# Optimal market thickness*

Simon Loertscher† Ellen V. Muir‡ Peter G. Taylor§

This version: October 24, 2021 First version: February 15, 2016

**Abstract**

Traders that arrive over time give rise to a dynamic tradeoff between the benefits of increasing gains from trade by accumulating traders and the associated cost of delay due to discounting. We analyze this tradeoff in a dynamic bilateral trade model in which a buyer and seller arrive in each period and draw their types independently from commonly known distributions. With symmetric binary types, the optimal market clearing policy can be implemented with posted prices and ex post budget balance, provided it is optimal to store at least one trader. While optimally thick markets involve storing a small number of traders, their performance is nevertheless close to that of a large market. In particular, irrespective of the type distributions, two-thirds of the gains from increased market thickness can be achieved by storing just one trader.

**Keywords:** market thickness, dynamic mechanisms, posted-price mechanisms, two-sided private information, (im)possibility of efficient trade
**JEL-Classification:** C72, D47, D82

*We are grateful to two editors (Tilman Börgers and Alessandro Pavan) and five anonymous reviewers of this journal for feedback that helped us improve the paper. We also thank Ivan Balbuzanov, Eric Budish, Gabriel Carroll, Yeon-Koo Che, Darrell Duffie, Alexander Frankel, Dan Garrett, Scott Kominers, Thomas Mariotti, Leslie Marx, Hayden Melton, Claudio Mezzetti, Paul Milgrom, Alex Nichifor, Carlos Oyarzun, Agathe Pernoud, Andy Skrzypacz, Alexander Teytelboym, Steve Williams, Jun Xiao, Leeat Yariv and audiences at the University of Melbourne, University of Queensland, Ecole Polytechnique Fédérale de Lausanne, Stanford University, the Melbourne IO and Theory Day, the Econometric Society Australasian Meeting in Sydney 2016, the Econometric Society European Meeting in Geneva 2016, the European Association for Research in Industrial Economics Meeting in Lisbon 2016, the Columbia/Duke/MIT/Northwestern IO Theory Conference in Chicago 2016, the Australasian Economic Theory Workshop in Auckland 2017, the University of North Carolina at Chapel Hill, the University of New South Wales, the Chicago Mercantile Exchange and the University of Chicago for comments and suggestions. Financial support provided by the Australian Research Council through Laureate Fellowship FL130100039 and Discovery Project DP200103574, the ARC Centre of Excellence for Mathematical and Statistical Frontiers, the Elizabeth and Vernon Puzey Foundation and the Samuel and June Hordern Endowment is also gratefully acknowledged. Toan Le and Jonathan Lim provided excellent research assistance. We also acknowledge support from the European Research Council under the Grant Agreement no. 340903.

†Department of Economics, Level 4, FBE Building, 111 Barry Street, University of Melbourne, Victoria 3010, Australia. Email: simonl@unimelb.edu.au.

‡Department of Economics, Stanford University. Email: evmuir@stanford.edu

§School of Mathematics and Statistics, University of Melbourne, Melbourne, Australia. Email: taylorpg@unimelb.edu.au.

# 1 Introduction

This paper determines the optimal degree of market thickness in an infinite horizon model in which one buyer and one seller arrive in every period and develops a measure of market thickness that permits comparisons across different environments, thereby contributing to the nascent field of dynamic market making. Assuming that each buyer's value and each seller's cost is persistent, we consider a social planner whose objective is to maximize expected discounted social surplus. The basic tradeoff the planner faces is that increasing market thickness by storing traders increases welfare in future periods through increased gains from trade. However, this comes at the cost of delaying consumption.

The key forces at work are most transparent and tractable in a symmetric setting with binary types, where each trader's type is either *efficient* or *suboptimal*. Trades between efficient types maximize the gains from trade, while trades between efficient and suboptimal types achieve less than half of this maximum. There are no gains from trade between suboptimal types. The optimal market clearing policy in this setting is very intuitive. Suboptimal types should never be stored as the new arrivals in the future can be no worse than these types. Likewise, it is optimal to execute any trade between efficient types as soon as it becomes available. Hence, the problem reduces to deciding when to forgo a trade between an efficient trader and a suboptimal trader and store the efficient trader. We refer to the optimal policy as a threshold policy, as it involves storing a number of efficient traders up to a *threshold* and not executing any suboptimal trades until this threshold is reached.

The optimal storage threshold is increasing in the discount factor $\delta$. Provided it is positive, the optimal market clearing policy can be implemented using posted prices. Until the storage threshold is reached, the only trades that are executed are trades between efficient buyers and sellers, which can be implemented by simply posting a price equal to the average of their types. If the arriving buyer and seller are willing to trade at this price, an efficient trade is executed. If only one of these agents wants to trade, thereby revealing that its type is efficient, then this agent is either stored or it trades with a stored agent. Once the storage threshold is reached, a trade between an efficient and a suboptimal type is executed if the arriving agents consist of an efficient type of the kind stored (say, a seller) and a suboptimal type on the other side of the market (a buyer). Importantly, such a trade can still be implemented via a posted price by ensuring that this price is sufficiently favorable for the side of the market from which no agents are stored. In contrast, a posted-price implementation is not possible with a storage threshold of zero. In this case, the efficient policy induces ex post efficient trade in every period and the market-clearing prices depend on the arriving agents' types on both sides of the market.

The posted-price implementation shows that, whenever the optimal storage threshold is positive, the optimal market clearing policy can be implemented regardless of whether arriving agents are privately informed of their types. That is, there is no Myerson and Satterthwaite (1983) problem. Intuitively, with private information and a storage threshold of zero, implementing the ex post efficient allocation in every period requires that arriving trades involving efficient types are always executed. Identifying the efficient types requires that the planner pays an information rent to these types on *both* sides of the market in every period, potentially resulting in a deficit. With a positive storage threshold, suboptimal types only trade with positive probability in periods where the storage threshold has been reached and traders are stored on the *other* side of the market. Hence, the social planner pays an information rent to efficient types on *at most one* side of the market in any given period.

We show analytically and numerically that optimally thick markets typically involve storing a small number of traders. Specifically, as the discount factor $\delta$ increases, the optimal storage threshold grows slowly, at a rate that is bounded by $(1-\delta)^{-1}$. Moreover, for a wide range of plausible parameterizations, the optimal storage threshold is a single digit. This raises the question of why optimal storage thresholds are typically small and optimally thick markets are seemingly thin. Addressing this question requires a quantitative measure of market thickness. The optimal threshold, while a natural candidate, is of limited use as it is specific to this dynamic setting and only offers an ordinal measure.

With this in mind, we propose a cardinal measure of market thickness. For each storage threshold we consider the average surplus per period under the stationary distribution, which is the standard welfare criterion for a perfectly patient social planner. The maximum gain from increasing market thickness is the difference between the average per-period surplus with an unbounded storage threshold and a storage threshold of zero.[1] Our market thickness measure determines the proportion of this maximum gain that is achieved by a given threshold policy.[2] This measure is independent of the discount factor, allowing us to meaningfully quantify how the discount factor affects the *optimal* degree of market thickness. According to our market thickness measure, a threshold policy with a threshold of one always achieves two-thirds of the maximum gain from increasing market thickness, and a threshold policy with a threshold of six achieves more than 90 percent of this maximum gain.

Many of our results related to threshold policies extend to continuous type distributions.

[1]The average stationary per-period surplus with an unbounded storage threshold corresponds to expected surplus per buyer-seller pair in the Walrasian benchmark involving a static market with a continuum of buyers and sellers of equal mass. With a storage threshold of zero it corresponds to welfare in the bilateral trade benchmark.

[2]This measure can easily be adapted to alternative environments and mechanisms by choosing an appropriate welfare criterion and benchmark for measuring relative market thickness.

Exploiting the static benchmark with a continuum of traders on each side of the market, buyers and sellers can still be categorized as "efficient" or "suboptimal" depending on which side of the Walrasian price their types are. Absent private information, two-thirds of the maximum gain from increasing market thickness is still achieved using a threshold policy with a threshold of one. Accounting for private information, we show that this gain is even larger. Posting the Walrasian price in any given period still allows the planner to execute trades between efficient agents. When the storage threshold is reached and a trade involving a suboptimal type may need to be executed, a second-best mechanism is required to determine whether there are any gains from trade.[3]

In the final section of the paper we also consider the case of a profit-maximizing market maker and show that such a market maker induces an excessively thick market relative to the benevolent social planner. While many important questions (such as the nature of optimal mechanisms away from the binary type setting) remain open, this paper provides a starting point for studying efficient and profit-maximizing market making in dynamic setups.

Our paper relates to two strands of literature. The first includes Gresik and Satterthwaite (1989), Satterthwaite and Williams (1989, 2002), McAfee (1992), Rustichini et al. (1994), and Cripps and Swinkels (2006) and analyzes double auctions in static settings. Motivated by the impossibility results of Vickrey (1961), Hurwicz (1972) and Myerson and Satterthwaite (1983), this literature has studied how quickly inefficiencies arising from private information vanish as the numbers of buyers and sellers increase. Our paper relates to this literature by providing a dynamic perspective, showing that benefits from increasing market thickness accrue even absent incentive problems, and by providing a measure of market thickness that is applicable across different mechanisms and environments.

The second strand is the recent literature on dynamic matching and mechanism design. The notions of periodic ex post incentive compatibility and individual rationality used in our mechanism design analysis were introduced by Bergemann and Välimäki (2010). We study a problem with a dynamic population of agents with persistent types, as do Parkes and Singh (2003), Gershkov and Moldovanu (2010) and Board and Skrzypacz (2016). Our contribution relative to these papers is that our focus is on the structure of the optimal allocation rule and deriving the optimal degree of market thickness. Within the dynamic matching literature, our paper is closest to Baccara et al. (2020), which, motivated by the problem of matching children and parents in an adoption "market," considers a dynamic, two-sided matching problem. We adhere to the standard assumptions in the dynamic mechanism design literature

[3]The performance of the second-best mechanism improves if one of the trading agents is known to be efficient, which explains why the gains from increasing market thickness are larger when types are traders' private information.

of geometric discounting and quasilinear payoffs,[4] which allows us to study a broad range of alternative questions and setups. The tradeoff between the benefits of increasing market thickness and the cost of delay has also been studied by Akbarpour et al. (2020) who, building on Ünver (2010) and Anderson et al. (2017), study efficiency in a dynamic matching model in which exchange possibilities have a network structure. This tradeoff also naturally arises in many market microstructure models (see, among other others, Vayanos (1999), Rostek and Weretka (2015) and Du and Zhu (2017)). Our paper complements this literature by deriving the optimal trading mechanism.

The remainder of this paper is organized as follows. Section 2 introduces the model. In Section 3 we characterize the optimal market clearing policy and optimal market thickness under a social planner with complete information. Section 4 deals with implementation when types are traders' private information. Section 5 concludes the paper. All proofs are provided in the appendix.

# 2 Model

We consider a discrete-time, infinite-horizon version of the classical bilateral trade problem of Myerson and Satterthwaite (1983) where a single buyer and seller arrive in each period. Each buyer (seller) has quasilinear utility and demands (supplies) at most one unit. The value of agents' outside option of not participating is zero. All agents and the social planner are risk-neutral geometric discounters, with a common discount factor $\delta \in [0, 1)$. Of course, this means that agents who arrive earlier are given more weight in the planner's objective.

For most of this paper we assume that agents draw their persistent types independently from symmetric, binary distributions. Buyers draw their values from a distribution with support $\{\underline{v}, \overline{v}\}$ and sellers draw their costs from a distribution with support $\{\underline{c}, \overline{c}\}$. We refer to buyers of type $\overline{v}$ and sellers of type $\underline{c}$ as *efficient* and buyers of type $\underline{v}$ and sellers of type $\overline{c}$ as *suboptimal*. We assume that $\Pr(c = \underline{c}) = \Pr(v = \overline{v}) = w$ for some $w \in (0, 1)$ and that types are symmetric in the sense that $\overline{v} - \overline{c} = \Delta$ and $\underline{v} - \underline{c} = \Delta$. So that we have a non-trivial problem we assume that $\overline{v} > \overline{c} > \underline{v} > \underline{c}$. Introducing the normalizations $\overline{v} = 1$ and $\underline{c} = 0$, this then implies that $\Delta \in (0, \frac{1}{2})$. We refer to $(\overline{v}, \underline{c})$ as an *efficient pair* or an *efficient trade* and $(\overline{v}, \overline{c})$ and $(\underline{v}, \underline{c})$ as *suboptimal pairs* or *suboptimal trades*.

For most of this paper we consider the problem faced by a social planner that maximizes expected discounted social surplus. Two static benchmarks are useful. The first of these is the Walrasian market involving a continuum of buyers and a continuum of sellers of equal

[4]See, for example, Athey and Miller (2007), Bergemann and Välimäki (2010), Athey and Segal (2013), Pavan et al. (2014) and Skrzypacz and Toikka (2015).

mass. Here, only efficient trades are executed under the efficient allocation. The market-clearing price $p = \frac{1}{2}$ implements this allocation, the probability $w$ represents the fraction of agents who trade, and social surplus per buyer-seller pair is $S_\infty := w$. The second benchmark we consider is ex post efficient bilateral trade involving a single buyer and seller. Since a trade is ex post efficient if and only if it involves an efficient trader, expected social surplus is given by $S_0 := w^2 + 2w(1-w)\Delta$. From the perspective of a perfectly patient social planner, the difference $S_\infty - S_0 = w(1-w)(1-2\Delta) > 0$ represents the maximum gain per buyer-seller pair (or, equivalently, per period) that can be achieved by increasing market thickness.

Many of our main results extend to continuous type distributions, as is more commonly seen in static mechanism design settings. Whenever we consider such a setting we assume that buyers (sellers) draw their values (costs) independently from the absolutely continuous distribution $F$ ($G$) whose density $f$ ($g$) has full support on $[0, 1]$.

# 3 Optimal market clearing

We now study the problem faced by a social planner that maximizes expected discounted social surplus. For the purpose of this section, we assume that the social planner knows the types of arriving traders.

## 3.1 Threshold policies and their properties

Whenever a suboptimal pair arrives, the social planner faces the following tradeoff. Executing this trade yields an immediate payoff of $\Delta$. On the other hand, storing the efficient trader means that when an efficient trader on the other side of the market eventually arrives as part of a suboptimal pair, the social planner can execute an efficient trade of value 1, instead of another suboptimal trade of value $\Delta$.[5] The market clearing policy that maximizes expected discounted social surplus optimally addresses this tradeoff.

Clearly, any efficient $(\overline{v}, \underline{c})$ trade is executed as soon as it becomes available because this yields the maximum gain from trade and delay is costly. Moreover, it is never optimal to store suboptimal types as the new arrivals in any future period cannot be worse than these types. The optimal policy therefore reduces to determining when the social planner should store efficient types (buyers *or* sellers) and when suboptimal trades should be executed.[6] Without

[5] By construction, if two suboptimal pairs $(\overline{v}, \overline{c})$ and $(\underline{v}, \underline{c})$ are simultaneously present the social planner's payoff is increased by rematching these pairs to create an efficient trade.

[6] There is a naturally analogy to storing traders in dynamic limit order books, which, in practice, may face additional constraints. For example, it may not be possible to clear traders from the market in a discriminatory fashion, as stipulated here, in which case all traders need to be cleared simultaneously. An additional constraint may be that markets clear traders periodically. These alternatives forms of market

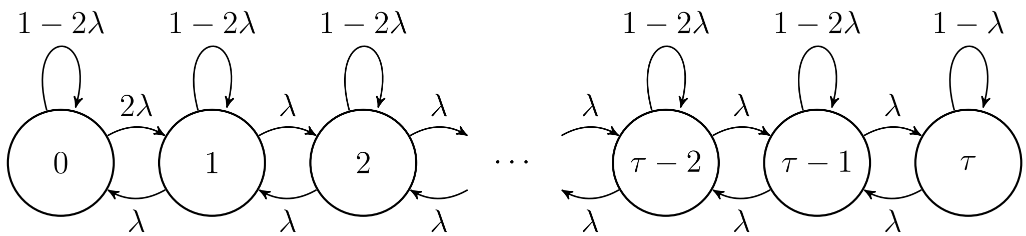


Figure 1: The Markov chain over the number of stored identical efficient traders induced by the threshold policy with threshold $\tau$, where $\lambda = w(1-w)$.

loss of generality we can restrict attention to stationary, deterministic market clearing policies (see Appendix A.1 for the formal argument). This implies that if the social planner ever arrives at the decision of whether or not to store $n$ efficient traders, the social planner must store $n-1$ efficient traders with probability 1 under the optimal market clearing policy.

Putting all of these observations together, the optimal market clearing policy involves storing efficient traders that arrive as part of a suboptimal pair up to some *threshold* $\tau \in \mathbb{Z}_{\geq 0}$. By symmetry, the same threshold applies regardless of whether buyers or sellers are stored. We call such a policy a *threshold policy.* Whenever less than $\tau$ efficient traders are stored, this policy only executes efficient trades. Efficient types that arrive as part of a suboptimal pair either trade with a stored trader or are themselves stored. Suboptimal trades are executed only when $\tau$ efficient traders are stored and the arriving agents consist of an efficient type of the kind stored and a suboptimal type on the other side of the market.[7]

Formally, in Appendix A.1 we map the dynamic optimization problem faced by the social planner to a *Markov decision process* and define threshold policies with respect to this process. The threshold policy with threshold $\tau$ induces a Markov chain $\{Y_t\}_{t\in\mathbb{N}}$ over $\{0,\ldots,\tau\}$, the number of stored identical efficient traders. As is illustrated in Figure 1, $\{Y_t\}_{t\in\mathbb{N}}$ is a finite birth-and-death process. We then have the following theorem.

**Theorem 1.** *The optimal market clearing policy is a threshold policy with threshold* $\tau \in \mathbb{Z}_{\geq 0}$. *The stationary distribution* $\boldsymbol{\kappa}$ *of the Markov chain* $\{Y_t\}_{t\in\mathbb{N}}$ *under this policy is*

$$\kappa_0 := \mathbb{P}(Y_t = 0) = \frac{1}{2\tau+1}, \quad \textit{and for all } i \in \{1,\ldots,\tau\}, \quad \kappa_i := \mathbb{P}(Y_t = i) = \frac{2}{2\tau+1}.$$

clearing are analyzed in a companion paper (Loertscher and Muir, 2021b), which shows that these alternatives provide a good approximation of the optimal policy when $\delta$ is sufficiently large.

[7]An alternative but equivalent way of expressing the trades that are executed under a threshold policy when the threshold is reached is to say that if, say, $\tau$ efficient buyers are stored and an efficient buyer arrives, then the arriving seller trades independently of whether it is efficient or suboptimal.

## 3.2 Measuring market thickness

We now characterize the *thickness* of the markets that are induced by threshold policies. While using the storage threshold itself as a measure of market thickness might seem like a natural approach, there are two problems with this. First, the optimality of threshold polices is specific to our dynamic setting with binary types. Consequently, measuring market thickness by the optimal threshold does not permit comparisons across different environments and policies. Second, the threshold is an ordinal measure and does not quantify market thickness relative to the two static benchmarks of ex post efficient bilateral trade and a Walrasian market involving a continuum of traders.

With that in mind, we now propose an alternative measure of market thickness. For a given threshold policy with threshold $\tau$, we let $S_\tau$ denote the average surplus per period under the stationary distribution. This is the standard welfare criterion for a perfectly patient social planner. Consequently,

$$\mathcal{T}_\tau := \frac{S_\tau - S_0}{S_\infty - S_0} \tag{1}$$

captures the proportion of the maximum gain from increasing market thickness that is achieved with a storage threshold of $\tau$. This provides a cardinal measure of market thickness that is independent of the discount factor $\delta$. In Section 3.3 this allows us to perform meaningful comparative statics concerning the *optimal* degree of market thickness. Of course, the complete information assumption underlying $\mathcal{T}_\tau$ is best thought of as a theoretical benchmark. With that in mind, in Section 4.3 we extend the analysis to second-best mechanisms that account for agents' private information and show that the analogously defined measure of market thickness that is based on second-best mechanisms, denoted $\mathcal{T}_\tau^{SB}$, is never smaller than $\mathcal{T}_\tau$.

All that remains is to compute $S_\tau$. Letting $s_\tau(y)$ denote the average payoff in each state $y \in \{0, \dots, \tau\}$, we have $S_\tau = \sum_{y=0}^{\tau} \kappa_y s_\tau(y)$. For $\tau = 0$, we have $s_0(0) = w^2 + 2w(1-w)\Delta$, $\kappa_0 = 1$ and $S_0 = w^2 + 2w(1-w)\Delta$. Naturally, this corresponds to surplus under ex post efficiency in the static bilateral trade benchmark. For $\tau > 0$, an efficient trade is executed in state 0 only if an efficient pair arrives and $s_\tau(0) = w^2$. For all states $y \in \{1, \dots, \tau - 1\}$, an efficient trade is executed whenever an efficient trader arrives on the appropriate side of the market, which implies $s_\tau(y) = w$. Thus, expected social surplus is the same as surplus per buyer-seller pair under the Walrasian benchmark in these states. In state $\tau$, a suboptimal trade is also executed with probability $w(1-w)$, implying $s_\tau(\tau) = w + w(1-w)\Delta$. Putting all of this together yields $S_\tau = \frac{1}{2\tau+1}[2\tau w + w^2 + 2w(1-w)\Delta]$ and $\lim_{\tau\to\infty} S_\tau = S_\infty$. This in turn implies $S_\tau - S_0 = \frac{2\tau}{2\tau+1} w(1-w)(1-2\Delta) = \frac{2\tau}{2\tau+1}(S_\infty - S_0)$, giving us the following proposition.

**Proposition 1.** *Under the threshold policy with threshold $\tau \in \mathbb{Z}_{\geq 0}$, we have $\mathcal{T}_\tau = \frac{2\tau}{2\tau+1}$.*

An equivalent measure of market thickness is provided by taking our previous measure of market thickness (1) and, for $\nu \in \mathbb{Z}_{\geq 0} \cup \{\infty\}$, replacing $S_\nu$ with the per period probability $\mathbb{P}_S(\nu)$ of executing a suboptimal trade under the stationary distribution. Suboptimal trades are never executed in the Walrasian benchmark so we have $\mathbb{P}_S(\infty) = 0$. Using $\mathbb{P}_S(0) = 2w(1-w)$ and $\mathbb{P}_S(\tau) = \frac{2}{2\tau+1}w(1-w)$, implies $\mathbb{P}_S(0) - \mathbb{P}_S(\tau) = \frac{4\tau}{2\tau+1}w(1-w)$ and

$$\mathcal{T}_\tau = \frac{\mathbb{P}_S(0) - \mathbb{P}_S(\tau)}{\mathbb{P}_S(0)} = \frac{2\tau}{2\tau+1}. \tag{2}$$

We return to this equivalent measure of market thickness in Section 4.1, where we will show that $\mathbb{P}_S(\nu)$ also corresponds to the probability that an arriving trader has a price impact.

Interestingly, $\mathcal{T}_\tau$ provides a meaningful measure of market thickness beyond our symmetric binary types setting. Suppose, temporarily, that traders draw their types from continuous distributions $F$ and $G$, as introduced in Section 2. The Walrasian price $p \in (0,1)$ satisfies $1 - F(p) = G(p)$. Setting $w = 1 - F(p) = G(p)$, we can approximate this environment using our binary types setting by categorizing buyers (sellers) as *efficient* if their value (cost) lies above (below) the Walrasian price $p$ and *suboptimal* otherwise. Ex post efficient trade between an efficient type and a suboptimal type then generates expected social surplus of

$$\Delta = \mathbb{E}_{v,c}[\max\{v - c, 0\} | v, c \geq p \text{ or } v, c \leq p].$$

Setting $\overline{v} = \mathbb{E}[v|p \leq v]$ and $\underline{c} = \mathbb{E}[c|c \leq p]$, the expected gain from trade between two efficient traders is $\overline{v} - \underline{c}$.[8] We can then consider a *two-class threshold* policy with threshold $\tau \in \mathbb{Z}_{\geq 0}$. This policy separates buyers and sellers into two classes as just described and implements the threshold policy with threshold $\tau$ with respect to these classes.

**Proposition 2.** *Assume that traders draw their types from continuous distributions $F$ and $G$, as introduced in Section 2. Then there exists a two-class threshold policy such that $\mathcal{T}_\tau$ captures the proportion of the maximum possible gains from increasing market thickness that is achieved by storing up to $\tau$ efficient traders.*

Proposition 1 shows that even for small values of $\tau$, our measure of market thickness is close to that of the Walrasian benchmark. For example, for $\tau = 1$, we have $\mathcal{T}_1 = \frac{2}{3}$. Proposition 2 shows that this result is not specific to our setting with symmetric binary

[8] For example, we have $w = \frac{1}{2}$ if $F = G$, and if $F$ and $G$ are both uniform on $[0,1]$, we have $\overline{v} = \frac{3}{4}$, $\underline{c} = \frac{1}{4}$ and $\Delta = \frac{1}{12}$.

types. Indeed, $\mathcal{T}_\tau$ is detail-free in the sense that it is the same for *any* type distributions if the social planner uses a two-class threshold policy that categorizes buyers and sellers according to whether their types are above or below the Walrasian price $p$. What the symmetric binary distribution then affords us is not so much that the benefits from increasing market thickness accrue quickly in $\tau$ but rather that these two-class threshold policies are optimal.

To illustrate how this measure applies to alternative settings and policies, we can consider a static setting in the tradition of Gresik and Satterthwaite (1989), where one varies market thickness by varying the number of buyer-seller pairs. Denoting by $S_{(n)}$ the per pair expected social surplus with $n+1$ pairs present, we have $S_{(0)} = w^2 + 2w(1-w)\Delta = S_0$ and $S_{(\infty)} = w = S_\infty$. Accordingly, our measure of market thickness applied to this setting would be $\mathcal{T}_{(n)} = \frac{S_{(n)} - S_{(0)}}{S_{(\infty)} - S_{(0)}}$.[9] As we illustrate in Section 4.3, this measure can also easily account for additional constraints such as incentive compatibility, individual rationality and budget balance, by simply computing social surplus under these constraints in each setting.

### 3.3 Optimally thick markets

Using dynamic programming we can analytically characterize the *optimal* threshold (see Lemma A1), which we denote by $\tau^*$.[10] Given $\tau^*$, the *optimal market thickness* is $\mathcal{T}^* := \mathcal{T}_{\tau^*}$.

**Proposition 3.** *$\tau^*$ and $\mathcal{T}^*$ increase in $\delta$ and $w(1-w)$ and decrease in $\Delta$.*

Unsurprisingly, decreasing $\Delta$ or increasing $\delta$ decreases the opportunity cost of storing efficient traders, resulting in an increase in $\tau^*$ and, thereby, in $\mathcal{T}^*$. Moreover, the very purpose of storing an efficient trader is to form and execute an efficient trade when an efficient trader on the other side of the market eventually arrives as part of a suboptimal pair. The probability that such a pair arrives in any given period being $w(1-w)$, it follows that optimal market thickness is increasing in this probability and maximized when $w = \frac{1}{2}$.

Letting $\rho = \frac{w(1-w)(1-2\Delta)}{\Delta}$, we show in Appendix 3 that

$$\tau^* \geq 1 \quad \Leftrightarrow \quad \delta \geq \delta^* := \frac{1}{1+\rho}. \tag{3}$$

[9] Straightforward calculations show that, for example, $S_{(1)} = w^2(1-w)^2(1-2\Delta) + w^2(1-2\Delta) + 2w\Delta$, which implies $\mathcal{T}_{(1)} = w(1-w)$. With continuous distributions, there are additional sorting benefits for $n \geq 2$ beyond those captured by the binary types approximation. Corollary 1 in Loertscher and Muir (2021a) derives $\mathcal{T}_{(n)}$ for the case when the buyer's and seller's distributions are the uniform on $[0,1]$ and shows that, using the present notation, $\mathcal{T}_{(n)} = \frac{2n}{2n+3}$. Note that $\mathcal{T}_{(1)} = \frac{2}{5}$, which is larger than the $\frac{1}{4}$ obtained for the uniform distribution with binary types.

[10] Due to the discrete nature of our setting, multiple thresholds can be optimal for some knife-edge cases. For ease of exposition, we set $\tau^*$ to be the largest optimal storage threshold throughout.

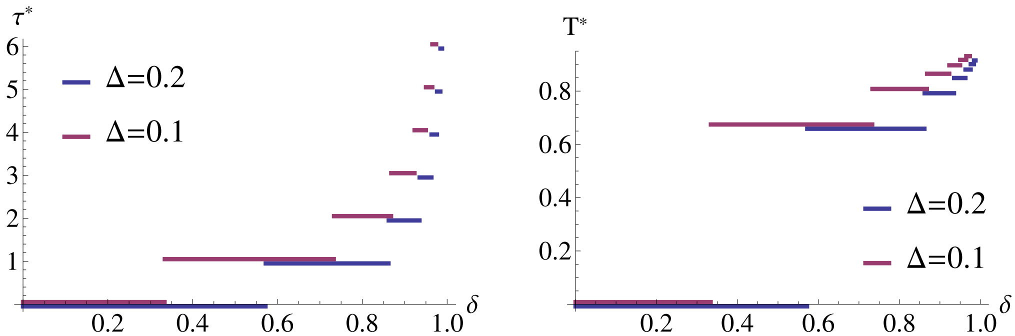


Figure 2: A numerical illustration of $\tau^*$ (left panel) and $\mathcal{T}^*$ (right panel) as a function of $\delta$ for $w = 0.5$ and $\Delta = 0.1, 0.2$.

Intuitively, by storing an efficient trader, the planner eventually executes an efficient trade rather than two suboptimal trades, yielding an undiscounted net return of $1 - 2\Delta$.[11] The decision to store an efficient trader therefore involves trading off an immediate reward of $\Delta$ and realizing the return of $1 - 2\Delta$ with probability $w(1 - w)$ in any period in which this trader is stored. Hence, $\rho$ can be thought of as the per-period rate of return associated with storing an efficient trade and $\delta^*$ is the discount factor that makes the social planner indifferent between earning a payoff of 1 and a payoff of $\delta(1 + \rho)$.

Optimally thick markets are thin when measured by the maximum number of traders present. For example, for the parameterizations displayed in Figure 2, $\tau^* \leq 6$ for all $\delta \leq 0.95$. While $\tau^* = 6$ might seem small relative to the Walrasian benchmark widely used to describe and analyze competitive markets, this threshold achieves more than 92% of the maximally achievable increase in market thickness. Thus, although a market with a storage threshold of six may seem thin against the backdrop of a market with a continuum of traders, its market thickness is remarkably close to that of the large Walrasian market. The marginal product $\mathcal{T}_{\tau+1} - \mathcal{T}_\tau = \frac{2}{1+2(1+\tau)^2}$ also decreases quickly in $\tau$, reflecting the steep increase in the expected number of periods until the gain from storing additional traders is realized. This shows that there are steeply diminishing returns from storing additional efficient traders and one might intuitively expect that market thickness grows slowly in $\delta$. Proposition 4 formalizes this intuition and shows that optimal market thickness grows slowly—at rate $(1 - \delta)^{-1}$—as $\delta \to 1$.

**Proposition 4.** *Letting* $\overline{\tau}(\delta, \Delta) = \frac{\log\left(\frac{1-\Delta}{\Delta}\right)}{1-\delta} + \frac{\log\left(\frac{\Delta}{1-\Delta}\right)}{2}$ *we have* $\tau^* \leq \overline{\tau}(\delta, \Delta) + O(1 - \delta)$.

[11] Heuristically, storing an efficient trader ensures that eventually a buyer of type $\overline{v}$ and a seller of type $\underline{c}$, that would have otherwise participated in suboptimal trades in different periods, are simultaneously present.

# 4 Implementation with private information

In this section we assume that values and costs are private information of the agents.

## 4.1 Posted-price mechanisms

We first study a posted-price implementation of the optimal market clearing policy. Posted-price mechanisms are simple and widely used in practice. As we will see shortly, provided $\tau^* \geq 1$, a posted-price implementation of the optimal market clearing policy is possible and endows traders with dominant strategies. Throughout this section we need to distinguish states where efficient buyers are stored from states where efficient sellers are stored. For a given threshold policy with threshold $\tau$ we now let the Markov chain $\{Z_t\}_{t \in \mathbb{N}}$ over $\{-\tau, \dots, \tau\}$ represent the number of stored efficient traders, where positive states $z > 0$ indicate that buyers are stored and negative states $z < 0$ indicate that sellers are stored.

Assuming $\tau^* \geq 1$, many pricing rules $p : \{-z, \dots, z\} \to [\underline{c}, \overline{v}]$ can be used as part of a posted-price mechanism to implement the optimal policy. Under this policy only efficient trades are executed in the states $z \in \{-\tau^* + 1, \dots, \tau^* - 1\}$. The planner can ensure that no other trades are executed by posting a price of $p(z) \in [\underline{v}, \overline{c}]$. In the state $z = \tau^*$ both efficient and suboptimal $(\overline{v}, \overline{c})$ trades are executed under the optimal policy. These trades can be executed by posting a price of $p(\tau^*) \in [\overline{c}, \overline{v}]$. Similarly, in the state $z = -\tau^*$ both efficient and suboptimal $(\underline{v}, \underline{c})$ trades can be executed by posting a price of $p(-\tau^*) \in [\underline{c}, \underline{v}]$. A queueing protocol (or tie-breaking rule) specifies the order in which efficient traders are cleared from the market when multiple efficient traders of the same type are present in a single period.[12] Accordingly, a posted-price mechanism that implements the optimal policy is characterized by the chosen pricing rule $p$, the threshold $\tau^*$ and the chosen queueing protocol.

Formally, under a *posted-price mechanism* that implements the optimal market clearing policy, the social planner posts a state-dependent price $p(z) \in [\underline{c}, \overline{v}]$ at the start of each period. Each trader reports a type to the social planner upon arrival and does not make any subsequent reports. Specifically, traders observe the state $z$ and the posted price $p(z)$ in the period they arrive. Buyers report a type $v \in \{\underline{v}, \overline{v}\}$ and sellers report a type $c \in \{\underline{c}, \overline{c}\}$. Agents have also the option of not participating in the mechanism, meaning that they permanently exit the market.

Next, the social planner implements the optimal market clearing policy with respect to the reports of all participating traders, using the queueing protocol to break any ties.

[12] For example, in period $t$ a first-in-first-out queueing protocol gives the lowest priority to the agents that arrived in period $t$ and clears any stored efficient traders in the order in which they arrived. A last-in-first-out queueing protocol gives the highest priority to period $t$ agents and clears any stored efficient types according to a reversed order of arrival.

Importantly, any trades are executed at the posted price determined at the *start* of the period, *before* the arriving traders report their types. Under an appropriately chosen pricing rule and queueing protocol, truthful reporting is a weakly dominant strategy for all traders.[13]

**Proposition 5.** *Suppose that $\tau^* \geq 1$ and, for $z \in \{-\tau^* + 1, \ldots, \tau^* - 1\}$, let $p$ be a pricing rule satisfying $p(z) \in [\underline{v}, \overline{c}]$, $p(\tau^*) \in [\overline{c}, \overline{v}]$ and $p(-\tau^*) \in [\underline{c}, \underline{v}]$. Then the optimal market clearing policy can be implemented using a posted-price mechanism with pricing rule $p$ and a last-in-first-out queueing protocol, where truthful reporting is a weakly dominant strategy for all traders.*

Allowing the planner to post separate prices $p^B$ for buyers and $p^S$ for sellers, the posted-price mechanism with $p^B(z) = \overline{v}$ and $p^S(z) = \underline{c}$ for $z \in \{-\tau^* + 1, \ldots, \tau^* - 1\}$, $p^B(\tau^*) = \overline{v}$, $p^S(\tau^*) = \overline{c}$, $p^B(-\tau^*) = \underline{v}$ and $p^S(-\tau^*) = \underline{c}$ and a last-in-first-out queueing protocol implements the optimal market clearing policy and, subject to allocating efficiently, maximizes the profit of the planner. Under the last-in-first out queueing protocol specified here and in Proposition 5, agents only trade with positive probability at the price that was posted in the period they arrived. This endows all traders with truthful reporting as a weakly dominant strategy because it ensures that they essentially act as price-takers.

An important implication of Proposition 5 is that, provided $\tau^* \geq 1$, the optimal market clearing policy can be implemented with a budget-balanced mechanism.

**Corollary 1.** *If $\delta \geq \delta^*$, the optimal market clearing policy can be implemented without running a deficit.*

Whenever $\delta \geq \delta^*$, suboptimal trades are executed with positive probability only in states where the storage threshold $\tau^* \geq 1$ is reached. The planner can ensure that only efficient trades are executed in all other states $z$ by posting prices of $p^B = \overline{v}$ and $p^S = \underline{c}$ for buyers and sellers, respectively. This is true regardless of whether arriving buyers and sellers are privately informed about their values and costs and shows that no information rents need to be paid to efficient traders in these states. In states where the storage threshold is reached, the planner induces trade for suboptimal types on one side of the market with positive

[13] There is also an indirect implementation where arriving traders indicate the quantities $q^B \in \{0, \frac{1}{2}, 1\}$ and $q^S \in \{0, \frac{1}{2}, 1\}$ that they are willing to trade at the current price, where $q^i = \frac{1}{2}$ indicates $i$ is indifferent between trading and not trading. Arriving traders *bid sincerely* if the arriving buyer with value $v$ bids $q^B(p) = 1$ if $v > p$, $q^B(p) = \frac{1}{2}$ if $v = p$ and $q^B(p) = 0$ otherwise and the arriving seller with cost $c$ bids $q^S(p) = 1$ if $c < p$, $q^S(p) = \frac{1}{2}$ if $c = p$ and $q^S(p) = 0$ otherwise. If any buyers or sellers make a report that is not consistent with sincere bidding, we assume that these agents are immediately cleared from the market without trading. Under an appropriately chosen pricing rule and queueing protocol, sincere bidding is a weakly dominant strategy for all traders and the social planner can correctly infer the information needed to implement the optimal market clearing policy.

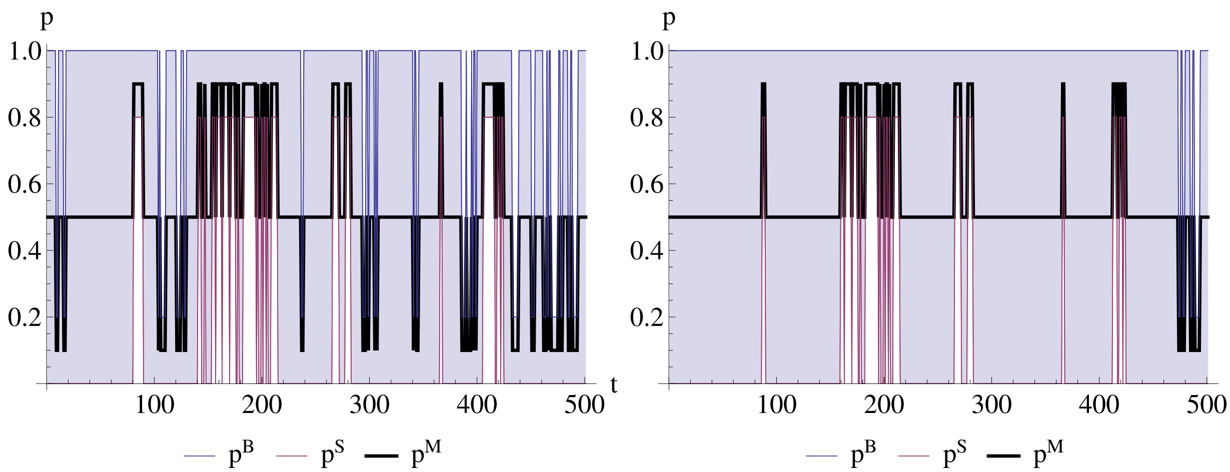


Figure 3: For $w = 0.5$ and $\Delta = 0.2$ and a single realization of the arrival process we plot the pricing rules $p^B$, $p^S$ and $p^M$, shading the bid-ask spread $p^B - p^S$. The left and right panels assume $\delta \in [0.94, 0.96]$ ($\tau^* = 3$) and $\delta \in [0.98, 0.99]$ ($\tau^* = 6$), respectively.

probability by posting a sufficiently favourable price for this side of the market.[14] While the impossibility result of Myerson and Satterthwaite (1983) may hold for the bilateral trade problem,[15] provided $\tau^* \geq 1$, the planner does not need to incur a deficit by simultaneously paying information rents to traders on both sides of the market in any state in our dynamic setting. Intuitively, the possibility of storing efficient types increases competition among these traders, reducing the information rents of these types in a manner that is reminiscent of the possibility results of Makowski and Mezzetti (1993) and Williams (1999).

Fixing a pricing rule we can also connect our measure of market thickness to the *price impact* of arriving agents. For example, consider the budget-balanced, midpoint pricing rule with $p^M(z) = \frac{1}{2}$ for $z \in \{-\tau^* + 1, \ldots, \tau^* - 1\}$, $p^M(\tau^*) = 1 - \frac{\Delta}{2}$ and $p^M(-\tau^*) = \frac{\Delta}{2}$. Letting $\mathbb{P}_I(\tau^*)$ denote the stationary per period probability that an agent's arrival changes the posted price from the Walrasian price of $\frac{1}{2}$ to one of the extremal prices of $\frac{\Delta}{2}$ or $1 - \frac{\Delta}{2}$, we have $\mathbb{P}_I(\tau^*) = \frac{2}{2\tau+1} w(1-w)$. This is precisely the per period probability $\mathbb{P}_S(\tau)$ of executing a suboptimal trade under the stationary distribution, and the price impact of arriving agents can therefore be connected to our measure of market thickness using (2).

Fixing a pricing rule also allows us to perform a quantitative analysis of the dynamic prices generated by the corresponding post-price mechanism (see Appendix C). Figure 3 provides an illustration of the price dynamics induced by the budget-balanced, midpoint pricing rule $p^M$, as well as the profit-maximizing posted-price implementation with the pricing rules

[14]If $\tau^*$ buyers are stored then the price on the seller's side of the market can be increased to $p^S = \bar{c}$ and if $\tau^*$ sellers are stored then the price on the buyer's side of the market can be decreased to $p^B = \underline{v}$.

[15]As we will see in Section 4.2 ex post efficient trade is not possible without running a deficit in the bilateral trade problem if and only if $w > 2\Delta$.

$p^B$ and $p^S$.

## 4.2 Profit maximization

We now move beyond posted-price mechanisms and consider the dynamic mechanism design problem faced by a profit-maximizing market maker. Once again arriving traders must be incentivized to both participate in the mechanism and reveal their private information. This requires that we continue to consider allocations rules which, in addition to specifying the market clearing policy, also specify a queueing protocol that is used to break ties.

For concreteness, we impose *periodic ex post incentive compatibility (P-IC)* and *periodic ex post individual rationality (P-IR)* constraints.These constraints require that truthful reporting and participation is optimal for the traders that arrive in period $t$, regardless of the history of reports up to $t-1$ and the report of the other trader that arrives in period $t$, assuming that all future traders report truthfully.

We defer the details of this analysis—including formally defining the class of direct mechanisms and the aforementioned notions of incentive compatibility and individual rationality and establishing a version of the payoff equivalence theorem—to Appendix B.3. The gist of this analysis is the following. By adapting standard mechanism design arguments, one can compute the virtual valuation and cost functions $\Phi$ and $\Gamma$ of buyers and sellers, respectively. These functions are given by

$$\Phi(\underline{v}) := \underline{v} - \frac{w}{1-w}(\overline{v} - \underline{v}), \quad \Phi(\overline{v}) = \overline{v}, \quad \Gamma(\underline{c}) = \underline{c} \quad \text{and} \quad \Gamma(\overline{c}) = \overline{c} + \frac{w}{1-w}(\overline{c} - \underline{c}), \tag{4}$$

and are such that whenever the market maker executes a trade of $(v, c)$, the immediate contribution to its profit is $\Phi(v) - \Gamma(c)$. Aside from the fact that the value of a suboptimal trade is now given by $\Delta_{PM} := \Phi(\overline{v}) - \Gamma(\overline{c}) = \Phi(\underline{v}) - \Gamma(\underline{c})$ rather than $\Delta$, the profit-maximizing market maker solves the same dynamic optimization problem as the social planner. Applying our analysis from Section 3 with $\Delta$ replaced by $\Delta_{PM}$ implies that threshold policies are still optimal and allows us to characterize the corresponding optimal threshold $\tau^*_{PM}$.[16]

Since expected discounted profit under the profit-maximizing mechanism can be computed directly from the value $\Delta_{PM}$ of a suboptimal trade and the optimal threshold policy, this also means that the choice of queueing protocol is irrelevant.[17] Finally, checking that

[16]To be precise, the analysis of Section 3 applies if $\Delta_{PM} > 0$. If $\Delta_{PM} \leq 0$ then the optimal market clearing policy is trivial and involves storing an unbounded number of efficient traders, which we can represent by setting $\tau^*_{PM} = \infty$.

[17]While the choice of queueing protocol played an important role for posted-price mechanisms, here the transfers paid to arriving traders are more flexible. The market maker can set transfers that elicit the types of arriving traders without affecting expected discounted profit, regardless of the chosen queueing protocol.

the optimal threshold policy always gives rise to an allocation rule that is monotone in the appropriate sense,[18] we have the following theorem.

**Theorem 2.** *Coupling the threshold policy with threshold $\tau^*_{PM}$ with any queueing protocol yields a monotone allocation rule $\boldsymbol{Q}^*_{PM}$ that maximizes the market maker's profit, subject to the P-IC and P-IR constraints. There is a profit-maximizing mechanism, satisfying P-IC and P-IR, that can be implemented using transfers that traders pay or receive upon arrival.*

A simple illustrative example provided at the end of the proof of Theorem 2 considers a queueing protocol that always prioritizes the arriving agents and otherwise uses a first-in-first-out protocol. A payment rule that implements the corresponding allocation rule and involves transfers that traders always pay or receive upon arrival is then as follows. A buyer of type $\underline{v}$ makes a payment of 0 unless it arrives with a seller of type $\underline{c}$ to a market in the state $z = -\tau^*_{PM}$, in which case the buyer trades and makes a payment of $\underline{v}$. Likewise, a seller of type $\overline{c}$ is paid 0 unless it arrives with a buyer of type $\overline{v}$ to a market in state $z = \tau^*_{PM}$, in which case the seller trades and receives a payment of $\overline{c}$. The payment of an arriving efficient buyer is $\overline{v}$ in the states $z \in \{-\tau^*_{PM}+1, \dots, -1, \tau^*_{PM}\}$ and $\underline{v}$ in the state $z = -\tau^*_{PM}$. In the states $z \in \{0, \dots, \tau^*_{PM}-1\}$ the buyer is stored, its expected discounted allocation is $\left(\frac{\delta\lambda}{1-\lambda+\delta\lambda}\right)^{z+1}$ and it makes a payment of $\overline{v}\left(\frac{\delta\lambda}{1-\lambda+\delta\lambda}\right)^{z+1}$. Analogously, the payment to an arriving efficient seller is $\underline{c}$ in the states $z \in \{-\tau^*_{PM}, 1, \dots, \tau^*_{PM}-1\}$ and $\overline{c}$ in the state $z = -\tau^*_{PM}$. In the states $z \in \{-\tau^*_{PM}+1, \dots, -1\}$ the seller is stored, its expected discounted allocation is $\left(\frac{\delta\lambda}{1-\lambda+\delta\lambda}\right)^{1-z}$ and it receives a payment of $\underline{c}\left(\frac{\delta\lambda}{1-\lambda+\delta\lambda}\right)^{1-z}$ upon arrival.

Since

$$\Delta_{PM} = \Delta - \frac{w}{1-w}(1-\Delta) < \Delta,$$

suboptimal trades are less valuable to the profit-maximizing market marker than to the social planner. Combining this fact with the comparative statics from Proposition 3 then shows that a profit-maximizing market maker induces an inefficiently thick market, relative to the social planner. This is reminiscent of Hotelling's (1931) finding that a monopolist extracts an exhaustible resource at a slower rate than a perfectly competitive industry.[19] More formally, taking $\mathcal{T}^*_{PM} := \mathcal{T}_{\tau^*_{PM}}$, we have the following corollary.

**Corollary 2.** *$\tau^*_{PM} \geq \tau^*$ and, consequently, $\mathcal{T}^*_{PM} \geq \mathcal{T}^*$.*

Corollary 2 also has implications for the regulation of profit-maximizing market makers. Suppose that a regulator can observe and tax these profits and that a specific tax of $\sigma > 0$ per

[18]Under a *monotone* allocation rule the expected discounted allocations of arriving buyers (sellers) are increasing (decreasing) in their reported types. In Appendix B.3 we show that an allocation rule can be implemented as part of a P-IC and P-IR mechanism if and only if it is monotone.

[19]Inefficiently few matches also take place under profit maximization in the dynamic matching model of Fershtman and Pavan (2019).

unit traded is imposed. The relative value of a suboptimal trade to the market maker then becomes $\frac{\Delta_{PM}-\sigma}{1-\sigma} < \Delta_{PM}$. Consequently, the market maker will induce an even thicker market in equilibrium, further increasing the allocative inefficiencies associated with such a market maker. In contrast, an ad valorem tax levied as a percentage on the market makers's profit does not affect the relative value of a suboptimal trade and equilibrium market thickness. Consequently, and in contrast to static market settings, ad valorem taxes are superior to specific taxes in markets in our dynamic setting.

## 4.3 Second-best mechanisms

We conclude this section by discussing the market thickness implications of privately informed traders under a social planner. Utilizing our expression for $\Delta_{PM}$ to compute profit under ex post efficient bilateral trade yields $w^2 + 2w(1-w)\Delta_{PM} = w(2\Delta - w)$. Whenever $w > 2\Delta$, this expression is negative and the ex post efficient outcome is not implementable. Moreover, we know from Section 4.1 that the optimal market clearing policy can be implemented without running a deficit whenever $\tau^* \geq 1$. Therefore, even with privately informed traders, second-best mechanisms only come into play when $\tau = 0$ and $w > 2\Delta$. Letting $S_0^{SB}$ denote welfare in the static bilateral trade setting under the second-best mechanism, when traders are privately informed our measure of market thickness becomes $\mathcal{T}_\tau^{SB} = \frac{S_\tau - S_0^{SB}}{S_\infty - S_0^{SB}}$.

When ex post efficiency is not possible under static bilateral trade, this directly impacts our measure of market thickness because, for any $\tau \geq 1$, $S_0^{SB} < S_0$ immediately implies that $\mathcal{T}_\tau^{SB} > \mathcal{T}_\tau$. There is also an indirect effect: since suboptimal trades are less valuable in the bilateral trade setting under the second-best mechanism, the social planner is more inclined to store at least one efficient trader. Relative to the complete information case, each of these effects increases our measure of market thickness. Formally, let $\tau_{SB}^*$ denote the optimal storage threshold with privately informed traders. Whenever $\delta > \delta^*$, the benefits and costs of storing multiple traders are the same with and without private information and we have $\tau^* = \tau_{SB}^*$. Analogously to how $\delta^*$ is defined in the complete information setting, let $\delta_{SB}^*$ be the value of the discount factor such that $\tau_{SB}^* = 0$ if and only if $\delta < \delta_{SB}^*$. We then have the following proposition.

**Proposition 6.** *We have $\mathcal{T}_\tau^{SB} \geq \mathcal{T}_\tau$, with strict inequality if $w > 2\Delta$ and $\tau \geq 1$. Furthermore, if $w > 2\Delta$ then $\delta_{SB}^* < \delta^*$ and $\tau_{SB}^* = 0$ for $\delta < \delta_{SB}^*$, $\tau_{SB}^* = 1$ for $\delta \in [\delta_{SB}^*, \delta^*)$ and $\tau_{SB}^* = \tau^*$ otherwise.*

Analogous results concerning the market thickness effects of private information also hold when traders draw their types from the continuous distributions $F$ and $G$, as introduced in Section 2. In this case the effects are more nuanced. Naturally, as shown by Myerson and

Satterthwaite (1983), there is an incentive problem when $\tau = 0$. However, there is also an incentive problem when $\tau > 0$ because the planner needs to accommodate the execution of suboptimal trades in states where the storage threshold is reached. In Appendix B.7 we provide a detailed construction of a *two-class threshold mechanism* that implements the two-class threshold policies introduced in Section 3.2, assuming that trader types are private and that the distributions $F$ and $G$ exhibit increasing virtual value and cost functions. This mechanism runs an expected deficit of zero in every period and executes only efficient trades when fewer than $\tau$ efficient traders are stored by simply posting the Walrasian price $p$. When the storage threshold is reached, a second-best mechanism is implemented whenever the arriving agent of the type stored is an efficient trader.[20] Let $S_\tau^{TC}$ denote expected welfare per period under the two-class threshold mechanism with threshold $\tau$ and $S_0^{SB}$ denote expected welfare under the second-best mechanism in the bilateral trade problem. The following proposition then shows that a result analogous to Proposition 6 holds for the market thickness measure $\mathcal{T}_\tau^{TC} = \frac{S_\tau^{TC} - S_0^{SB}}{S_\infty - S_0^{SB}}$.

**Proposition 7.** *The two-class threshold mechanism satisfies P-IC and P-IR and balances the budget in expectation in every period. Moreover, for any $\tau \geq 1$, we have $\mathcal{T}_\tau^{TC} > \mathcal{T}_\tau$.*

Intuitively, the last statement of Proposition 7 holds because a static, second-best mechanism is only required when the storage threshold is reached and a trade that potentially involves a suboptimal type needs to be executed. Positive selection ensures that this second-best mechanism performs better than the second-best mechanism that is run in every period when $\tau = 0$. Under the former mechanism the planner knows that a trader on a given side of the market is efficient, whereas the latter mechanism uses the prior distributions.

# 5 Conclusions

In dynamic environments in which traders arrive over time, optimally thick markets balance the gains from trade associated with accumulating traders against the cost of delay. In a dynamic extension of Myerson and Satterthwaite (1983) in which traders draw their types from symmetric binary distributions, we show that the optimal market clearing policy of a social planner is characterized by the maximum number of efficient traders that are stored in any period. Optimally thick markets are thin in the sense that they typically involve storing a small number of traders. Yet, a market clearing policy that involves storing at most one efficient trader achieves two-thirds of the maximum gain associated with increasing market thickness. Accordingly, the textbook model of a competitive market involving a continuum

[20] This reflects the alternative interpretation of threshold policies given in footnote 7.

of traders may provide a good approximation for real-world markets that clear over time, and typically have few traders present at any given point in time. Whenever traders' types are private and the optimal market clearing policy involves storing at least one efficient trader, this policy can be implemented using a posted-price mechanism. Many of these results extend to the case in which traders draw their types from continuous distributions, allowing us to quantify the large gains that can be reaped by employing simple, coarse mechanisms in many dynamic environments.

## References


Akbarpour, M., S. Li, and S. O. Gharan (2020): "Thickness and Information in Dynamic Matching Markets," *Journal of Political Economy*, 128, 783–815.

Anderson, R., I. Ashlagi, D. Gamarnik, and Y. Kanoria (2017): "Efficient Dynamic Barter Exchange," *Operations Research*, 65, 1446–1459.

Athey, S. and D. A. Miller (2007): "Efficiency in repeated trade with hidden valuations," *Theoretical Economics*, 2, 299–354.

Athey, S. and I. Segal (2013): "An Efficient Dynamic Mechanism," *Econometrica*, 81, 2463–2485.

Baccara, M., S. Lee, and L. Yariv (2020): "Optimal Dynamic Matching," *Theoretical Economics*, 15, 1221–1278.

Bergemann, D. and J. Välimäki (2010): "The Dynamic Pivot Mechanism," *Econometrica*, 78, 771–789.

Board, S. and A. Skrzypacz (2016): "Revenue Management with Forward-Looking Buyers," *Journal of Political Economy*, 124, 1046–1087.

Börgers, T. (2015): *An Introduction to the Theory of Mechanism Design*, New York: Oxford University Press.

Cripps, M. W. and J. M. Swinkels (2006): "Efficiency of Large Double Auctions," *Econometrica*, 74, 47–92.

Du, S. and H. Zhu (2017): "What is the Optimal Trading Frequency in Financial Markets?" *Review of Economic Studies*, 84, 1606–1651.

Elkind, E. (2007): "Designing and Learning Optimal Finite Support Auctions," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, Society for Industrial and Applied Mathematics, 736–745.

Fershtman, D. and A. Pavan (2019): "Matching Auctions," Working paper.

Gershkov, A. and B. Moldovanu (2010): "Efficient Sequential Assignment with Incomplete Information," *Games and Economic Behavior*, 68, 144–154.

Gresik, T. and M. Satterthwaite (1989): "The Rate at which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms," *Journal of Economic Theory*, 48, 304–332.

Hotelling, H. (1931): "The Economics of Exhaustible Resources," *Journal of Political Economy*, 39, 137–175.

Hurwicz, L. (1972): "On Informationally Decentralized Systems," in *Decision and Organization*, ed. by C. B. McGuire and R. Radner, Amsterdam: North-Holland, 297–336.

Lauermann, S. (2013): "Dynamic Matching and Bargaining Games: A General Approach," *American Economic Review*, 103, 663–689.

Loertscher, S. and E. V. Muir (2021a): "The benefits of market thickness for niche products," Working paper.

——— (2021b): "Optimal dynamic market making," Working paper.

Makowski, L. and C. Mezzetti (1993): "The Possibility of Efficient Trade for Trading an Indivisible Object," *Journal of Economic Theory*, 59, 451–465.

McAfee, R. P. (1992): "A Dominant Strategy Double Auction," *Journal of Economic Theory*, 56, 434–450.

Myerson, R. (1981): "Optimal Auction Design," *Mathematics of Operations Research*, 6,

58–78.
Myerson, R. and M. Satterthwaite (1983): "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory*, 29, 265–281.
Parkes, D. C. and S. Singh (2003): "An MDP-Based Approach to Online Mechanism Design," in *Proceedings of the 17th Annual Conference on Neural Information Processing Systems (NIPS 03)*.
Pavan, A., I. Segal, and J. Toikka (2014): "Dynamic Mechanism Design: A Myersonian Approach," *Econometrica*, 82, 601–653.
Puterman, M. L. (1994): *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, New York: Wiley.
Rostek, M. and M. Weretka (2015): "Dynamic Thin Markets," *Review of Financial Studies*, 28, 2946–2992.
Rustichini, A., M. A. Satterthwaite, and S. R. Williams (1994): "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica*, 62, 1041–1063.
Satterthwaite, M. and A. Shneyerov (2007): "Dynamic Matching,Two-sided Incomplete Information, and Participation Costs: Existence and Convergence to Perfect Competition," *Econometrica*, 75, 155–200.
Satterthwaite, M. and S. R. Williams (1989): "Bilateral Trade with the Sealed $k$-Double Auction: Existence and Efficiency," *Journal of Economic Theory*, 48, 107–133.
——— (2002): "The Optimality of a Simple Market Mechanism," *Econometrica*, 70, 1841–1863.
Skrzypacz, A. and J. Toikka (2015): "Mechanisms for Repeated Trade," *The American Economic Journal: Microeconomics*, 7, 252–293.
Ünver, M. U. (2010): "Dynamic Kidney Exchange," *Review of Economic Studies*, 77, 372–414.
Vayanos, D. (1999): "Strategic Trading and Welfare in a Dynamic Market," *Review of Economic Studies*, 66, 219–254.
Vickrey, W. (1961): "Counterspeculation, Auction, and Competitive Sealed Tenders," *Journal of Finance*, 16, 8–37.
Williams, S. R. (1999): "A Characterization of Efficient, Bayesian Incentive Compartible Mechanisms," *Economic Theory*, 14, 155–180.

# Appendix

## A Proofs for Section 3

### A.1 Preliminaries: Markov decision process representation

In this appendix we map the dynamic optimization problem faced by the social planner to a Markov decision process $\langle \mathcal{X}, \mathcal{A}, P, r, \delta \rangle$. We also formally define the class of threshold policies with respect to this process.

We construct the Markov decision process by initially assuming that the social planner always stores and clears traders as part of a buyer-seller *pair*.[21] This assumption is without loss of generality because we are dealing with a bilateral trade setting in which a single buyer and seller arrive in each period. To identify the *state* in any given period we can think of the planner as first determining the number of efficient $(\overline{v}, \underline{c})$ pairs that are present before then determining the number of identical suboptimal $(\overline{v}, \overline{c})$ or $(\underline{v}, \underline{c})$ pairs among the remaining set of traders. This is without loss of generality because the non-identical suboptimal pairs $(\overline{v}, \overline{c})$ and $(\underline{v}, \underline{c})$ can be rematched to form an efficient $(\overline{v}, \underline{c})$ pair and an infeasible $(\underline{v}, \overline{c})$ pair. Infeasible pairs do not need to be counted because they do not generate positive surplus and cannot be rematched to create efficient pairs. We can therefore assume without loss of generality that such pairs are immediately cleared from the market without trading. By the symmetric nature of our setting we can characterize the optimal market clearing policy without distinguishing the type of suboptimal pairs that are present. The *state space* of the social planner's Markov decision process is therefore two-dimensional and given by $\mathcal{X} := \{(x_E, x_S) : x_E, x_S \in \mathbb{Z}_{\geq 0}\}$, where $x_E$ and $x_S$ respectively denote the number of efficient pairs and the number of suboptimal pairs present.

The set of *actions* available to the social planner in state $\boldsymbol{x}$ is given by $\mathcal{A}_{\boldsymbol{x}} := \{(a_E, a_S) : a_E, a_S \in \mathbb{Z}_{\geq 0}, a_E \leq x_E, a_S \leq x_S\}$, where $a_E$ and $a_S$ respectively denote the number of efficient pairs and the number of suboptimal pairs that are cleared from the market. Taking $\mathcal{A} := \cup_{\boldsymbol{x} \in \mathcal{X}} \mathcal{A}_{\boldsymbol{x}}$ then specifies the *action space* of the Markov decision process. The *reward function* $r : \mathcal{A} \to \mathbb{R}$ maps an action $\boldsymbol{a} \in \mathcal{A}$ to the immediately reward $r(\boldsymbol{a}) = a_E + \Delta a_S$ for the social planner.

Let $\boldsymbol{X}_t \in \mathcal{X}$ denote the state of the market after the arrival of the period $t$ agents and $\boldsymbol{A}_t$ denote the action taken by the social planner in period $t \in \mathbb{N}$. The *transition probability*

$$P(\boldsymbol{x'}, \boldsymbol{x}, \boldsymbol{a}) := \mathbb{P}(\boldsymbol{X}_{t+1} = \boldsymbol{x'} | \boldsymbol{X}_t = \boldsymbol{x}, \boldsymbol{A}_t = \boldsymbol{a})$$

[21] Alternatively, and in line with Section 3.1, we could assume that the social planner only stores efficient *traders*. As we will see shortly, assuming that the social planner always stores and clears pairs of traders is also without loss of generality and permits a simple representation of the Markov decision process.

then specifies the probability that the state in period $t+1$ is $\boldsymbol{x'}$, conditional on a period $t$ state of $\boldsymbol{x}$ and action of $\boldsymbol{a}$. Fixing a state $\boldsymbol{x} = (x_E, x_S)$ and a feasible action $\boldsymbol{a} = (a_E, a_S)$, an efficient pair arrives in the following period with probability $w^2$, while an infeasible one arrives with probability $(1-w)^2$. We therefore have

$$P(\boldsymbol{x} - \boldsymbol{a} + (1,0), \boldsymbol{x}, \boldsymbol{a}) = w^2 \quad \text{and} \quad P(\boldsymbol{x} - \boldsymbol{a}, \boldsymbol{x}, \boldsymbol{a}) = (1-w)^2.$$

A suboptimal pair arrives in the following period with probability $w(1-w)$. If $x_S - a_S = 0$ then we have

$$P(\boldsymbol{x} - \boldsymbol{a} + (0,1), \boldsymbol{x}, \boldsymbol{a}) = 2w(1-w)$$

and if $x_S - a_S > 0$, we have

$$P(\boldsymbol{x} - \boldsymbol{a} + (0,1), \boldsymbol{x}, \boldsymbol{a}) = P(\boldsymbol{x} - \boldsymbol{a} + (1,-1), \boldsymbol{x}, \boldsymbol{a}) = w(1-w).$$

Given the Markov decision process $\langle \mathcal{X}, \mathcal{A}, P, r, \delta \rangle$, a stationary and deterministic *policy* $\pi : \mathcal{X} \to \mathcal{A}$ is such that $\pi(\boldsymbol{x}) \in \mathcal{A}_{\boldsymbol{x}}$ specifies the action taken by the social planner in state $\boldsymbol{x}$. Since $r$ is a deterministic function, $\mathcal{X}$ is a countable set and, for all $\boldsymbol{x} \in \mathcal{X}$, $\mathcal{A}_{\boldsymbol{x}}$ is a finite set, this implies that a stationary deterministic optimal policy exists and is characterized by the appropriate Bellman equation (see, for example, Theorem 6.2.6 and Theorem 6.2.10 in Puterman, 1994). By construction, the *optimal policy* $\pi^*$ of the Markov decision process maximizes expected discounted social welfare. The dynamic optimization problem of the social planner therefore reduces to determining the optimal policy $\pi^*$ of $\langle \mathcal{X}, \mathcal{A}, P, r, \delta \rangle$.

In characterizing the optimal policy, we introduce a simple class of policies that we call *threshold policies*. Threshold policies immediately clear efficient (and infeasible) pairs from the market. Identical suboptimal pairs are stored up to a threshold of $\tau \in \mathbb{Z}_{\geq 0}$. Any additional identical suboptimal pairs are cleared immediately from the market. Therefore, given a threshold $\tau \in \mathbb{Z}_{\geq 0}$, the associated threshold policy $\pi_\tau$ of the Markov decision process $\langle \mathcal{X}, \mathcal{A}, P, r, \delta \rangle$ is given by

$$\pi_\tau(x_E, x_S) = (x_E, 0) \quad \text{if} \quad x_S \leq \tau \quad \text{and} \quad \pi_\tau(x_E, x_S) = (x_E, x_S - \tau) \quad \text{if} \quad x_S > \tau.$$

Representing the Markov decision process in terms of pairs of traders provides a convenient formalism. However, equivalently—and in line with Section 3.1—we can assume without loss of generality that the social planner *only stores (identical) efficient traders* that arrive as part of suboptimal pairs and immediately clears the suboptimal traders from the market.

## A.2 Proof of Theorem 1

The following proof uses the Markov decision process representation of the social planner's dynamic optimization problem that is introduced in Appendix A.1. This representation assumes, without loss of generality, that traders are always stored and cleared from the market in *pairs*. Appendix A.1 also provides a formal definition of threshold policies.

*Proof.* Appendix A.1 shows that there exists a deterministic, stationary optimal policy $\pi^*$ of the Markov decision process $\langle \mathcal{X}, \mathcal{A}, P, r, \delta \rangle$. Efficient pairs of traders yield the maximum possible gain from trade and there is no benefit associated with storing such pairs. Therefore, $\pi^*$ must immediately clear any efficient trades form the market and, for all $\boldsymbol{x} = (x_E, x_S) \in \mathcal{X}$, we have $\pi^*(x_E, x_S) \geq (x_E, 0)$. Moreover, sample paths of the Markov decision process under the optimal policy $\pi^*$ are such that if $x_S$ identical suboptimal pairs are present in period $t$ then $x_S - 1$ suboptimal pairs must have been stored with positive probability under $\pi^*$ in some previous period. Since $\pi^*$ is stationary and deterministic, this implies that if it is optimal to store $x_S$ suoptimal pairs under $\pi^*$ then it is necessarily optimal to store $x_S - 1$ suboptimal pairs under $\pi^*$. Summarizing, either the optimal market clearing policy $\pi^*$ is such that

$$\pi^*(x_E, x_S) = (x_E, 0) \quad \text{for all } \boldsymbol{x} \in \mathcal{X},$$

or there exists a threshold $\tau^* \in \mathbb{Z}_{\geq 0}$ such that

$$\pi^*(x_E, x_S) = (x_E, 0) \quad \text{if} \quad x_S \leq \tau^* \quad \text{and} \quad \pi^*(x_E, x_S) = (x_E, x_S - \tau^*) \quad \text{if} \quad x_S > \tau^*.$$

It only remains to rule out the first of these cases and show that an unbounded number of suboptmal pairs cannot be stored under $\pi^*$. Observe that as the number of stored suboptimal pairs diverges to infinity, the expected number of periods until all stored suboptimal pairs are rematched also diverges to infinity. Therefore, as the number of stored pairs increases, the expected discounted benefit from storing an additional suboptimal pair converges to zero. However, the benefit of immediately clearing a suboptimal pair is always $\Delta > 0$. Putting all of this together, there exists a maximum number $\tau^*$ of suboptimal trades which can be optimally stored and the optimal policy $\pi^*$ is a threshold policy.[22]

We conclude by computing the stationary distribution of the Markov chain $\{Y_t\}_{t \in \mathbb{N}}$ over the number of stored suboptimal pairs that is induced by a given threshold policy $\pi_\tau$. The

[22]See Proposition 4 for an explicit upper bound on $\tau^*$.

transition matrix $\boldsymbol{P}$ of this Markov chain is given by

$$\boldsymbol{P} = \begin{pmatrix} 1-2\lambda & 2\lambda & 0 & \cdots & 0 & 0 & 0 \\ \lambda & 1-2\lambda & \lambda & & 0 & 0 & 0 \\ 0 & \lambda & 1-2\lambda & & 0 & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & & 1-2\lambda & \lambda & 0 \\ 0 & 0 & 0 & & \lambda & 1-2\lambda & \lambda \\ 0 & 0 & 0 & \dots & 0 & \lambda & 1-\lambda \end{pmatrix}.$$

Writing the stationary distribution as a row vector $\boldsymbol{\kappa}$, $\boldsymbol{\kappa}$ satisfies $\sum_{i=0}^{\tau} \kappa_i = 1$ and solves the linear system $\boldsymbol{\kappa P} = \boldsymbol{\kappa}$. Solving this system is straightforward and yields $\kappa_i = \frac{2}{2\tau+1}$ for $i > 0$ and $\kappa_0 = \frac{1}{2\tau+1}$, as required. □

## A.3 Proof of Proposition 2

*Proof.* In the static bilateral trade setting, expected per period surplus is

$$\begin{aligned} S_0 &= \int_0^1 \int_0^v (v-c)\, dg(c) dF(v) \\ &= (1-F(p))G(p)(\overline{v} - \underline{c}) + G(p)F(p)\Delta_L + (1-F(p))(1-G(p))\Delta_H, \end{aligned}$$

where $\Delta_L := \frac{\int_0^p \int_0^v (v-c) dG(c) dF(v)}{F(p)G(p)}$, $\Delta_H := \frac{\int_p^1 \int_p^v (v-c) dG(c) dF(v)}{(1-F(p))(1-G(p))}$ and $(\overline{v} - \underline{c}) := \frac{\int_p^1 \int_0^p (v-c) dG(c) dF(v)}{(1-F(p))G(p)}$. Recalling that $G(p) = w = 1 - F(p)$, this is the same as $S_0 = w^2(\overline{v} - \underline{c}) + 2w(1-w)\tilde{\Delta}$, where $\tilde{\Delta} = \frac{\Delta_H + \Delta_L}{2}$ is the expected surplus of a suboptimal trade.

Consider now the threshold policy that separates buyers and sellers into two classes: efficient traders, meaning $v \geq p$ and $c \leq p$, and suboptimal traders, meaning $v < p$ and $c > p$. The market clearing policy immediately executes trades between efficient trader pairs and stores identical efficient traders—buyers with $v \geq p$ or sellers with $c \leq p$—up to some threshold $\tau$. Note that this market clearing policy does not distinguish the nature of these traders (i.e. whether they arrived as part of a suboptimal pair that creates surplus $\Delta_H$ or $\Delta_L$). However, since each type of suboptimal pair arrives with an equal probability, a suboptimal trade creates an expected surplus of $\tilde{\Delta}$. Setting $\Delta = \frac{\tilde{\Delta}}{\overline{v} - \underline{c}}$, our analysis from the binary type setting applies and we have, in particular, that the relative increase in market thickness from storing $\tau$ efficient traders is $\mathcal{T}_\tau$. □

## A.4 Proof of Proposition 3

Before proceeding with the proof of Proposition 3, we prove the following useful lemma.

**Lemma A1.** *For any* $\tau \geq 1$, $\tau^* \geq \tau$ *if and only if* $\left(\frac{z_+ z_-}{2}\right)^\tau \geq \Delta\left(z_+^\tau + z_-^\tau\right)$, *where*

$$z_+ = 2 + \frac{\sqrt{(1-\delta)(1-\delta+4\delta w(1-w))} + (1-\delta)}{\delta w(1-w)}, \tag{5}$$

$$z_- = 2 - \frac{\sqrt{(1-\delta)(1-\delta+4\delta w(1-w))} - (1-\delta)}{\delta w(1-w)}. \tag{6}$$

*For the special case of* $\tau = 1$ *we have* $\tau^* \geq 1$ *if and only if* $\delta \geq \delta^*$, *where* $\delta^*$ *is defined in* (3).

*Proof.* We prove this lemma by deriving and solving the Bellman equations that characterize the optimal threshold policy. For a given threshold $\tau \in \mathbb{Z}_{\geq 0}$ and $y \in \{0, \ldots, \tau\}$, let $V_\tau(y)$ denote the expected discounted value associated with having $y$ efficient traders stored at the *end* of a given period (i.e. after executing the action and earning the corresponding instantaneous reward for that period) under the threshold policy with threshold $\tau$. Let $\lambda = w(1-w)$. Then, for $\tau = 0$, we have

$$V_0(0) = \frac{\delta(w^2 + 2\lambda\Delta)}{(1-\delta)}. \tag{7}$$

For $\tau \geq 1$, the set of value functions $\{V_\tau(0), \ldots, V_\tau(\tau)\}$ are given by the solution to a system of $\tau + 1$ linear equations. For $y \in \{1, \ldots, \tau - 1\}$, $V_\tau(y)$ satisfies the recursion

$$V_\tau(y) = \delta\left[w^2(1 + V_\tau(y)) + \lambda(1 + V_\tau(y-1) + V_\tau(y+1)) + (1-w)^2 V_\tau(y)\right] \tag{8}$$

with boundary conditions

$$V_\tau(0) = \delta\left[w^2(1 + V_\tau(0)) + 2\lambda V_\tau(1) + (1-w)^2 V_\tau(0)\right], \tag{9}$$

$$V_\tau(\tau) = \delta\left[w^2(1 + V_\tau(\tau)) + \lambda(1 + V_\tau(\tau-1) + \Delta + V_\tau(\tau)) + (1-w)^2 V_\tau(\tau)\right]. \tag{10}$$

Recall that throughout this paper we take $\tau^*$ to be the largest optimal storage threshold when multiple storage thresholds are optimal. Since $\Delta$ is the instantaneous reward from clearing a suboptimal trade, the optimal threshold $\tau^*$ is then characterized by the stopping conditions

$$V_{\tau^*+1}(\tau^* + 1) - V_{\tau^*+1}(\tau^*) < \Delta \quad \text{and} \quad V_{\tau^*}(\tau^*) - V_{\tau^*}(\tau^* - 1) \geq \Delta. \tag{11}$$

Note that for the special case of $\tau^* = 0$, only the first stopping condition $V_1(1) - V_1(0) < \Delta$ applies. For any $\tau \geq 1$ we have $\tau^* \geq \tau$ if and only if the second stopping condition $V_\tau(\tau) - V_\tau(\tau - 1) \geq \Delta$ holds.

Given any $\tau \geq 1$, we now characterize when $\tau^* \geq \tau$. For any $y \in \{2, \ldots, \tau\}$, we set $\widetilde{V}_\tau(y) := V_\tau(y) - V_\tau(y-1)$. The stopping conditions given in (11) that characterize $\tau^*$ then become

$$\widetilde{V}_{\tau^*+1}(\tau^*+1) < \Delta \quad \text{and} \quad \widetilde{V}_{\tau^*}(\tau^*) \geq \Delta. \tag{12}$$

Moreover, we have $\tau^* \geq \tau$ if and only if $\widetilde{V}_\tau(\tau) \geq \Delta$. For $\tau = 1$, $V_1(0)$ and $V_1(1)$ satisfy (9) and (10). We have

$$\begin{aligned} V_1(0) &= \delta\left[w^2(1+V_1(0)) + 2\lambda V_1(1) + (1-w)^2 V_1(0)\right], \\ V_1(1) &= \delta\left[w^2(1+V_1(1)) + \lambda(1+V_1(0)+\Delta+V_1(1)) + (1-w)^2 V_1(1)\right], \end{aligned}$$

and solving these linear equations yields

$$\begin{aligned} V_1(0) &= \frac{\delta\left(2\delta\Delta\lambda^2 + w^2(\delta(\lambda-1)+1) + 2\delta\lambda w\right)}{(1-\delta)(\delta(3\lambda-1)+1)}, \\ V_1(1) &= \frac{\delta\left(\Delta\lambda(\delta(2\lambda-1)+1) + \delta\lambda w^2 + w(\delta(2\lambda-1)+1)\right)}{(1-\delta)(\delta(3\lambda-1)+1)}, \end{aligned}$$

which implies that

$$\widetilde{V}_1(1) = V_1(1) - V_1(0) = \frac{\lambda\delta(1+\Delta)}{1-\delta+3\lambda\delta}. \tag{13}$$

More generally, for $\tau \geq 2$ and $y \in \{2, \ldots, \tau-1\}$, the "difference" value function $\widetilde{V}_\tau(y)$ satisfies the recursion

$$\widetilde{V}_\tau(y) = \delta\left[(1-2\lambda)\widetilde{V}_\tau(y) + \lambda\widetilde{V}_\tau(y-1) + \lambda\widetilde{V}_\tau(y+1))\right], \tag{14}$$

with boundary conditions

$$\widetilde{V}_\tau(1) = \delta\left[\lambda + (1-3\lambda)\widetilde{V}_\tau(1) + \lambda\widetilde{V}_\tau(2)\right], \tag{15}$$

$$\widetilde{V}_\tau(\tau) = \delta\left[\lambda\Delta + (1-2\lambda)\widetilde{V}_\tau(\tau) + \lambda\widetilde{V}_\tau(\tau-1)\right]. \tag{16}$$

Solving the recursion (14) with the boundary condition (15) in Mathematica yields

$$\widetilde{V}_\tau(y) = \left(\frac{z_-}{2}\right)^\tau k_0 + \left(\frac{z_+}{2}\right)^\tau \left(\frac{2-(z_-+2)k_0}{z_++2}\right),$$

where $z_+$ and $z_-$ are defined in (5) and (6) and the constant $k_0$ is pinned down by the boundary condition (16). Elementary calculations (which we omit here for the sake of brevity) show that $z_+ > 2$ and $z_- \in (0,2)$. Imposing the boundary condition (16) yields

$$k_0 = \frac{2z_+^{\tau-1}(2\delta\lambda z_+ - \delta z_+ + z_+ - 2\delta\lambda) - (z_+ + 2)\delta\Delta\lambda 2^m}{(z_- + 2)z_+^{\tau-1}(2\delta\lambda z_+ - \delta z_+ + z_+ - 2\delta\lambda) - (z_+ + 2)z_-^{\tau-1}(2\delta\lambda z_- + z_- - \delta z_- - 2\delta\lambda)}.$$

and putting everything together we have

$$\widetilde{V}_\tau(\tau) = \frac{(z_+ - 2)\left(\frac{z_+ z_-}{2}\right)^\tau + \Delta z_+\left(2z_+^{\tau-1} - z_-^\tau\right)}{z_+^{\tau+1} - 2z_-^\tau}.$$

Plugging $\tau = 1$ into this last expression yields (13), which shows that the solution is in fact valid for any $\tau \geq 1$. For any $\tau \geq 1$, we therefore have

$$\widetilde{V}_\tau(\tau) - \Delta = \frac{(z_+ - 2)\left[\left(\frac{z_+ z_-}{2}\right)^\tau - \Delta\left(z_+^\tau + z_-^\tau\right)\right]}{z_+^{\tau+1} - 2z_-^\tau}. \tag{17}$$

Combining this with fact that $z_+ > 2$ and $z_- \in (0,2)$ then shows that $\widetilde{V}_\tau(\tau) \geq \Delta$ and $\tau^* \geq \tau$ if and only if

$$\left(\frac{z_+ z_-}{2}\right)^\tau - \Delta\left(z_+^\tau + z_-^\tau\right) \geq 0$$

as required. For the special case of $\tau = 1$ this condition simplifies to $\delta \geq \delta^*$, where $\delta^*$ is defined in (3). □

We are now ready to proceed with the proof of Proposition 3.

*Proof.* Since $\mathcal{T}^* = \frac{2\tau^*}{2\tau^*+1}$ is increasing in $\tau^*$, it suffices to prove each of the comparative statics results for $\tau^*$.

First, we show that $\tau^*$ is increasing in $\delta$ on $[0,1)$. Given any $\tau \geq 1$, each term in the recursion that characterizes the "difference" value function $\widetilde{V}_\tau(\tau)$ is increasing in $\delta$. Hence, for every $\tau \geq 1$, $\widetilde{V}_\tau(\tau)$ is increasing in $\delta$ and the stopping conditions given in (12) then immediately imply that $\tau^*$ is increasing in $\delta$ on $[0,1)$ as required. Second, we show that $\tau^*$ is decreasing in $\Delta$ on $(0, \frac{1}{2})$ and increasing in $\lambda = w(1-w)$ on $(0, \frac{1}{4})$. Using (17) we have, for any $\tau \geq 1$,

$$\frac{d\,(\widetilde{V}_\tau(\tau) - \Delta)}{d\,\Delta} = -\frac{(z_+^\tau - 2)(z_+^\tau + z_-^\tau)}{z_+^{\tau+1} - 2z_-^\tau} < 0,$$

$$\frac{d\,\widetilde{V}_\tau(\tau)}{d\,\lambda} = \frac{1-\delta}{\lambda\sqrt{(1-\delta)(1-\delta+4\delta\lambda)}} \left( \frac{\left(2\tau z_-^\tau + (\tau+1)z_+^{\tau+1}\right)\left(\Delta\left(2z_+^\tau - z_+ z_-^\tau\right) + 2^\tau(z_+ - 2)\right)}{\left(z_+^{\tau+1} - 2z_-^\tau\right)^2} + \frac{4\Delta(\tau-1)z_-^\tau + 8\Delta\tau z_+^{\tau-1} + 2^{\tau+2}}{z_-\left(z_+^{\tau+1} - 2z_-^\tau\right)} \right) > 0,$$

where the inequalities follows from $z_+ > 2$ and $z_- \in (0, 2)$. The stopping conditions given in (12) then immediately imply that $\tau^*$ is decreasing in $\Delta$ on $(0, \frac{1}{2})$ and increasing in $\lambda$ on $(0, \frac{1}{4})$ as required. □

## A.5 Proof of Proposition 4

*Proof.* Consider the decision to store the $j$th efficient trader. The instantaneous reward from clearing this trader immediately is $\Delta$, while an upper bound on the payoff associated with storing this trader is $\delta^j(1-\Delta)$.[23] For storing the $j$th trader to be profitable, we must have $\delta^j(1-\Delta) > \Delta$. Hence, $\tau^*$ must be such that

$$\tau^* < \frac{\log\left(\frac{\Delta}{1-\Delta}\right)}{\log(\delta)}.$$

In the limit as $\delta \to 1$ (i.e. taking a Laurent series expansion of $\frac{1}{\log(\delta)}$ about $\delta = 1$), we have an upper bound on $\tau^*$ of

$$\tau^* \leq \frac{\log\left(\frac{1-\Delta}{\Delta}\right)}{1-\delta} + \frac{\log\left(\frac{\Delta}{1-\Delta}\right)}{2} + O(1-\delta).$$

□

[23] Note that this is an upper bound because the value associated with storing the $j$th trader cannot be realized until $j$ appropriate suboptimal pairs have arrived, which cannot occur for at least $j$ subsequent periods. When stored efficient traders are rematched and cleared, this produces an instantaneous reward of 1 for the social planner. However, if the $j$th efficient trader was not stored then the efficient trader from the $j$th appropriate suboptimal pair would instead be stored and subsequently generate an expected discounted payoff of at least $\Delta$ (otherwise it would not be optimal to store even one efficient trader, let alone $j$ of them).

# B Proofs for Section 4

## B.1 Proof of Proposition 5

*Proof.* We consider the posted-price mechanism characterized by the pricing rule $p$ introduced in the proposition statement, the optimal threshold policy with threshold $\tau^* \geq 1$ and a last-in-first-out queueing protocol. By construction, the posted-price mechanism implements the optimal market clearing policy under truthful reporting. Suppose that agents report truthfully under this mechanism. In the states $z \in \{-\tau^* + 1, \ldots, \tau^* - 1\}$, the mechanism then posts a price of $p(z) \in [\underline{v}, \overline{c}]$ and only executes efficient trades at the price. By construction, only $\underline{c}$ and $\overline{v}$ types earn a positive payoff from trading at this price. In the state $z = \tau^*$, the mechanism posts a price of $p(\tau^*) \in [\overline{c}, \overline{v}]$. Any efficient trade that arises is executed and if a buyer of type $\overline{v}$ arrives with a seller of type $\overline{c}$, then a suboptimal $(\overline{v}, \overline{c})$ trade is also executed. By construction, only $\underline{c}$, $\overline{c}$ and $\overline{v}$ types earn a non-negative payoff from trading at this price. In the state $z = -\tau^*$, the mechanism posts a price of $p(-\tau^*) \in [\underline{c}, \underline{v}]$. Any efficient trade that arises is executed and if a buyer of type $\underline{v}$ arrives with a seller of type $\underline{c}$, then a suboptimal $(\underline{v}, \underline{c})$ trade is also executed. By construction, only $\underline{c}$, $\underline{v}$ and $\overline{v}$ types earn a non-negative payoff from trading at this price.

It remains to check that truthful reporting is a weakly dominant strategy for all traders. Suppose that an arriving buyer faces a price of $p \in [\underline{c}, \underline{v}]$. Then the arriving buyer will trade at this price and earn a non-negative payoff regardless of their report. Consequently, truthful reporting is a weakly dominant strategy for both types. Next, suppose that an arriving buyer faces a price of $p \in [\underline{v}, \overline{c}]$. Any buyer that reports type $\underline{v}$ then guarantees themselves a payoff of zero. The last-in-first-out queueing protocol ensures that any buyer that reports to be of type $\overline{v}$ will eventually trade at this price. Truthful reporting is then a weakly dominant strategy of type $\underline{v}$ since this guarantees that type a payoff of zero rather than a non-positive expected discounted payoff. Similarly, truthful reporting is a weakly dominant strategy of type $\overline{v}$ since this guarantees that type a strictly positive expected discounted payoff, rather than a payoff of zero. Finally, suppose that an arriving buyer faces a price of $p \in [\overline{c}, \overline{v}]$. The last-in-first-out queuing protocol ensures that the arriving buyer immediately trades at this price upon reporting to be of type $\overline{v}$ and is immediately cleared from the market without trading upon reporting to be of type $\underline{v}$. Truthful reporting is then a dominant strategy for buyers of type $\underline{v}$ since this guarantees a payoff of zero, rather than a negative payoff. Similarly, truthful reporting is a weakly dominant strategy for buyers of type $\overline{v}$ since this guarantees a non-negative payoff, rather than a payoff of zero. Putting all of this together, we see that truthful reporting is a weakly dominant strategy for all buyers. An analogous argument shows that truthful reporting is also a weakly dominant strategy for sellers. □

## B.2 Proof of Corollary 1

*Proof.* By construction, the posted price mechanism from Proposition 5 does not run a deficit. Combining this with Lemma A1, which shows that $\tau^* \geq 1$ if and only if $\delta \geq \delta^*$ then yields the desired result. $\square$

## B.3 Preliminaries: Dynamic mechanism design analysis

In this appendix we provide the formal mechanism design details that underpin the proof of Theorem 2. Without loss of generality we restrict attention to direct, deterministic mechanisms. A direct, deterministic mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ consists of an allocation rule $\boldsymbol{Q} = \{\boldsymbol{Q}_t\}_{t \in \mathbb{N}}$ and a payment rule $\boldsymbol{M} = \{\boldsymbol{M}_t\}_{t \in \mathbb{N}}$. Let $\mathcal{H}_t \coloneqq (\{\underline{v}, \overline{v}\} \times \{\underline{c}, \overline{c}\})^t$ denote the set of histories of traders' reports up to and including period $t$. The period $t$ allocation rule $\boldsymbol{Q}_t : \mathcal{H}_t \to \{0,1\}^{2t}$ maps the period $t$ history of trader reports $\boldsymbol{h}_t$ to the set of period $t$ allocations, and similarly, the period $t$ transfer rule $\boldsymbol{M}_t : \mathcal{H}_t \to \mathbb{R}^{2t}$ maps this history to the set of period $t$ transfers.

In contrast to the Markov decision process constructed in Appendix A.1, which does not differentiate between different traders of the same type, we now need to keep track of traders' identities. We label buyers and sellers according to the period in which they arrive. With this in mind, given a period $t$ history $\boldsymbol{h}_t \in \mathcal{H}_t$, we denote the respective period $t$ allocations of buyer and seller $i \in \{1, \ldots, t\}$ by $Q_t^{B_i}(\boldsymbol{h}_t)$ and $Q_t^{S_i}(\boldsymbol{h}_t)$. Similarly, $M_t^{B_i}(\boldsymbol{h}_t)$ and $M_t^{S_i}(\boldsymbol{h}_t)$ denote the respective expected transfers from $B_i$ and to $S_i$ in period $t$ given $\boldsymbol{h}_t$. Due to the anonymity of the Markov decision process, mapping a market clearing policy $\pi$ to an allocation rule $\boldsymbol{Q}$ requires augmenting $\pi$ with a queueing protocol, denoted by $\mu$, that serves as a tie-breaking rule among traders of the same type.

Although the mechanism design problem is dynamic because the optimal policy varies with the state, the problem of incentivizing the traders to reveal their private information is essentially static as the traders' private information does not evolve over time. The *periodic ex post incentive compatibility (P-IC)* and *periodic ex post individual rationality (P-IR)* constraints require that truthful reporting and participation is optimal for every period $t$ trader and every history $\boldsymbol{h}_{t-1}$, regardless of the report of the other period $t$ trader, assuming that all future traders report truthfully.[24] Formally, given a direct mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ let $q(\hat{\theta}, \theta, \boldsymbol{h}_{t-1})$ and $m(\hat{\theta}, \theta, \boldsymbol{h}_{t-1})$ denote the discounted probability of trade and expected discounted payment, respectively, for a trader that reports $\hat{\theta}$ at history $\boldsymbol{h}_{t-1}$ when the other

[24] In a companion paper (Loertscher and Muir, 2021b), we also consider weaker constraints. *Interim* constraints require that truthful reporting and participation is optimal for every period $t$ trader and every history $\boldsymbol{h}_{t-1} \in \mathcal{H}_{t-1}$, assuming the other period $t$ trader and all future traders report truthfully. *Bayesian* constraints require that truthful reporting and participation is optimal for every period $t$ trader, assuming all other traders report truthfully.

period $t$ trader reports $\theta$.[25] For every history $\boldsymbol{h}_{t-1} \in \mathcal{H}_{t-1}$, $v \in \{\underline{v}, \overline{v}\}$ and $c \in \{\underline{c}, \overline{c}\}$, P-IC requires

$$\begin{aligned} v &= \underset{\hat{\theta} \in \{\underline{v}, \overline{v}\}}{\arg\max} \left\{ vq(\hat{\theta}, c, \boldsymbol{h}_{t-1}) - m(\hat{\theta}, c, \boldsymbol{h}_{t-1}) \right\}, \\ c &= \underset{\hat{\theta} \in \{\underline{c}, \overline{c}\}}{\arg\max} \left\{ m(\hat{\theta}, v, \boldsymbol{h}_{t-1}) - cq(\hat{\theta}, v, \boldsymbol{h}_{t-1}) \right\}, \end{aligned} \tag{P-IC}$$

while P-IR requires

$$vq(v, c, \boldsymbol{h}_{t-1}) - m(v, c, \boldsymbol{h}_{t-1}) \geq 0 \quad \text{and} \quad m(c, v, \boldsymbol{h}_{t-1}) - cq(c, v, \boldsymbol{h}_{t-1}) \geq 0. \tag{P-IR}$$

Under the mechanism that maximizes the market maker's expected discounted profit, the individual rationality constraints bind for buyers of type $\underline{v}$ and sellers of type $\overline{c}$ and the incentive compatibility constraints bind locally downward for buyers of type $\overline{v}$ and locally upward for sellers of type $\underline{c}$ (see, for example, Elkind, 2007). These binding constraints yield

$$\begin{aligned} m(\underline{v}, c, \boldsymbol{h}_{t-1}) &= \underline{v}q(\underline{v}, c, \boldsymbol{h}_{t-1}), \quad m(\overline{c}, v, \boldsymbol{h}_{t-1}) = \overline{c}q(\overline{c}, v, \boldsymbol{h}_{t-1}) \\ m(\overline{v}, c, \boldsymbol{h}_{t-1}) &= \overline{v}(q(\overline{v}, c, \boldsymbol{h}_{t-1}) - q(\underline{v}, c, \boldsymbol{h}_{t-1})) + \underline{v}q(\underline{v}, c, \boldsymbol{h}_{t-1}), \\ m(\underline{c}, c, \boldsymbol{h}_{t-1}) &= \underline{c}(q(\underline{c}, v, \boldsymbol{h}_{t-1}) - q(\overline{c}, v, \boldsymbol{h}_{t-1})) + \overline{c}q(\overline{c}, v, \boldsymbol{h}_{t-1}). \end{aligned} \tag{18}$$

The incentive compatibility constraints for buyers of type $\underline{v}$ and seller of type $\overline{c}$ are satisfied and the allocation rule is *implementable* as part of a P-IC and P-IR mechanism if and only if the allocation rule is *monotone* in the sense that $q(\overline{v}, c, \boldsymbol{h}_{t-1}) \geq q(\underline{v}, c, \boldsymbol{h}_{t-1})$ and $q(\underline{c}, v, \boldsymbol{h}_{t-1}) \geq q(\overline{c}, v, \boldsymbol{h}_{t-1})$.[26] The virtual type functions are then given by

$$\Phi(\underline{v}) \coloneqq \underline{v} - \frac{w}{1-w}(\overline{v} - \underline{v}) \quad \text{and} \quad \Gamma(\overline{c}) \coloneqq \overline{c} + \frac{w}{1-w}(\overline{c} - \underline{c}),$$

with $\Phi(\overline{v}) = \overline{v}$ and $\Gamma(\underline{c}) = \underline{c}$.[27]

Next, for $i \leq t$, let $v^{B_i}(\boldsymbol{h}_t) \in \{\underline{v}, \overline{v}\}$ and $c^{S_i}(\boldsymbol{h}_t) \in \{\underline{c}, \overline{c}\}$ denote the types of buyer $B_i$ and seller $S_i$ given history $\boldsymbol{h}_t \in \mathcal{H}_t$, respectively. Following standard mechanism design arguments (see, for example, Myerson, 1981; Börgers, 2015; Loertscher and Muir, 2021b) expected discounted profit under any direct mechanism that implements the allocation rule

[25] Since we are dealing with deterministic mechanisms, for the period $t$ buyer we have $q(\hat{v}, c, \boldsymbol{h}_{t-1}) = \sum_{i=t}^{\infty} \delta^{i-1} Q_i^{B_t}(\boldsymbol{h}_i) \mathbb{P}(\boldsymbol{H}_i = \boldsymbol{h}_i | \boldsymbol{H}_t = (\hat{v}, c, \boldsymbol{h}_{t-1}))$ and $m(\hat{v}, c, \boldsymbol{h}_{t-1}) = \sum_{i=t}^{\infty} \delta^{i-1} M_i^{B_t}(\boldsymbol{h}_i) \mathbb{P}(\boldsymbol{H}_i = \boldsymbol{h}_i | \boldsymbol{H}_t = (\hat{v}, c, \boldsymbol{h}_{t-1}))$. We can analogously define $q(\hat{c}, v, \boldsymbol{h}_{t-1})$ and $m(\hat{c}, v, \boldsymbol{h}_{t-1})$ for the period $t$ seller.

[26] Without loss of generality we can then set, for all $t \in \mathbb{N}$ and $\boldsymbol{h}_t \in \mathcal{H}_t$, $M_t^{B_t}(\boldsymbol{h}_t) = m(\hat{v}, c, \boldsymbol{h}_{t-1})$ and, for all $i \neq t$, $M_t^{B_i}(\boldsymbol{h}_t) = 0$. Analogous expressions hold for the sellers.

[27] With binary types, $\Phi(\overline{v}) > \Phi(\underline{v})$ and $\Gamma(\underline{c}) < \Gamma(\overline{c})$, and so the regularity condition of Myerson (1981) is always satisfied.

$\boldsymbol{Q}$ under the binding P-IC and P-IR constraints described above[28] is given by

$$R(\boldsymbol{Q}) := \sum_{t=1}^{\infty}\sum_{i=1}^{t}\sum_{\boldsymbol{h}_t\in\mathcal{H}_t} \delta^{t-1}\left(\Phi(v^{B_i}(\boldsymbol{h}_t))Q_t^{B_i}(\boldsymbol{h}_t) - \Gamma(c^{S_i}(\boldsymbol{h}_t))Q_t^{S_i}(\boldsymbol{h}_t)\right)\mathbb{P}(\boldsymbol{H}_t = \boldsymbol{h}_t). \qquad (19)$$

The problem of a profit-maximizing market maker therefore reduces to determining the implementable allocation rule $\boldsymbol{Q}^*_{PM}$ that maximizes (19) respectively subject to the feasibility constraints that are inherent in our dynamic setting.[29]

Expected discounted welfare under any direct, truthful mechanism that implements the allocation rule $\boldsymbol{Q}$ is given by

$$W(\boldsymbol{Q}) := \sum_{t=1}^{\infty}\sum_{i=1}^{t}\sum_{\boldsymbol{h}_t\in\mathcal{H}_t} \delta^{t-1}\left(v^{B_i}(\boldsymbol{h}_t)Q_t^{B_i}(\boldsymbol{h}_t) - c^{S_i}(\boldsymbol{h}_t)Q_t^{S_i}(\boldsymbol{h}_t)\right)\mathbb{P}(\boldsymbol{H}_t = \boldsymbol{h}_t). \qquad (20)$$

Comparing (20) and (19) shows that welfare-maximization and profit-maximization are isomorphic insofar as the latter objective function has the same structure as the former, with the virtual types replacing the true types. Let $\Delta_{PM} := \Phi(\overline{v}) - \Gamma(\overline{c}) = \Phi(\underline{v}) - \Gamma(\underline{c})$ denote the value of a suboptimal trade to a profit-maximizing market maker. By construction, any allocation rule that implements the optimal threshold policy $\pi_{\tau^*}$, as derived in Section 3, maximizes (20) subject to feasibility. Therefore, applying our analysis from Section 3 with $\Delta$ replaced by $\Delta_{PM}$ shows that threshold policies are still optimal under profit-maximization. We let $\tau^*_{PM}$ denote the corresponding optimal threshold. Note that whenever $w \geq \Delta$, we have $\Delta_{PM} \leq 0$ and in this case the profit-maximizing market clearing policy $\pi_{\tau^*_{PM}}$ is such that $\tau^*_{PM} = \infty$.[30]

The allocation rule that implements the profit-maximizing threshold policy $\pi^*_{\tau^*_{PM}}$ is unique only up to the identities of traders that are cleared from the market when more than one trader of a given type is present in a single period. In expressing the market maker's optimization problem in terms of the Markov decision process representation constructed in Appendix A.1, we have essentially shown that the market maker's payoff does not vary with the treatment of individual traders for a given market clearing policy. Therefore, we can

[28]In a companion paper (Loertscher and Muir, 2021b), we show that the expected discounted profit of the market maker is invariant under the weaker constraints outlined in footnote 24. Thus, the market maker has no incentive to conceal the history $\boldsymbol{h}_{t-1} \in \mathcal{H}_t$ from arriving traders.

[29]*Feasibility* requires that, for all $t \in \mathbb{N}$ and all $\boldsymbol{h}_t \in \mathcal{H}_t$, $\sum_{i=1}^{t} Q_t^{B_i}(\boldsymbol{h}_t) \leq \sum_{i=1}^{t} Q_t^{S_i}(\boldsymbol{h}_t)$ and, for all $t \in \mathbb{N}$ and sequence of nested histories $\{\boldsymbol{h}_i\}_{i=t}^{\infty}$, $\sum_{i=t}^{\infty} Q_i^{B_t}(\boldsymbol{h}_i) \leq 1$ and $\sum_{i=t}^{\infty} Q_i^{S_j}(\boldsymbol{h}_i) \leq 1$. The first set of feasibility constraints ensure that aggregate demand never exceeds aggregate supply, while the second set of constraints are the dynamic analog of the feasibility constraints from a standard assignment game.

[30]Intuitively, $\Delta_{PM} \leq 0$ implies the profit associated with executing a suboptimal trade is non-positive. Thus, the market maker only executes efficient trades and is willing to store an unbounded number of efficient traders.

map the profit-maximizing market clearing policy $\pi_{\tau^*_{PM}}$ to a deterministic profit-maximizing allocation rule $\boldsymbol{Q}^*_{PM}$ by coupling it with any stationary, deterministic queueing protocol $\mu$ that specifies how ties are broken.

## B.4 Proof of Theorem 2

The proof of Theorem 2 relies on the dynamic mechanism design analysis from Appendix B.3. We start by formally stating and proving the following lemma.

**Lemma B2.** *Suppose that the market maker implements the profit-maximizing market clearing policy $\pi_{\tau^*_{PM}}$. Then its revenue does not vary with the choice of queueing protocol $\mu$.*

*Proof.* Given any market clearing policy $\pi$, we can compute expected discounted social surplus under $\pi$ by using the reward function $r(x_E, x_S) = x_E + \Delta x_S$, as specified in Appendix A.1. Similarly, by construction, we can compute the maximal expected discounted profit of the market maker when $\pi$ is implemented using a P-IC and P-IR mechanism by replacing the original reward function $r$ with the reward function $r_{PM}(x_E, x_S) = x_E + \Delta_{PM} x_S$. This immediately shows that the expected discounted profit of the market maker is invariant to the queueing protocol $\mu$. □

We are now in a position to prove Theorem 2.

*Proof.* Putting everything in Appendix B.3 together, a profit-maximizing dynamic mechanism can be constructed as follows. First, by Lemma B2, we can select any queueing protocol $\mu$. Together with the optimal policy $\pi_{\tau^*_{PM}}$ this gives rise to a profit-maximizing allocation rule $\boldsymbol{Q}^*_{PM}$. Second, given $\boldsymbol{Q}^*_{PM}$, one can compute the associated expected discounted allocation rule $\boldsymbol{q}^*_{PM}$. Provided the expected discounted allocation rule is monotone, the transfers under the optimal P-IC and P-IR mechanism can then be computed using (18) and footnote 26.

To complete the proof of the first statement of the theorem, it only remains to show that the optimal policy $\pi_{\tau^*_{PM}}$ together with *any* queueing protocol $\mu$ gives rise to a monotone allocation rule $\boldsymbol{Q}^*_{PM}$. For ease of exposition, from this point forward we simply write the allocation rule as $\boldsymbol{Q}$ and the expected discounted allocation rule as $\boldsymbol{q}$. Recall from Appendix B.3 that $q(\hat{\theta}, \theta, \boldsymbol{h}_{t-1})$ denotes the discounted probability of trade for a trader that reports $\hat{\theta}$ at history $\boldsymbol{h}_{t-1}$ when the other period $t$ trader reports $\theta$. However, given the policy $\pi_{\tau^*_{PM}}$ and the queueing protocol $\mu$, we can simplify the state space. In particular, rather than using the complete period $t-1$ history $\boldsymbol{h}_{t-1}$, we can simply consider states $z \in \{-\tau^*_{PM}, \ldots, \tau^*_{PM}\}$ that identify the number of efficient traders stored and whether these traders are buyers or sellers. Monotonicity then requires that, for any state $z \in \{-\tau^*_{PM}, \ldots, \tau^*_{PM}\}$, $v \in \{\underline{v}, \overline{v}\}$ and $c \in \{\underline{c}, \overline{c}\}$,

$$q(\overline{v}, c, z) \geq q(\underline{v}, c, z) \quad \text{and} \quad q(\underline{c}, v, z) \geq q(\overline{c}, v, z).$$

Note that, regardless of the queueing protocol, $q(\underline{v}, \underline{c}, -\tau^*_{PM}) = q(\overline{v}, \underline{c}, -\tau^*_{PM}) = 1$ and, otherwise, $q(\underline{v}, c, z) = 0$ and $q(\overline{v}, c, z) \geq 0$. Thus, $q(\overline{v}, c, z) \geq q(\underline{v}, c, z)$ holds for all $z \in \{-\tau^*_{PM}, \ldots, \tau^*_{PM}\}$ and $c \in \{\underline{c}, \overline{c}\}$ as required. Similarly, regardless of the queueing protocol, we have $q(\overline{c}, \overline{v}, \tau^*_{PM}) = q(\underline{c}, \overline{v}, \tau^*_{PM}) = 1$ and $q(\overline{c}, v, z) = 0$ and, otherwise, $q(\underline{c}, v, z) \geq 0$. Thus, $q(\overline{c}, v, z) \geq q(\underline{c}, v, z)$ holds for all $z \in \{-\tau^*_{PM}, \ldots, \tau^*_{PM}\}$ and $v \in \{\underline{v}, \overline{v}\}$ and we have a monotone allocation rule as required.

We prove the second statement of the theorem by assuming that when multiple identical efficient traders are present in period $t$, the traders that arrived in period $t$ have absolute priority. Otherwise, a first-come-first-served queueing protocol is used. This is a convenient choice of queueing protocol because it ensures that in *most* of the cases we delineate below, agents trade or leave the market upon arrival. Moreover, a stored efficient type trades in a given period if and only if it is first in the queue and an appropriate suboptimal pair arrives. Suppose that $y \in \{1, \ldots \tau^*_{PM}\}$ efficient traders are stored. Then the expected number of periods $Z(y)$ until the $y$th stored efficient agent trades follows a negative binomial distribution that counts the number of periods until $y$ appropriate suboptimal pairs have arrived. Setting $\lambda = w(1-w)$, for $i \geq y$, the probability mass function of $Z(y)$ is

$$\mathbb{P}(Z(y) = i) = \binom{i-1}{y-1} \lambda^y (1-\lambda)^{i-y}.$$

Therefore, using Mathematica we can compute the discounted probability of trade (or, equivalently, the expected discounted allocation) for the $y$th trader in the queue. We have

$$\sum_{i=y}^{\infty} \delta^i \mathbb{P}(Z(y) = i) = \left( \frac{\delta\lambda}{1 - \lambda + \delta\lambda} \right)^y.$$

We first explicitly compute $\boldsymbol{q}$ for arriving buyers. Specifically, for $z = \tau^*_{PM}$ and any $c \in \{\underline{c}, \overline{c}\}$,

$$q(\overline{v}, c, \tau^*_{PM}) = 1 > q(\underline{v}, c, \tau^*_{PM}) = 0;$$

for any $z \in \{0, \ldots, \tau^*_{PM} - 1\}$,

$$q(\overline{v}, \underline{c}, z) = 1 > q(\underline{v}, \underline{c}, z) = 0 \quad \text{and} \quad q(\overline{v}, \overline{c}, z) = \left(\frac{\delta\lambda}{1 - \lambda + \delta\lambda}\right)^{z+1} > q(\underline{v}, \overline{c}, z) = 0;$$

for any $z \in \{-\tau^*_{PM} + 1, \ldots, -1\}$ and $c \in \{\underline{c}, \overline{c}\}$,

$$q(\overline{v}, c, z) = 1 > q(\underline{v}, c, z) = 0;$$

and, finally, for $z = -\tau^*_{PM}$,

$$q(\overline{v}, \underline{c}, -\tau^*_{PM}) = 1 = q(\underline{v}, \underline{c}, -\tau^*_{PM}) = 1 \quad \text{and} \quad q(\overline{v}, \overline{c}, -\tau^*_{PM}) = 1 > q(\underline{v}, \overline{c}, -\tau^*_{PM}) = 0.$$

Analogous calculations apply for sellers.

A profit-maximizing P-IC and P-IR mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ that implements $\pi_{\tau^*_{PM}}$ is then completely specified once we define a payment rule. To show that there exists a payment rule involving transfers that traders pay or receive upon arrival, we compute $\boldsymbol{M}_t$ using the expected discounted allocation rule $\boldsymbol{q}$ we just computed, (18) and footnote 26. We start by deriving the payments made by buyers. Specifically, for $z = \tau^*_{PM}$ and any $c \in \{\underline{c}, \overline{c}\}$,

$$M_t^{B_t}(\overline{v}, c, z) = \overline{v} \quad \text{and} \quad M_t^{B_t}(\overline{v}, c, z) = 0;$$

for any $z \in \{0, \ldots, \tau^*_{PM} - 1\}$,

$$M_t^{B_t}(\overline{v}, \underline{c}, z) = \overline{v},\ M_t^{B_t}(\underline{v}, \underline{c}, z) = 0,\ M_t^{B_t}(\overline{v}, \overline{c}, z) = \overline{v}\left(\frac{\delta\lambda}{1 - \lambda + \delta\lambda}\right)^{z+1}, \text{ and } M_t^{B_t}(\underline{v}, \overline{c}, z) = 0;$$

for any $z \in \{-\tau^*_{PM} + 1, \ldots, -1\}$ and any $c \in \{\underline{c}, \overline{c}\}$,

$$M_t^{B_t}(\overline{v}, c, z) = \overline{v} \quad \text{and} \quad M_t^{B_t}(\underline{v}, c, z) = 0;$$

and, for $z = -\tau^*_{PM}$,

$$M_t^{B_t}(\overline{v}, \underline{c}, z) = M_t^{B_t}(\underline{v}, \underline{c}, z) = \underline{v}, \quad M_t^{B_t}(\overline{v}, \overline{c}, z) = \overline{v} \quad \text{and} \quad M_t^{B_t}(\underline{v}, \overline{c}, z) = 0.$$

Finally, we have $M_t^{B_i}(\hat{v}, c, \boldsymbol{h}_{t-1}) = 0$ for any $i \neq t$. The derivation of these payments for the sellers proceeds along similar lines. □

## B.5 Proof of Corollary 2

*Proof.* Since $\Delta_{PM} < \Delta$, the result follows immediately from Proposition 3. □

## B.6 Proof of Proposition 6

*Proof.* We start by assuming that $w > 2\Delta$ and deriving the second-best mechanism in the static bilateral trade problem. This mechanism involves always executing efficient trades and executing suboptimal trades with probability $q^{SB}$, which generates profit of $R(q^{SB}) = w^2 + 2w(1-w)q^{SB}\Delta_{PM} = w(w + 2q^{SB}(\Delta - w))$. The probability $q^{SB}$ is then pinned down by the requirement that it is the largest $q^{SB}$ that satisfies $R(q^{SB}) \geq 0$, yielding $q^{SB} = \frac{w}{2(w-\Delta)}$. Noting that $2(w - \Delta) = w + w - \Delta > w + \Delta > w$, where the first inequality follows from $2\Delta < w$ and the second from the assumption $\Delta > 0$, we have $q^{SB} < 1$.

The second-best mechanism in the static bilateral trade problem generates a per period expected surplus, denoted $S_0^{SB}$, of $S_0^{SB}(w) = w^2 + 2w(1-w)q^{SB}\Delta$. Hence, for any $\tau \geq 1$, modifying our measure of market thickness to account for the second-best mechanism in the static problem we have

$$\mathcal{T}_\tau^{SB} = \frac{S_\tau(w) - S_0^{SB}(w)}{S_\infty(w) - S_0^{SB}(w)} = \frac{2}{2\tau+1}\left(\tau + \frac{(1-q^{SB})\Delta}{1-2q^{SB}\Delta}\right) > \frac{2\tau}{2\tau+1} = \mathcal{T}_\tau,$$

where the inequality follows because $2\Delta < 1$.

Next, we can derive $\delta_{SB}^*$ by repeating our calculations from the proof of Lemma A1 for the special case of $\tau = 1$. We simply have to assume that the second-best mechanism is used in the static bilateral trade problem and replace the value function $V_0(0)$ with

$$V_0(0) = \frac{\delta(w^2 + 2w(1-w)q^{SB}\Delta)}{(1-\delta)}.$$

This yields that $\tau_{SB}^* \geq 1$ provided

$$\frac{(2\delta\Delta\lambda^2 + w^2(\delta(\lambda-1)+1) + 2\delta\lambda w)}{(1-\delta)(\delta(3\lambda-1)+1)} \geq \frac{w^2 + 2\lambda q^{SB}\Delta}{1-\delta}$$

or, equivalently,

$$\frac{(\delta-1)\Delta(w-1)w^2 + 2\delta\Delta\lambda^2(\Delta - w) - \delta\lambda(w-1)w(2\Delta + (3\Delta-2)w)}{(1-\delta)(-1+\delta-3\delta\lambda)(w-\Delta)} \geq 0.$$

Now, since $\delta \in [0,1)$, $\lambda > 0$ and $w > \Delta > 0$ we have $1 - \delta > 0$, $-1 + \delta - 3\delta\lambda < 0$ and

$w - \Delta > 0$. Hence, this last inequality is equivalent to

$$(\delta - 1)\Delta(w-1)w^2 + 2\delta\Delta\lambda^2(\Delta - w) - \delta\lambda(w-1)w(2\Delta + (3\Delta - 2)w) \leq 0.$$

Rearranging, we have $\tau^*_{SB} \geq 1$ if and only if

$$\begin{aligned}\delta \geq \delta^*_{SB} &= \frac{\Delta}{\lambda + \Delta(1-2\lambda) + (1-w)(w-2\Delta)(1+\Delta)} \\ &= \frac{1}{1 + w(1-w)\frac{(1-2\Delta)}{\Delta} + (1-w)(w-2\Delta)(1+\Delta)\frac{1+\Delta}{\Delta}}.\end{aligned}$$

Note that $(1-w)(w-2\Delta)(1+\Delta) > 0$ since $w > 2\Delta > 0$ and $w < 1$. Thus, comparing $\delta^*_{SB}$ to the expression for $\delta^*$ in (3), we have that $\delta^*_{SB} < \delta^*$ as required.

Finally, that $\tau^*_{SB} = 1$ for $\delta \in [\delta^*_{SB}, \delta^*)$ and $\tau^*_{SB} = \tau^*$ for $\delta \geq \delta^*$ follows immediately from Proposition 5. That is, since the first-best mechanism can be implemented without running a deficit whenever $\tau^* \geq 1$, using the second-best mechanism in the static bilateral trade problem only affects the decision of whether or not to store one efficient trader. □

## B.7 Proof of Proposition 7

Before we can prove Proposition 7, we need to formally introduce the class of *two-class threshold mechanisms*. Under the mechanism characterized by the threshold $\tau \geq 1$, identical efficient traders (sellers willing to sell at the Walrasian price $p$ or buyers willing to buy at $p$) are stored up to the threshold $\tau$. The two-class threshold mechanism then commences each period by posting the Walrasian price $p$ and what occurs following this depends on the state. First, consider states $z \in \{-\tau + 1, \ldots, \tau - 1\}$. Traders who arrive observe the price $p$ and state $z$ and indicate their demands and supplies as in the posted-price mechanisms analyzed in Section 4.1. If there is a pair willing to trade at $p$, an (efficient) trade is executed using a last-in-first-out queueing protocol (in the event of excess demand or excess supply).

Second, in states $z = \tau$ and $z = -\tau$, the price $p$ is posted and a sequential game is played as follows. In state $z = -\tau$, the arriving seller first indicates whether it is willing to trade at the price $p$.[31] If the arriving seller does not accept the price $p$, then that seller is cleared from the market. The buyer observes this and is given the option of trading with a stored seller at $p$. If the buyer is unwilling to trade at this price then it is cleared from the market. Otherwise, if the seller indicates willingness to trade at price $p$, the arriving buyer and the seller participate in the dominant-strategy implementation of the (one-shot)

[31] Note that accepting this price does *not* imply that the seller will receive this price in the event that they trade.

second-best mechanism for the case where the seller's cost distributions is $\frac{G(c)}{G(p)}$ for $c \in [0,p]$ and the buyer's distribution is $F(v)$ for $v \in [0,1]$. Thus, the traders that arrive to the market in state $z = -\tau$ always end up leaving the market during that period, regardless of any trading outcomes. Analogously, in state $z = \tau$, the arriving buyer first indicates whether it is willing to trade at $p$. If the buyer rejects this price, the arriving seller observes this and is given the option of trading with a stored buyer at the price $p$. Otherwise, if the buyer indicates willingness to trade at $p$, the arriving seller and the buyer participate in the dominant-strategy implementation of the (one-shot) second-best mechanism for the case when the seller's cost distribution is $G(c)$ for $c \in [0,1]$ and the buyer's distribution is $\frac{F(v)}{1-F(p)}$ for $v \in [p,1]$. Whether or not they trade, both arriving traders are cleared from the market.

*Proof.* The proof that two-class threshold mechanisms satisfy P-IC and P-IR follows along the same lines as for the proof that truthful reporting is a weakly dominant strategy for traders in the posted-price mechanisms from Proposition 5. This works because traders all face a price of $p$, except for those that participate in the second-best mechanism (this will be addressed shortly). The last-in-first-out queueing protocol ensures that no efficient type can misreport and subsequently participate in the second-best mechanism. If a suboptimal type misreports and ends up participating in the second-best mechanism, this guarantees that trader a non-positive payoff. Consider then the traders who participate in the second-best mechanism. Notice that no trader has the choice between participating in the second-best mechanism or the posted-price mechanism but rather only between participating in the second-best mechanism and not participating at all. Hence, from the perspective of these traders, this is a one-shot game and the dominant-strategy properties of these mechanisms imply that the overall mechanism satisfies P-IC and P-IR as required.

Given this equilibrium behaviour, the dynamics of the Markov decision problem are the same as those derived in Section 3 and the stationary distribution of the Markov chain $\{Y_t\}_{t\in\mathbb{N}}$ over the number of stored efficient traders is still as specified in Theorem 1. It only remains to establish that $\mathcal{T}_\tau^{TC} > \mathcal{T}_\tau$ for $\tau \geq 1$.

Denote by $\Phi(v) = v - \frac{1-F(v)}{f(v)}$ and $\Gamma(c) = c + \frac{G(c)}{g(c)}$ the virtual types and, for $\alpha \in [0,1]$, by $\Phi_\alpha(v) = v - \alpha\frac{1-F(v)}{f(v)}$ and $\Gamma_\alpha(c) = c + \alpha\frac{G(c)}{g(c)}$ the weighted virtual types. To simplify the proof we assume these functions are increasing.[32] Let $Q^\alpha(v,c) = 1$ if $\Phi_\alpha(v) \geq \Gamma_\alpha(c)$ and $Q^\alpha(v,c) = 0$ otherwise. It is well-known (see, e.g., Myerson and Satterthwaite, 1983) that the second-best mechanism when the traders draw their types from the prior distributions $F$ and $G$ is characterized by the smallest value $\alpha \in [0,1]$ such that $\int_0^1\int_0^1(\Phi(v) - \Gamma(c))Q^\alpha(v,c)dG(c)dF(v) = 0$. Denote by $\alpha_0$ this value of $\alpha$, where $\alpha_0 > 0$

[32]If these functions are not increasing, one simply replaces $\Phi_\alpha(v)$ and $\Gamma_\alpha(c)$ by their ironed counterparts, and proceed with the rest of the proof in precisely the same manner.

holds due to the impossibility of ex post efficient trade. Likewise, when the seller's cost is drawn from the distribution that is truncated at $p$, the second-best mechanism is characterized by an $\alpha_L$ such that

$$\frac{1}{G(p)}\int_0^1\int_0^p(\Phi(v)-\Gamma(c))Q^{\alpha_L}(v,c)\,dG(c)\,dF(v)=0.$$

Notice that $\int_0^1\int_0^p(\Phi(v)-\Gamma(c))Q^{\alpha_0}(v,c)\,dG(c)\,dF(v)>0$ since the original distribution of the seller's cost first-order stochastic dominates the truncated distribution. Hence, $\alpha_L<\alpha_0$. Similarly, when the buyer's value is drawn from the distribution that is truncated at $p$, the second-best mechanism is characterized by an $\alpha_H<\alpha_0$ such that

$$\frac{1}{1-F(p)}\int_p^1\int_0^1(\Phi(v)-\Gamma(c))Q^{\alpha_H}(v,c)\,dG(c)\,dF(v)=0.$$

Finally, $S_0^{SB}=\int_0^1\int_0^1(v-c)Q^{\alpha_0}(v,c)\,dG(c)\,dF(v)$ is expected welfare under second-best in the one-shot bilateral trade problem. This welfare can be expressed as $S_0^{SB}=w^2(\overline{v}-\underline{c})+w(1-w)(\Delta_L^0+\Delta_H^0)-w^2\varepsilon^0$ where, of course, $G(p)=w=1-F(p)$ and

$$\overline{v}-\underline{c}=\frac{\int_p^1\int_0^p(v-c)dG(c)dF(v)}{(1-F(p))G(p)},\quad \varepsilon^0=\frac{\int_p^1\int_0^p(v-c)(1-Q^{\alpha_0}(v,c))dG(c)dF(v)}{(1-F(p))G(p)},$$
$$\Delta_L^0=\frac{\int_0^p\int_0^p(v-c)Q^{\alpha_0}(v,c)dG(c)dF(v)}{F(p)G(p)},\quad \Delta_H^0=\frac{\int_p^1\int_p^1(v-c)Q^{\alpha_0}(v,c)dG(c)dF(v)}{(1-F(p))(1-G(p))}.$$

As before, $S_\infty=w(\overline{v}-\underline{c})$, so

$$S_0^{SB}=wS_\infty+w(1-w)(\Delta_L^0+\Delta_H^0)-w^2\varepsilon^0. \tag{21}$$

Letting $s^{TC}(y)$ denote expected surplus in state $y$ under the two-class mechanism characterized by the threshold $\tau\geq 1$, we have $s^{TC}(0)=wS_\infty$ and, for any $y\in\{1\ldots,\tau-1\}$, $s^{TC}(y)=S_\infty$. Let $s_L^{TC}(\tau)$ and $s_H^{TC}(\tau)$ denote expected surplus in state $\tau$ when sellers and buyers are stored, respectively. For $i\in\{L,H\}$ we have

$$s_i^{TC}(\tau)=w(\overline{v}-\underline{c})+w(1-w)(\overline{v}-\underline{c}+\Delta_i^{TC}-\varepsilon_i)=S_\infty+w(1-w)(\overline{v}-\underline{c}+\Delta_i^{TC}-\varepsilon_i)$$

where

$$\varepsilon_L = \frac{\int_p^1 \int_0^p (v-c)(1-Q^{\alpha_L}(v,c))dG(c)dF(v)}{(1-F(p))G(p)},\ \varepsilon_H = \frac{\int_p^1 \int_0^p (v-c)(1-Q^{\alpha_H}(v,c))dG(c)dF(v)}{(1-F(p))G(p)},$$

$$\Delta_L^{TC} = \frac{\int_0^p \int_0^p (v-c)Q^{\alpha_L}(v,c)dG(c)dF(v)}{F(p)G(p)}, \quad \Delta_H^{TC} = \frac{\int_p^1 \int_p^1 (v-c)Q^{\alpha_H}(v,c)dG(c)dF(v)}{(1-F(p))(1-G(p))}.$$

Letting $S_\tau^{TC}$ denote the expected per-trader surplus for the two-class mechanism under the stationary distribution we thus have

$$s^{TC}(\tau) = \frac{1}{2}(s_L^{TC}(\tau) + s_H^{TC}(\tau)) = S_\infty + \frac{w(1-w)}{2}(2(\overline{v}-\underline{c}) + \Delta_L^{TC} + \Delta_H^{TC} - \varepsilon_L - \varepsilon_H)$$
$$\Rightarrow S_\tau^{TC} = (1-\kappa_0)S_\infty + \kappa_0 w S_\infty + \kappa_0(1-w)(2S_\infty + w(\Delta_L^{TC} + \Delta_H^{TC} - \varepsilon_L - \varepsilon_H)),$$

which yields

$$\mathcal{T}_\tau^{TC} = \frac{S_\tau^{TC} - S_0^{SB}}{S_\infty - S_0^{SB}} = 1 - \kappa_0(1-w) + \frac{\kappa_0(1-w)\left(2S_\infty w(\Delta_L^{TC} + \Delta_H^{TC} - \varepsilon_L - \varepsilon_H) - S_0^{SB}\right)}{S_\infty - S_0^{SB}}.$$

To show that $\mathcal{T}_\tau^{TC} > \mathcal{T}_\tau = 1 - \kappa_0$ it thus suffices to show that the second term on the right-hand side of this last equation is positive. Using (21) and regrouping terms we have

$$2S_\infty + w(\Delta_L^{TC} + \Delta_H^{TC} - \varepsilon_L - \varepsilon_H) - S_0^{SB} = (1-w)(S_\infty - w\varepsilon^0) + (S_\infty - w\varepsilon_H)$$
$$+ w(\Delta_L^{TC} + \Delta_H^{TC} - (\Delta_L^0 + \Delta_H^0)) + w(\varepsilon^0 - \varepsilon_L) + w^2(\Delta_L^0 + \Delta_H^0).$$

Observe that for all $i \in \{0, L, H\}$ we have $\varepsilon_i < \overline{v} - \underline{c}$ which implies that $S_\infty - w\varepsilon_i > 0$. Moreover, for $i \in \{L, H\}$ we have $\Delta_i^0 < \Delta_i^{TC}$ and $\varepsilon_i < \varepsilon^0$ since $\alpha_i < \alpha_0$. Therefore, the right-hand side of the last equation is positive and we have $\mathcal{T}_\tau^{TC} > \mathcal{T}_\tau$ as required. □

# C Posted-price dynamics

Given a pricing rule, we can characterize the stationary price distribution and derive price dynamics. For example, consider the midpoint pricing rule $p^M$ with $p^M(z) = \frac{1}{2}$ for $z \in \{-\tau^*+1, \ldots, \tau^*-1\}$, $p^M(\tau^*) = 1 - \frac{\Delta}{2}$ and $p^M(-\tau^*) = \frac{\Delta}{2}$. Let $P_t$ denote the price posted in period $t$. By Theorem 1, under the stationary distribution we have $\mathbb{P}(P_t = \underline{v}) = \mathbb{P}(P_t = \overline{c}) = \frac{1}{2\tau^*+1}$ and $\mathbb{P}(P_t = \frac{1}{2}) = \frac{2\tau^*-1}{2\tau^*+1}$. It follows that the variance of the posted prices under the stationary distribution is $\mathrm{Var}(P_t) = \frac{(1-2\Delta)^2}{2(2\tau^*+1)}$. We then have the following comparative statics.

**Lemma C3.** *The probability $\mathbb{P}(P_t = \frac{1}{2})$ increases in $\delta$ and $w(1-w)$ and decreases in $\Delta$.*

*The variance* $\mathrm{Var}(P_t)$ *decreases in* $\delta$ *and* $w(1-w)$*. Moreover,* $\lim_{\delta\to 1}\mathbb{P}(P_t=\frac{1}{2})=1$ *and* $\lim_{\delta\to 1}\mathrm{Var}(P_t)=0$.

This lemma show that, as can be seen in Figure 3, the volatility of the posted prices decreases as the storage threshold $\tau^*$ increases. The comparative statics results of Lemma C3 largely mirror the comparative statics of $\tau^*$. The exception is the effect of the value of a suboptimal trade, $\Delta$, on the variance, which cannot be signed in general.[33] These results also resonate with convergence results in the literature on the microfoundation of competitive equilibria. Satterthwaite and Shneyerov (2007) and Lauermann (2013) provide sufficient conditions for equilibria in dynamic search and matching settings to converge to the (static) Walrasian equilibrium as search frictions—parametrized by a discount factor—vanish. A subtle but important difference in our setting is that the posted-price mechanism is efficient for any $\delta$, provided only that $\delta \geq \delta^*$.

[33] An increase in $\Delta$ both shifts probability mass away from $\frac{1}{2}$ and narrows the gap between the lowest price $\Delta$ and the highest price $1-\Delta$ in the support.