shows the degree to which the informational reliance in size-discovery sessions on prior double-auction market prices worsens the allocative efficiency of the double-auction markets.

6 Mechanisms Only

In the previous sections, we showed that augmenting a price-discovery market with future size-discovery sessions never increases welfare, and strictly reduces welfare if the size-discovery platform operator relies on the price-discovery market for information about aggregate inventory imbalances. It is then natural to ask whether simply getting rid of the price-discovery market,

[23] This follows from (31) since b is negative and increases monotonically in λ.

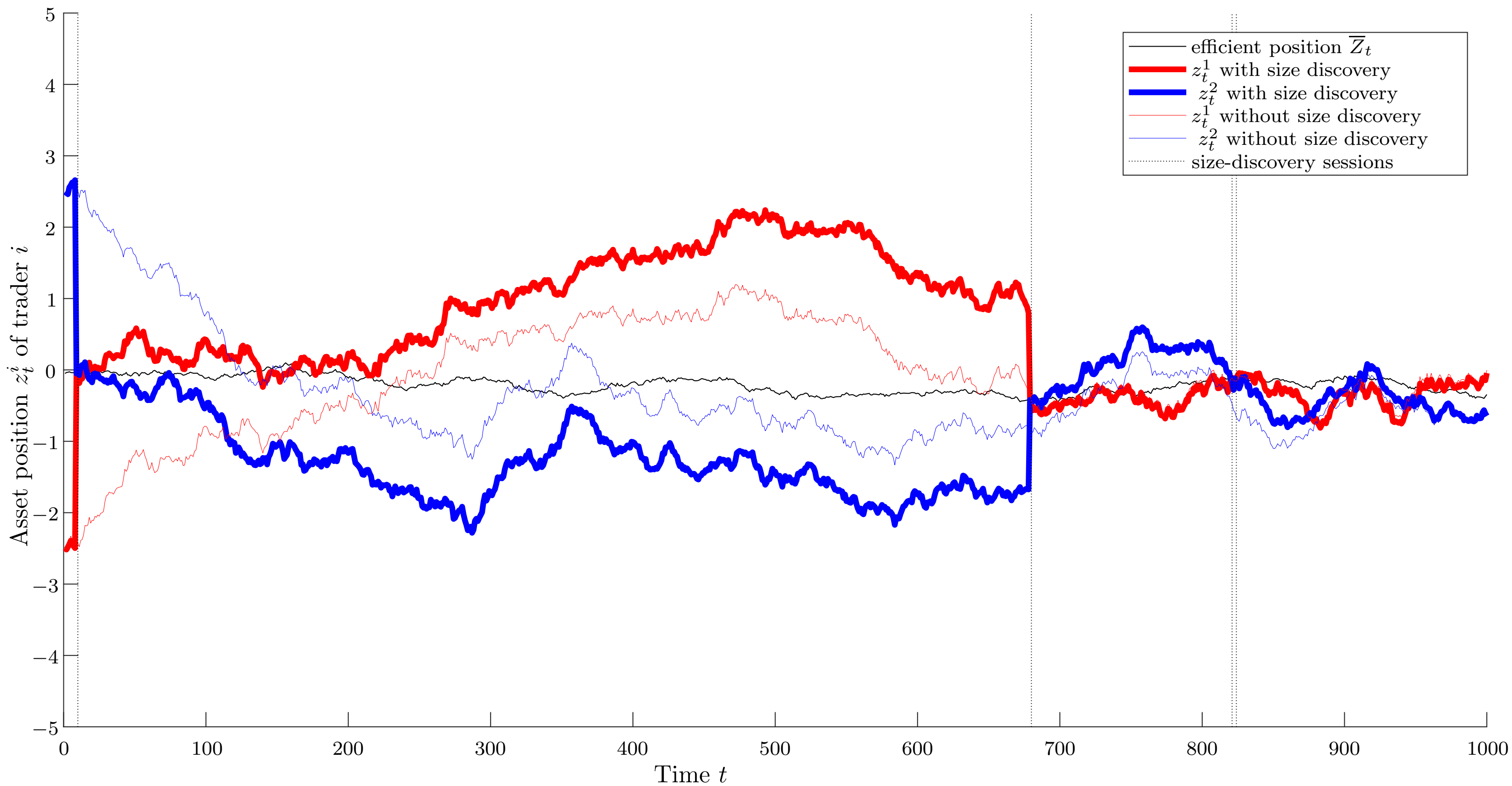

Figure 3: Inventory sample paths with and without price-based size discovery. This figure plots the inventory sample paths of 2 of the $n = 10$ traders, with and without size-discovery mechanisms, based on the equilibria characterized by Propositions 5 and 3, respectively. For each agent, the inventory shock process is an independent Brownian motion with standard-deviation parameter $\sigma_i = 0.05$. The aggregate inventory is an independent Brownian motion with standard-deviation parameter $\sigma_Z = 0.15$. The other parameters are mean asset payoff $v = 0.5$, mean rate of arrival of asset payoff $r = 0.1$, inventory cost coefficient $\gamma = 0.1$, initial aggregate market inventory $Z_0 = -0.5$, an initial asset position of trader 1 of $z_0^1 = -2.5$, an initial asset position of trader 2 of $z_0^2 = 2.5$, and a mean frequency $\lambda = 0.1167 = 0.99\bar{\lambda}$ of size-discovery sessions. The graphs of the asset positions are shown in heavy line weights for the market with optimal size-discovery mechanisms, and in light line weights for the market with no size-discovery sessions.

and running only size-discovery sessions, could improve welfare, relative to a setting with price discovery. When stand-alone size discovery is feasible and is run sufficiently frequently, it strictly improves welfare, and indeed is strictly preferred by each trader individually. From a practical viewpoint, however, it could be difficult to arrange for the abandonment of price-discovery markets. Moreover, the size-discovery sessions that we analyze might be difficult to implement in practice without information coming out of the price-discovery market.

In this section, we consider a pure size-discovery market, for an economy with observable aggregate inventory. For example, it suffices that Z is a deterministic constant. We exploit the same perfect-reallocation size-discovery sessions developed earlier. As before, these sessions are run at the event times of an independent Poisson process N with mean arrival rate $\lambda > 0$.

Again, traders submit mechanism report processes $\hat{z} = (\hat{z}^1, \ldots, \hat{z}^n)$. The resulting excess-inventory process z^i of trader i is then determined by

$$z_t^i = z_0^i + H_t^i + \int_0^t \left(\frac{\sum_{j=1}^n \hat{z}_s^j}{n} - \hat{z}_s^i \right) dN_s. \tag{37}$$

As in Section 4, we assume that the aggregate inventory Z_t is common knowledge for all t. The size-discovery mechanism design (Y, T_κ) uses the asset reallocation determined by (3). We again apply the cash-transfer function T_κ defined by (4) for some coefficient $\kappa_0 < 0$, with

$$\kappa_1(Z_t) = v - \frac{2\gamma}{r}\overline{Z}_t \tag{38}$$

and

$$\kappa_2(Z_t) = -\overline{Z}_t - \frac{\kappa_1(Z_t)}{2\kappa_0 n^2}. \tag{39}$$

By the same reasoning used in Propositions 1 and 2, one can show these are the unique affine choices for $\kappa_1(\cdot)$ and $\kappa_2(\cdot)$ such that an equilibrium exists. Moreover, we must restrict attention to affine $\kappa_1(\cdot), \kappa_2(\cdot)$ in this dynamic setting in order to guarantee a linear-quadratic continuation-value function.

We seek a truth-telling equilibrium of the dynamic reporting game, in which each trader optimally chooses to report $\hat{z}_t^i = z_t^i$ and in which mechanism participation is always individually rational. The exact stochastic control problem solved by each trader is an obvious simplification of the control problem of Section 4, which appears in the appendix. The next proposition confirms that this equilibrium exists and provides a calculation of the continuation value for each trader.

Proposition 6. *For any $\kappa_0 < 0$, consider the size-discovery session mechanism design (Y, T_κ) of (3)-(4), with (38)-(39). The truth-telling equilibrium, that with reports $\hat{z}_t^i = z_t^i$, exists and* *has the following properties.*

1. *At each session time τ_k, each trader i achieves the efficient post-session position $z^i(\tau_k) = \overline{Z}(\tau_k)$, almost surely.*

2. *For each trader i, the equilibrium continuation value $V_M^i(z_t^i, Z_t)$ at time t is*

$$V_M^i(z_t^i, Z_t) = \tilde{\theta}_i + v\overline{Z}_t - \frac{\gamma}{r}\overline{Z}_t^2 + \kappa_1(Z_t)\left(z_t^i - \overline{Z}_t\right) - \frac{\gamma}{r+\lambda}\left(z_t^i - \overline{Z}_t\right)^2,$$

where

$$\tilde{\theta}_i = \frac{1}{r}\left(\frac{\gamma}{r}\frac{\sigma_Z^2}{n^2} - \frac{\gamma}{r+\lambda}\left(\frac{\sigma_Z^2}{n^2} + \sigma_i^2 - 2\frac{\rho^i}{n}\right) - \frac{2\gamma}{r}\frac{\rho^i}{n}\right).$$

As the mean frequency λ of reallocation sessions approaches infinity, the equilibrium welfare approaches the first-best welfare $W_{fb}(Z)$. This follows from the fact that the equilibrium total expected holding costs associated with excess inventory, relative to the holding costs at first best, approaches zero[24] as $\lambda \to \infty$. This is immediate from the fact that the quadratic coefficient $\gamma/(r+\lambda)$ of the indirect utility V_M^i approaches zero as $\lambda \to \infty$. These properties hold for any choice of $\kappa_0 < 0$, but setting $\kappa_0 = -\gamma(n-1)/(n^2(r+\lambda))$ makes each trader indifferent to instantaneous deviations by other traders.[25]

7 Discussion and Concluding Remarks

We conclude by discussing some implications for market designs involving both price discovery and size discovery.

7.1 Some discouraging market-design observations

The central result of the paper is that augmenting a price-discovery market (an exchange, in our case a dynamic double-auction market) with optimal size-discovery mechanisms does not improve allocative efficiency. Actually, for the more realistic case in which the size-discovery platform operator relies on the price-discovery market to help set the terms of compensation

[24] This convergence is also intuitively obvious from the fact that $\delta_t^i \equiv (z_t^i - \overline{Z}_t)^2$ jumps to zero at each of the event times of N. The duration of time between these successive perfect reallocations has expectation $1/\lambda$, which goes to zero. Between these perfect reallocations, δ_t^i has a mean that is continuous in t and grows in expectation at a bounded rate.

[25] Formally, if we consider the static mechanism report game with the continuation value corresponding to proposition 6, for this κ_0 truth-telling is a dominant strategy.

in size-discovery sessions, welfare is strictly lowered by adding size-discovery. Although the total welfare of market participants jumps up at each size-discovery session, the prospect of subsequent size-discovery sessions reduces the expected gains from trade in the price-discovery market between size-discovery sessions. The net effect is to leave welfare at least as low as that achieved without size-discovery, and strictly lower when the size-discovery operator relies on price information from the price-discovery market.

From a normative market-design viewpoint, this result is discouraging.

We do show that the first-best allocation can be achieved in principle by relying entirely on size-discovery, and simply dispensing with price-discovery markets. Even if such a radical redesign of markets could be realistically contemplated, it would require that the size-discovery platform operator is able to compute what would have been the market-clearing price $\phi_t = v - 2\overline{Z}_t\gamma/r$ in a double-auction market, were one to exist. This price-related information may be difficult to obtain in practice without actually opening the price-discovery market. The pieces of information needed to construct this hypothetical price ϕ_t are the mean payoff v of the asset, the average current excess inventory $\overline{Z}_t$ of market participants, the inventory cost coefficient γ, and the mean duration of time r^{-1} before the asset payoff occurs. In addition to its allocative role, the price-discovery market serves the role of constructing and revealing this price information.

We also showed that a market designer cannot rely on the price-discovery market merely to learn the price ϕ_t, and then achieve nearly full efficiency by running size-discovery sessions arbitrarily frequently. As λ rises, market participants become less and less active in the price-discovery market, in anticipation of the next size-discovery session, given the very low effective trading "cost-impact" of order submission in size-discovery sessions. If λ exceeds a specific threshold $\bar{\lambda}$, there would be no reliable price information coming out of the double-auction market. This is so because the resulting extremely low trade volume would make it so cheap to "push the price," in order to benefit from improved compensation in the subsequent size-discovery mechanism, that the price-discovery market would break down. The terms of compensation in the size-discovery sessions would thus need to be obtained from some other source.

Ye (2016) offers a model in which a dark pool can indeed harm the formation of informative prices. For a different model, Zhu (2014) obtains the opposite result for cases that do not involve large-trader price impact.

7.2 Cross-venue competition and stability

The observations of the previous subsection also imply that there may be a tenuous relationship between the operators of size-discovery and price-discovery platforms, respectively. Barring nearly omniscient alternative information sources, the size-discovery platform operator may need to rely heavily on the prices ϕ_t being produced in price-discovery markets. The size-discovery venue operator can draw more and more volume away from the price-discovery market by holding more and more frequent size-discovery sessions. In theory, the size-discovery venue could in some cases capture an arbitrarily large fraction of the total volume of trade across the two venues. In practice, however, the size-discovery operator would stop short, or be stopped short by others, out of practical business or regulatory concerns. CFA Institute (2012) address general concerns in this area, summarizing with the comment "The results of our analysis show that increases in dark pool activity and internalization are associated with improvements in market quality, but these improvements persist only up to a certain threshold. When a majority of trading occurs in undisplayed venues, the benefits of competition are eroded and market quality will likely deteriorate."

This concern may in some cases lead toward integration of the sponsors of price-discovery platforms and size-discovery platforms for trading the same asset, along the lines of BrokerTec, which operates both of these protocols for treasuries trading on the same screen-based platform. Even in this case, however, Schaumburg and Yang (2016) point to some interference arising from price information arriving during size-discovery sessions from the simultaneous operation of treasury futures trading on the Chicago Mercantile Exchange.

Zhu (2014) has shown that in a setting with asymmetric information about asset payoffs, there tends to be a selection bias by which relatively informed investors migrate toward price-discovery markets and relatively less informed investors migrate toward dark pools. This seems to suggest support for robust trade volumes on both types of venues. On the other hand, Zhu (2014) addressed the case of dark pools that promote this selection effect with delays in dark-pool order execution caused by rationing, because rationing discourages informed investors who want to act quickly on their information. As we have pointed out, dark-pool rationing is a relatively crude mechanism design for size-discovery. Although we have not analyzed the implications in our setting of adding asymmetric information about asset payoffs, one may anticipate from our results that more efficient mechanism designs than those currently used in dark pools would be less discouraging to informed investors. This could call into question the robustness of a market design that allows size-discovery venues to free-ride on the price information coming from lit exchanges, while also having a significant ability to draw volume away from lit exchanges.

As of late 2017, according to Rosenblatt Securities, dark pools account for about 15% of

U.S. equity trading volume.[26]

7.3 Intentional impairment of size-discovery mechanisms

One might be drawn to conjecture that our mechanism design for size-discovery is "too efficient." Indeed, we have shown that the reallocative efficiency and low effective price impact of our size-discovery mechanism design offer such an attractive alternative for executing trades, relative to submitting orders into the price-discovery market, that they reduce price-discovery market depth enough to offset all of the benefit of adding size-discovery. We have shown that adding size-discovery can actually worsen overall market efficiency.

Given this tension, one might hope to impair the efficiency of the size-discovery design just enough to raise overall market efficiency. By this line of enquiry, one would look for a loss of size-discovery efficiency that is more than offset by a gain in price-discovery allocative efficiency through an improvement of market depth.

We have discovered that this approach does not work, at least among linear-quadratic schemes for size-discovery. In the appendix, we calculate a mechanism design in which the imbalance $z^i_{t-} - \bar{Z}_t$ in the inventory of trader i is not completely eliminated in the size-discovery session. Instead, only a specified fraction ξ of this imbalance is erased by size discovery. Any parameter ξ between 0 and 1 can be supported in an equilibrium with the same properties (other than full efficiency)[27] shown in Section 2, which treats the special case $\xi = 1$. The appendix provides a corresponding generalization of the dynamic trading model of Section 4. In this setting, overall welfare is invariant to the effectiveness ξ of size-discovery. That is, welfare is the same whether one runs perfect reallocation mechanisms ($\xi = 1$), arbitrarily imperfect size-discovery mechanisms ($0 < \xi < 1$), or no size-discovery mechanisms at all.[28]

[26] See "Let There be Light, Rosenblatt's Monthly Dark Liquidity Tracker," September 2017, at http://rblt.com/letThereBeLight.aspx?year=2017.

[27] We must, however, slightly modify our notion of budget balance. Given the equilibrium strategies, the mechanism is budget balanced with probability 1, but this might not be the case for arbitrary off-equilibrium reports.

[28] We find in unreported numerical examples that if the Z_t is unobservable, and in what is otherwise the setting of Proposition 5, welfare is strictly lower with impaired mechanisms than with no mechanisms at all.

References

Almgren, R., and N. Chriss. 2001. Optimal execution of portfolio transactions. *Journal of Risk* 3:5–40.

Arrow, K. 1979. The property rights doctrine and demand revelation under incomplete information. In M. Boskin, ed., *Economics and Human Welfare*. Academic Press, Cambridge, Massachusetts.

Athey, S., and I. Segal. 2013. An efficient dynamic mechanism. *Econometrica* 81:2463–85.

Bergemann, D., and J. Välimäki. 2010. The dynamic pivot mechanism. *Econometrica* 78:771–89.

CFA Institute. 2012. Dark pools, internalization, and equity market quality. White paper, CFA Institute.

d'Aspremont, C., and L.-A. Gérard-Varet. 1979. Incentives and incomplete information. *Journal of Public economics* 11:25–45.

Di Tella, S. 2017. Optimal regulation of financial intermediaries. Working Paper, National Bureau of Economic Research.

Di Tella, S., Y. Sannikov, et al. 2016. Optimal asset management contracts with hidden savings. *(Working Paper, Stanford University)* .

Du, S., and H. Zhu. 2017. What is the optimal trading frequency in financial markets? *Review of Economic Studies* 84:1606–51.

Duffie, D., and H. Zhu. 2017. Size discovery. *The Review of Financial Studies* 30:1095–150.

Dworczak, P. 2017. Mechanism design with aftermarkets: Cutoff mechanisms. Working paper, Becker Firedmand Institute, University of Chicago.

Fleming, M., and G. Nguyen. 2015. 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.

Klemperer, P. D., and M. A. Meyer. 1989. Supply function equilibria in oligopoly under uncertainty. *Econometrica* 57:1243–77.

Myerson, R. B. 1981. Optimal auction design. *Mathematics of operations research* 6:58–73.

Ollár, M., M. Rostek, and J. H. Yoon. 2017. Privacy in markets. Working paper, Department of Economics, University of Wisconsin, Madison.

Pancs, R. 2014. Workup. *Review of Economic Design* 18:37–71.

Pavan, A., I. Segal, and J. Toikka. 2014. Dynamic mechanism design: A myersonian approach. *Econometrica* 82:601–53.

Protter, P. 2005. *Stochastic integration and differential equations, second edition.* Heidelberg: Springer.

Rostek, M., and M. Weretka. 2012. Price inference in small markets. *Econometrica* 80:687–711.

———. 2015. Dynamic thin markets. *Review of Financial Studies* 28:2946–92.

Sannikov, Y., and A. Skrzypacz. 2016. Dynamic trading: Price inertia and front-running. Working paper, Graduate School of Business, Stanford University.

Schaumburg, E., and R. Yang. 2016. The workup, technology, and price discovery in the interdealer market for U.S. treasury securities. Liberty Street Economics, Federal Reserve Bank of New York.

Vayanos, D. 1999. Strategic trading and welfare in a dynamic market. *The Review of Economic Studies* 66:219–54.

Vives, X. 2011. Strategic supply function competition with private information. *Econometrica* 79:1919–66.

Wilson, R. 1979. Auctions of shares. *Quarterly Journal of Economics* 93:675–89.

Ye, L. 2016. Understanding the impacts of dark pools on price discovery. Working paper, The Chinese University of Hong Kong, Shenzhen.

Zhu, H. 2014. Do dark pools harm price discovery? *Review of Financial Studies* 27:747–89.