Optimal market thickness and clearing*

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

This version: July 3, 2020 First version: February 15, 2016

Abstract

Traders that arrive over time give rise to a dynamic tradeoff between the benefits of increasing market thickness by accumulating traders and the associated cost of delayed trade 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. By storing traders and allowing them to trade in the future, market thickness increases. For symmetric binary type distributions, optimally thick markets involve storing surprisingly few traders. Nevertheless, their performance is close to that of a large market: for any type distributions, two-thirds of the gains from increased market thickness are achieved by storing just one trader. The benefits of storing traders are even larger if traders' types are their private information and second-best mechanisms are used. For large discount factors, clearing the market at optimally chosen fixed frequencies reaps most gains from dynamic market clearing.

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

*We are grateful to the editor (Alessandro Pavan) and five anonymous reviewers of this journal for feedback that helped us improve the paper. We also thank 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, Carlos Oyarzun, Andy Skrzypacz, Alexander Teytelboym, 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

Thicker markets reduce individual traders' price impacts and their incentives and scope for exerting market power. By improving sorting, they also increase the probability of efficient trades. Not surprisingly, increasing market thickness is at the heart of a wide range of policy areas, from competition and antitrust policy to market design, as well as economic theory and analysis. Notwithstanding its prominence, little is known about the optimal degree of market thickness and the factors that determine it when increasing market thickness comes with costs—such as delayed trade when agents arrive over time—as well as benefits.[1] Accordingly, practitioners have received little to no guidance from economists concerning this important aspect of market design and have, perhaps as a consequence, paid little attention to the economic tradeoffs that are involved.[2]

To shed light on the dynamic tradeoff between the benefits and costs of increasing market thickness by accumulating traders that arrive over time, this paper studies a dynamic version of the classical bilateral trade problem of Myerson and Satterthwaite (1983) in which a buyer-seller pair arrives in each period. The designer has the option of storing arriving traders, thereby increasing market thickness by allowing them to trade in future periods.

For concreteness, assume buyers and sellers draw their types independently from symmetric binary distributions. Two static benchmarks that are useful, among other things, to illustrate the basic assumptions. First, in the *perfectly thick* large market limit an infinite number of traders are present at once and only high-value buyers and low-cost sellers trade under ex post efficiency. Second, in the *perfectly thin* bilateral trade setting, a single buyer and seller are present and the only trade that is not executed under ex post efficiency is one between a low-value buyer and a high-cost seller. In our dynamic setting, the designer thus has an incentive to accumulate efficient traders (high-value buyers or low-cost sellers) in the hope of rematching them in the future to create an efficient trade (between a high-value buyer and a low-cost seller).

We show that in the symmetric binary type setting the optimal market clearing policy

[1] While we focus on dynamic markets, this tradeoff arises in many other contexts. For example, market thickness can be increased by merging geographically dispersed markets and allowing distant traders to interact, by "conflation" that treats similar yet distinct goods as identical and by increasing market transparency with improved information about offers from buyers and sellers.

[2] For example, the transition from paper- to computer-organized trading at the New York Stock Exchange was exclusively driven by the programmer's desire to execute trades as fast as possible without any consideration of the tradeoff between speed and market thickness. Likewise, the Native Vegetation Exchange for Victoria, Australia, was designed to execute compatible trades instantaneously, not on the grounds that this would be optimal but due to computational complexity. Similarly, eBay's clearing mechanism does not allow traders to accumulate, which, as documented by Hendricks and Sorensen (2018), results in substantial welfare losses.

is pinned down by a threshold that specifies how many identical, efficient traders to store. Efficient trades are executed as soon as they become available. Suboptimal trades (between a high-value buyer and a high-cost seller or a low-value buyer and a low-cost seller) are only executed once the storage threshold is reached. This allows us to characterize the optimal degree of market thickness. When measured by the maximum number of stored traders (i.e. the optimal storage threshold), optimally thick markets are surprisingly thin and appear to be very different from the perfectly thick large market limit.[3] Alternatively, market thickness can be measured by how much of the difference between per trader-pair welfare in the large market limit and the bilateral trade setting is captured by storing a threshold number of efficient traders. By this measure, the performance of seemingly thin markets is surprisingly similar to that of perfectly thick markets: two-thirds of the maximum increase in market thickness is achieved as the storage threshold increases from zero to just one. Perhaps even more surprisingly, this result does not depend on distributional assumptions. We show that for continuous type distributions, two-thirds of these gains can be captured by a dynamic policy that stores at most one analogously defined efficient trader, and that these gains are even larger when accounting for incentive issues and second-best mechanisms.

In the binary type setting, the efficient policy can be implemented with a posted-price mechanism that always balances the budget as soon as it is efficient to store at least one efficient trader. This provides a dynamic reconciliation of the Coase Theorem with the insight, due to Vickrey (1961), Hurwicz (1972), and Myerson and Satterthwaite (1983), that in static settings private information can be an insurmountable transaction cost. In our setting when the efficient policy is non-trivially dynamic, all transaction costs that may hinder efficient allocation in static settings are eliminated.[4] Another way of conveying the extent to which dynamics matter is our finding that, provided the discount factor is sufficiently large, a designer that maximizes its expected discounted profit generates more expected discounted social surplus than periodic ex post efficient bilateral trade. These results also highlight a fundamental difference between the driving forces behind market thickness in our setting and those in the double-auctions literature, whose focus has been on inefficiencies due to incentive compatibility, individual rationality and no-deficit constraints. In particular, if these were the only constraints that mattered, optimal market thickness would be achieved by storing one pair of traders. In our setting, this is not the case as the storage threshold increases with the discount factor.

[3] Perfectly thick markets with infinitely many traders serve as the canonical model of a competitive market and are often portrayed as the ideal market.

[4] Of course, with discrete types, whether efficient allocation is possible in the static setting depends on the details of the model. Remarkably, this possibility result as well as our measures of and results pertaining to market thickness are detail-free.

The posted-price implementation of the efficient dynamic policy also allows us to define an agent's price impact as the probability that, under the stationary distribution, the arrival of an agent changes the price that will be posted in the next period. Interestingly, price impact, which is widely used in applied, often atheoretical work, is an equivalent measure of market thickness in our setting. Finally, in the limit as the discount factor approaches one, the precise nature of market clearing—discriminatory, which is first-best, uniform or fixed frequency—does not affect the asymptotic efficiency of the market, provided these different dynamic market clearing regimes are determined optimally. In sharp contrast, the welfare loss from periodic ex post efficient bilateral trade relative to an optimal dynamic market clearing regime diverges as the discount factor approaches one.

The remainder of this paper is organized as follows. Section 2 introduces the model. In Section 3 we derive the optimal degree of market thickness by solving the Markov decision process of a planner who maximizes expected discounted social surplus. Section 4 maps this solution to dynamic mechanisms and accounts for incentive issues. Section 5 analyzes alternative but asymptotically optimal market clearing regimes. Related literature is discussed in Section 6 while Section 7 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 and draw their types independently from symmetric, binary distributions. For most of the paper, we assume that buyers draw their values from a distribution with support $\mathcal{V} = \{\underline{v}, \overline{v}\}$ and sellers draw their costs from a distribution with support $\mathcal{C} = \{\underline{c}, \overline{c}\}$. 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 $\overline{v} - \overline{c} = \Delta_0$ and $\underline{v} - \underline{c} = \Delta_0$. We normalize $\overline{v} = 1$ and $\underline{c} = 0$. In line with Myerson and Satterthwaite (1983), we ensure that some (but not all) trade is ex post efficient in the static bilateral trade problem by assuming $\overline{v} > \overline{c} > \underline{v} > \underline{c}$. Under the stated assumptions and normalizations this amounts to assuming $\Delta_0 \in (0, 1/2)$ and ensures that both the static and dynamic problems are non-trivial.[5]

Each buyer and seller has quasilinear utility and demands and supplies at most one unit, respectively. The value of agents' outside option of not participating is zero. Following Myerson and Satterthwaite (1983) we further assume that the agents can trade only via an

[5]If $\Delta_0 > 1/2$ all trade would be efficient in both the static bilateral setting and the static large market limit with infinitely many traders. If $\Delta_0 < 0$, only trades between high-value buyers and low-cost sellers would be efficient in both the bilateral trade problem and the large market limit; in either case, it could be implemented by setting a price of $1/2$ and rationing the long side of the market.

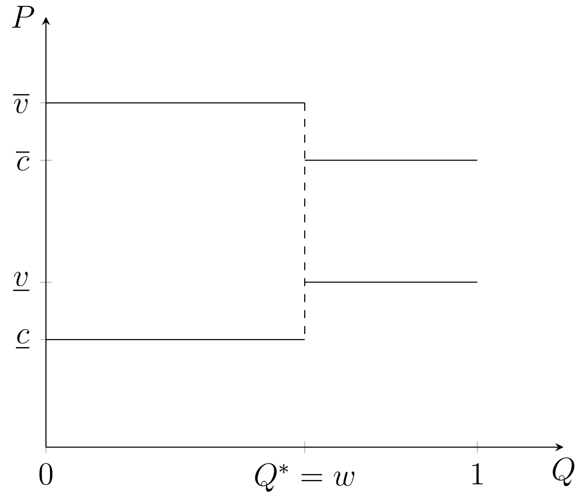


Figure 1: The Walrasian quantity in the perfectly thick large market limit.

exchange that is operated by an entity that we call the designer. All agents and the designer are risk-neutral geometric discounters, with a common discount factor $\delta \in [0, 1)$, and the arrival process, discount factor δ and type distributions are common knowledge. Arrivals are observable.

Two static benchmarks are useful. In the *perfectly thick large market limit*, illustrated in Figure 1, there are an infinite number of traders. Half of them are buyers and half of them sellers, each drawing their type independently from the distributions introduced above. This implies that the fraction of agents who trade is w, and the average social surplus per buyer-seller pair is $S_\infty(w) = w$. Further, the price $p = 1/2$ will be market-clearing and implement the efficient allocation.[6] In contrast, in the *perfectly thin bilateral trade* setup all trades are ex post efficient with the exception of the trade between a low-value buyer and a high-cost seller. The expected gain in social surplus from the static bilateral trade problem is thus $S_0(w) = w^2 + 2w(1 - w)\Delta_0$. The difference

$$S_\infty(w) - S_0(w) = (1 - w)(w - 2w\Delta_0) > 0 \tag{1}$$

is the maximum gain, per trading pair, that can be achieved from increasing market thickness.

Given the central role of efficiency in economics, a natural objective for the designer is to maximize expected discounted social surplus. Given this objective, it is always optimal to immediately execute any available *efficient trade*—that is, a trade between a high-value

[6]Of course, any price $p \in [\underline{v}, \overline{c}]$ will be market-clearing if one allows the agents to express demand and supply correspondences, but it is useful to focus on market clearing prices in the interior of the Walrasian price gap.

buyer and a low-cost seller. In contrast, for any $\delta \in (0,1)$, the decision to execute an available *suboptimal trade*—that is, a trade between a high-value buyer and a high-cost seller or between a low-value buyer and a low-cost seller—requires trading off the instantaneous gain of Δ_0 versus the discounted, expected value of being able to execute a trade of value 1 in the future. Suboptimal $(\overline{v}, \overline{c})$ and $(\underline{v}, \underline{c})$ trades occur with probability 0 in the large static market and with probability $2w(1-w)$ under ex post efficiency in the static bilateral trade setup. In this sense, the reduction in the probability of suboptimal trades provides an alternative, observation-based measure of market thickness.

Apart from dynamics and symmetry, the above assumptions depart from the classical Myerson-Satterthwaite setup only by assuming binary types, which considerably simplifies the exposition and analysis. We now show how the binary type setup connects to the setup with continuous distributions that is standard in static mechanism design. Suppose, temporarily, that the buyers and sellers draw their values and costs independently from continuous distributions F and G, whose densities are positive on their supports. Without loss of generality, we can assume they have the same support, which we take to be $[0,1]$.[7] This implies that the large market Walrasian price p, which satisfies $1 - F(p) = G(p)$, is in the interior of the support. This is in line with the Myerson-Satterthwaite assumption that some but not all trade is ex post efficient in the one-shot bilateral trade problem. The connection to the binary setup is then that $w = G(p) = 1 - F(p)$ and the continuous-distribution analogue to buyers with type $\overline{v}$ and sellers with cost $\underline{c}$ are buyers with value $v \geq p$ and sellers with costs $c \leq p$, respectively. The analogue to the welfare gains $\overline{v} - \underline{c}$ from an efficient trade is $\int_p^1 \int_0^p (v-c) dG(c) dF(v)/w^2$. An average suboptimal "trade"—which accounts for the possibility that not trading is efficient—creates expected social surplus of

$$\frac{\int_0^p \int_c^1 (v-c) dF(v) dG(c) + \int_p^1 \int_0^v (v-c) dG(c) dF(v)}{2w(1-w)}$$

and is the analogue of Δ_0. Varying w provides a parsimonious way of capturing changes in distributions in the setting with continuous types as well. For example, $w = 1/2$ occurs if $F = G$. We will return to this specification with continuous distributions and show that most of our key insights, other than optimality, extend to this setting. Assuming that the binary types are symmetric simplifies the derivation of optimal market thickness as it means the optimal policy will be the same with regards to storing either kind of *suboptimal* pair.[8]

[7] As long as the supports overlap, the Walrasian price p in the large market limit remains in the interior of both supports, and some but not all trade is ex post efficient in the one-shot bilateral trade problem. Without overlapping supports, there is no problem to solve.

[8] The continuous-distribution analogue is that F and G satisfy $\int_0^p \int_c^1 (v-c) dF(v) dG(c) = \int_p^1 \int_0^v (v-c) dG(c) dF(v)$, which is, for example, the case when F and G are uniform (and more generally when $F = G$

In Section 3, we will formally define threshold policies and show that these are optimal and induce a unique stationary distribution. Denoting by $S_\tau(w)$ the expected per trader-pair social surplus gain under the stationary distribution given a policy with threshold τ, our welfare-based measure of market thickness is

$$T_\tau := \frac{S_\tau(w) - S_0(w)}{S_\infty(w) - S_0(w)}, \tag{2}$$

where the difference $S_\infty(w) - S_0(w)$ is given by (1). This is a natural measure of market thickness that expresses the increase in market thickness under the policy with threshold τ as a fraction of the maximum achievable increase.[9] An alternative, observation-based measure of market thickness is one that involves the probability of a suboptimal trade occurring. Let $\mathbb{P}_S(\tau)$ denote the per period probability of such a trade under the policy with threshold τ and $\mathbb{P}_S(0) = 2w(1-w)$ denote this probability in the static bilateral trade problem. Then noting that suboptimal trades occur with probability zero in the static large market limit, an alternative measure of market thickness is given by

$$\rho_\tau := \frac{\mathbb{P}_S(0) - \mathbb{P}_S(\tau)}{\mathbb{P}_S(0)} = 1 - \frac{\mathbb{P}_S(\tau)}{2w(1-w)}. \tag{3}$$

As we will show, under any policy with threshold τ, the two measures are equivalent and we have $T_\tau = \rho_\tau$. Furthermore, these measures of market thickness are also tightly connected to an agent's likely price impact.

As is standard, we assume that values and costs are private information of the agents. Ultimately, we will therefore solve the dynamic mechanism design problem in which the designer elicits this information via an incentive compatible and individually rational mechanism. In Section 4 we will be precise about the notions of incentive compatibility and individual rationality and derive, among other things, the profit-maximizing mechanisms that satisfy these constraints. However, before doing so, we formulate and solve the Markov decision problem of a planner who maximizes expected discounted social surplus. Since the mechanism design problem of maximizing profit is isomorphic to that of maximizing social surplus (once we replace the true types with virtual types), this analysis will be equally valid and useful for solving for the profit-maximizing mechanisms.

and their density f is symmetric in the sense that, for all $x \in [0,1]$, $f(x) = f(1-x)$).

[9] Although T_τ is defined with respect to a threshold policy, it is clear from its construction that in essence it is a measure that is independent of policies and mechanisms: Any policy that gives rise to a stationary distribution allows one to compute the expected per period welfare, denoted $S_\tau(w)$, and thereby T_τ. Moreover, it is independent of the type space.

3 Optimal market thickness

We now derive the optimal market thickness when the designer's objective is to maximize expected discounted social surplus. For the purposes of this section, we assume that the designer knows the types of agents who have arrived while, of course, facing uncertainty about the types that will arrive in the future.

3.1 Threshold policies and their properties

At a fundamental level, the designer's problem is to determine which pairs of traders should be cleared from the market in each period in order to maximize expected discounted social surplus. Since $\Delta_0 < 1/2$, if suboptimal $(\overline{v}, \overline{c})$ and $(\underline{v}, \underline{c})$ pairs are present, an increase in the designer's payoff is achieved by rematching these pairs to create a $(\overline{v}, \underline{c})$ pair that generates a gain of 1 rather than $2\Delta_0$. Thus, when a suboptimal $(\overline{v}, \overline{c})$ or $(\underline{v}, \underline{c})$ pair is present, the designer has an incentive to wait, and not clear such a pair from the market, in the hope of eventually rematching pairs to create an efficient $(\overline{v}, \underline{c})$ pair. In principle, this decision depends on the entire history of agent arrivals. However, the state space can be simplified considerably by mapping the dynamic optimization problem to a Markov decision process.

The underlying state at time t is identified as follows. First, determine the number of efficient pairs present and then determine the number of identical suboptimal $(\overline{v}, \overline{c})$ or $(\underline{v}, \underline{c})$ pairs present among the remaining set of agents. The observations above imply that it cannot be optimal that non-identical suboptimal pairs, $(\overline{v}, \overline{c})$ and $(\underline{v}, \underline{c})$, are simultaneously present as these pairs can be split and rematched to form one efficient $(\overline{v}, \underline{c})$ pair and one infeasible $(\underline{v}, \overline{c})$ pair. Infeasible pairs can be ignored since these do not generate positive surplus and cannot be rematched to create efficient pairs. Thus, the *state space* of the designer's Markov decision process is 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 are the number of efficient pairs and suboptimal pairs present, respectively.

Let $\boldsymbol{X}_t \in \mathcal{X}$ denote the state of the market after the arrival of period t agents and $\mathcal{A}_{\boldsymbol{x}}$ denote the set of *actions* available to the designer in state $\boldsymbol{x}$, with $\mathcal{A} := \cup_{\boldsymbol{x} \in \mathcal{X}} \mathcal{A}_{\boldsymbol{x}}$. We have then $\mathcal{A}_{\boldsymbol{x}} = \{(a_E, a_S) : a_E, a_S \in \mathbb{Z}_{\geq 0}, a_E \leq x_E, a_S \leq x_S\}$, where a_E and a_S denote the respective number of efficient pairs and suboptimal pairs being cleared from the market. Denote by $\boldsymbol{A}_t$ the action taken by the designer in period $t \in \mathbb{N}$ and by

$$P_{\boldsymbol{a}}(\boldsymbol{x}, \boldsymbol{y}) := \mathbb{P}(\boldsymbol{X}_{t+1} = \boldsymbol{y} | \boldsymbol{X}_t = \boldsymbol{x}, \boldsymbol{A}_t = \boldsymbol{a})$$

the *transition probability* that, if the designer takes the action $\boldsymbol{a}$ in state $\boldsymbol{x}$ in period t, the state in period $t + 1$ will be $\boldsymbol{y}$. Fix a state $\boldsymbol{x}$ and a feasible action $\boldsymbol{a} = (a_E, a_S)$. Then an efficient pair arrives in the following period with probability w^2, while an infeasible one arrives with probability $(1-w)^2$. We thus have

$$P_{\boldsymbol{a}}(\boldsymbol{x},(x_E-a_E+1,x_S-a_S))=w^2 \quad \text{and} \quad P_{\boldsymbol{a}}(\boldsymbol{x},(x_E-a_E,x_S-a_S))=(1-w)^2.$$

If $x_S-a_S=0$ and a suboptimal pair arrives in the following period, then we end up in state $\boldsymbol{y}=(x_E-a_E,1)$ and have

$$P_{\boldsymbol{a}}(\boldsymbol{x},(x_E-a_E,1))=2w(1-w).$$

If $x_S-a_S>0$ and a suboptimal pair arrives, the arrival of an identical suboptimal pair will lead to the state (x_E-a_E,x_S-a_S+1), while a non-identical suboptimal pair will be rematched with a stored suboptimal pair to create an efficient pair. Consequently, we have

$$P_{\boldsymbol{a}}(\boldsymbol{x},(x_E-a_E,x_S-a_S+1))=P_{\boldsymbol{a}}(\boldsymbol{x},(x_E-a_E+1,x_S-a_S-1))=w(1-w).$$

Finally, $r(\boldsymbol{a})=a_E+\Delta a_S$ denotes the immediate *reward* when taking the action $\boldsymbol{a}\in\mathcal{A}$, where Δ is the value of a suboptimal trade to the designer. We set $\Delta=\Delta_0$ when the designer's objective is to maximize expected discounted social surplus.

Given a Markov decision process $\langle\mathcal{X},\mathcal{A},P,r,\delta\rangle$, a *policy* $\pi:\mathcal{X}\to\mathcal{A}$ is such that $\pi(\boldsymbol{x})\in\mathcal{A}_{\boldsymbol{x}}$ specifies the action taken by the designer in state $\boldsymbol{x}$. The *optimal policy* π^* of this Markov decision process maximizes the expected discounted reward earned by the planner, which by construction maximizes expected discounted social welfare. Thus, the designer's dynamic optimization problem reduces to determining the optimal policy π^* of $\langle\mathcal{X},\mathcal{A},P,r,\delta\rangle$. Since the state space $\mathcal{X}$ is countable, the feasible action sets $\mathcal{A}_{\boldsymbol{x}}$ are finite for all states $\boldsymbol{x}$ and the reward function is deterministic, 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).

To determine the optimal policy we begin by defining a simple class of policies, which we call *threshold policies*. Threshold policies immediately clear efficient pairs from the market. Identical suboptimal pairs are stored up to a threshold $\tau\in\mathbb{N}$, and any additional suboptimal pairs are cleared immediately from the market. That is, given a threshold $\tau\in\{0,1\ldots\}$, the associated threshold policy π_τ of the Markov decision process $\langle\mathcal{X},\mathcal{A},P,r,\delta\rangle$ is such that

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

Theorem 1. *The optimal policy is a threshold policy.*

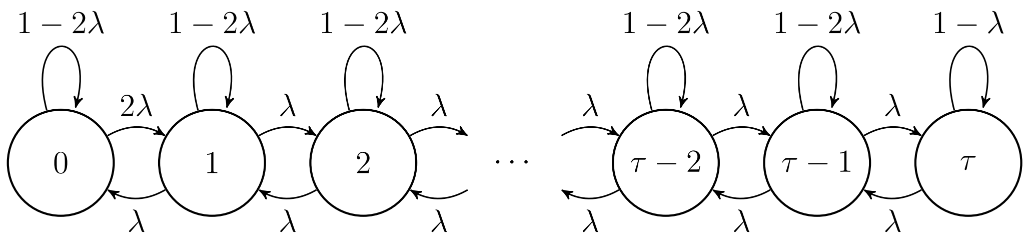


Figure 2: The Markov chain over the number of stored identical efficient traders induced by the optimal policy π^*, where $\lambda = w(1-w)$.

Whenever the arrival of a non-identical suboptimal pair leads to a stored suboptimal pair being rematched to produce an efficient trade, this also produces an infeasible $(\underline{v}, \overline{c})$ trade that is cleared from the market. Thus, we can, without loss of generality, assume from this point forward that the designer *only stores (identical) efficient traders* that arrive as part of suboptimal pairs and immediately clears the inefficient trader from the market.

Each threshold policy π_τ induces a Markov chain $\{Y_t\}_{t\in\mathbb{N}}$ over $\{0,\dots,\tau\}$, the number of identical efficient traders stored. As is illustrated in Figure 2, $\{Y_t\}_{t\in\mathbb{N}}$ is a finite birth-and-death process.

Lemma 1. *The stationary distribution $\boldsymbol{\kappa}$ of the Markov chain $\{Y_t\}_{t\in\mathbb{N}}$ under the threshold policy π_τ is given by, for all $i \in \{0,\dots,\tau\}$,*

$$\kappa_i := \mathbb{P}(Y_t = i) = \frac{1+\min\{i,1\}}{2\tau+1}.$$

Letting $s(y)$ be expected social surplus in a given period in state $y \in \{0,\dots,\tau\}$, we have $s(0) = wS_\infty(w)$, $s(y) = S_\infty(w)$ for $y \in \{1,\dots,\tau-1\}$ and $s(\tau) = S_\infty(w) + w(1-w)\Delta_0$. Since $\kappa_i = 2\kappa_0$ for $i > 0$ by Lemma 1, expected per period payoff under the stationary distribution, denoted $S_\tau(w)$, is $S_\tau(w) = (1-\kappa_0)S_\infty(w) + \kappa_0 S_0(w)$. Thus,

$$S_\tau(w) - S_0(w) = (1-\kappa_0)(S_\infty(w) - S_0(w)) \quad \text{and} \quad T_\tau = 1-\kappa_0.$$

In other words, T_τ is a unit-free measure and independent of δ, w and Δ because it is determined by the stationary distribution, which does not depend on these parameters. Substituting κ_0 and $S_\infty(w) = w$ then yields the following proposition:

Proposition 1. *Given threshold policy π_τ, we have*

$$S_\tau(w) = w^2 + \frac{2w(1-w)(\Delta+\tau)}{2\tau+1} \quad \textit{and} \quad T_\tau = \frac{2\tau}{2\tau+1}.$$

The fact that T_τ increases so quickly in τ — for example, $T_1 = 2/3$ — is as remarkable as its unit- and detail-free nature. To show that the steep increases from storing suboptimal trades do *not* depend on the distribution being discrete, let alone on it being binary, we now temporarily assume that buyers and sellers draw their values and costs independently from continuous distributions F and G that satisfy the properties introduced in Section 2, and consider a *two-class threshold* policy. This policy separates buyers and sellers into two classes: efficient traders, meaning values above and costs below the Walrasian price p in the large market limit and inefficient traders, which are those with values below and costs above p. Any trades involving an efficient buyer and seller are immediately executed and identical efficient traders are stored up to some threshold τ.

Proposition 2. *Assume that buyers and sellers draw their values and costs independently from continuous distributions F and G. Then there exists a two-class threshold policy such that T_τ as defined in* (2) *is the proportion of the maximum possible gains from increasing market thickness that is captured by storing τ identical efficient traders.*

Proposition 2 shows that the steep increases in the benefits to market thickness from storing efficient traders do not depend on the distribution being binary or symmetric. Indeed, as shown in Proposition 2, T_τ is the same for any distribution if the designer uses a two-class threshold policy that categorizes buyers and sellers according to whether their types are above or below p. What the symmetric binary distribution then affords us is not so much that the benefits from increasing market thickness accrue fast with two-class threshold policies—this is also the case for continuous type distributions—but rather that these policies are optimal.

Although the context and the conditions differ, Propositions 1 and 2 are similar in spirit to McAfee's (2002) finding that in a static, two-sided matching problem with supermodular payoffs and a continuum of agents at least half of the maximal gain (which requires infinitely many categories or matching submarkets) is achieved by creating only two categories.[10]

Under the stationary distribution, the respective per period probabilities of a suboptimal trade and an efficient trade are

$$\mathbb{P}_S(\tau) = \frac{2w(1-w)}{2\tau+1} \quad \text{and} \quad \mathbb{P}_E(\tau) = w^2 + \frac{2\tau w(1-w)}{2\tau+1}.$$

[10] If the Myerson-Satterthwaite assumption that some but not all trade is ex post efficient holds in a market setting with quasilinear payoffs, then there is an element of supermodularity in that surplus increases if buyers with values above the market clearing price are exclusively matched to sellers with costs below it. However, aggregate surplus is not affected by how these agents are matched to each other. In that sense, a Walrasian market is a special case of what McAfee (2002) studies: an appropriately structured coarse matching achieves one hundred percent of all the welfare gains.

Note that $\mathbb{P}_S(\tau)$ is decreasing in τ and satisfies $\mathbb{P}_S(0) = 2w(1-w)$ and $\lim_{\tau\to\infty} \mathbb{P}_S(\tau) = 0$, while $\mathbb{P}_E(\tau)$ increases in τ and satisfies $\mathbb{P}_E(0) = w^2$ and $\lim_{\tau\to\infty} \mathbb{P}_E(\tau) = w$.[11] Thus, the decrease in this probability as a percentage of $\mathbb{P}_S(0)$, denoted ρ_τ and introduced in (3), satisfies

$$\rho_\tau = \frac{2\tau}{2\tau+1} = T_\tau. \tag{4}$$

Therefore, these two natural measures of market thickness are equivalent.

3.2 Optimal policies

Theorem 1 gives rise to a tractable dynamic programming approach that can be used to characterize the optimal threshold, which we denote by τ^*. This is laid out in detail in the proof of Proposition 3, which includes an analytic characterization of τ^* in Corollary A1. Given τ^*, the *optimal market thickness* is $T^* := T_{\tau^*}$. The following proposition describes how T^* varies with the parameters w, Δ and δ:

Proposition 3. *The optimal market thickness T^* increases in $w(1-w)$ and is maximized by the uniform distribution, that is, by $w = 1/2$. It also increases in δ and decreases in Δ.*

Not surprisingly, T^* increases as Δ decreases and δ increases since either effect decreases the opportunity cost of waiting. In contrast, the fact that T^* is maximized at $w = 1/2$ is more subtle as the benefits of increasing market thickness decrease in w. To see this, notice that $S_0(w)/S_\infty(w) = w + 2w(1-w)\Delta$ measures the fraction of the per trader-pair surplus in the perfectly thick market that is captured under ex post efficiency in the bilateral trade problem. It is increasing in w, equal to 2Δ at $w = 0$ and equal to 1 at $w = 1$. Thus,

$$\frac{1 - S_0(w)/S_\infty(w)}{1 - 2\Delta} = 1 - w$$

is a natural measure of the benefits of increasing market thickness as a function of w. That T^* is maximized at $w = 1/2$ therefore relates to the costs of increasing market thickness (that is, to the expected delay). The very point of storing an efficient trader (say, a high-value buyer) who arrives as part of a suboptimal pair is to form and execute an efficient trade in the next event in the future that a suboptimal pair of the opposite kind (say, containing a low-cost seller) arrives. The probability that such a pair arrives in any given period being

[11] Interestingly, the per period probability of *any* trade, $\mathbb{P}_E(\tau) + \mathbb{P}_S(\tau) = w^2 + w(1-w)(2\tau+2)/(2\tau+1)$, is decreasing in τ. So, seemingly paradoxically, seeing less frequent trade is an indication of market thickness. However, the overall probability of trade is an unreliable measure of efficiency even in static settings. For example, if the buyers and sellers draw their types from the same continuous distribution, each agent trades with probability 1/2 both under ex post efficiency in the bilateral trade setup and in the large market limit.

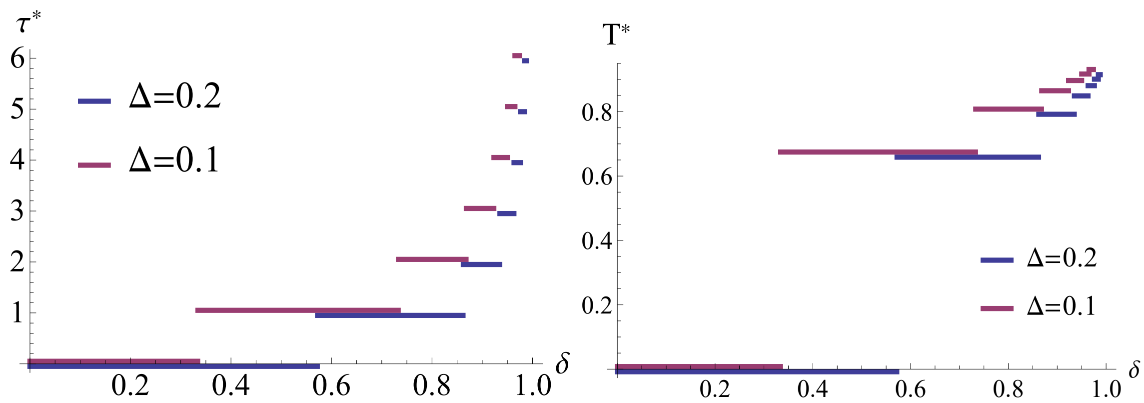


Figure 3: A numerical illustration of τ^* (left panel) and T^* (right panel) as a function of δ for $w = 0.5$ and $\Delta = 0.1, 0.2$.

$w(1-w)$, it follows that the expected delay is minimized by the uniform distribution. In other words, optimal market thickness is maximized under the uniform distribution because it maximizes the probability of a mismatch.[12]

3.3 The apparent thinness of optimally thick markets

As noted, market thickness, measured by T_τ, increases quickly in τ and satisfies $T_1 = 2/3$. In other words, storing only one suboptimal trade achieves two-thirds of the maximally achievable increase in market thickness. Moreover, the marginal product of τ, $T_{1+\tau} - T_\tau = 2/(1+2(1+\tau)^2)$, decreases fast in τ, which explains why optimally thick markets are rather thin when measured by the maximum number of traders present. For example, for the parameterizations displayed in Figure 3, $\tau^* \leq 6$ for all $\delta \leq 0.95$. While $\tau^* = 6$ is a small number, particularly in comparison to the large market limit widely used to describe and analyze competitive markets, by $T_6 = 0.923$, that is, a threshold of six already 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 large market with infinitely many traders, its welfare properties are remarkably close to those of the large market limit. This also implies that there are steeply diminishing returns from storing additional efficient traders and one might therefore intuitively expect that market

[12]The feature that the designer is more willing to wait if the probability of a mismatch is large is not specific to the present setup. As shown in Loertscher et al. (2017) for a setup with two periods and continuous distributions, relative to trades the designer would execute in a static setting, the designer restricts trades the least in the first period when the buyer's and seller's (virtual) types are close.

thickness grows slowly in δ. Formally, letting

$$\overline{\tau}(\delta,\Delta) = \frac{\log\left(\frac{1-\Delta}{\Delta}\right)}{1-\delta} + \frac{\log\left(\frac{\Delta}{1-\Delta}\right)}{2},$$

the following proposition provides an upper bound on τ^*:

Proposition 4. $\tau^* \le \overline{\tau}(\delta,\Delta) + O(1-\delta)$.

Proposition 4 shows that optimal market thickness grows slowly—at rate $1/(1-\delta)$—as $\delta \to 1$. Measured by the number of traders, optimally thick markets are thin.[13] From a practical perspective, Proposition 4 implies computational tractability.

4 Dynamic mechanisms

We now derive and analyze dynamic mechanisms that account for the agents' private information and implement the welfare- and profit-maximizing policies. We also analyze posted-price mechanisms and their dynamics.

4.1 Direct mechanisms

By the revelation principle, we can restrict attention to direct, truthful mechanisms without loss of generality. A direct 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 $\mathcal{H}_t \coloneqq (\mathcal{V}\times\mathcal{C})^t$ denote the set of histories of agents' 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 agent 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. Without loss of generality we can also restrict attention to deterministic mechanisms.

In contrast to the Markov decision process in Section 3, which does not differentiate between different agents of the same type, we now need to distinguish agents not only by their type but also by their arrival time, which, given the pairwise arrival process, coincides with their identity. 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 $i \in \{1,\dots,t\}$ 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 policy

[13]For example, $\overline{\tau}(0.95, 0.4) = 7.01$. Even as δ increases to 0.99, we have $\overline{\tau}(0.95, 0.4) = 39.45$, meaning that it will not be optimal to store more than 40 identical suboptimal pairs even with a discount factor so close to 1. In contrast, measured by T_τ, storing 40 suboptimal achieves 98.7 percent of the maximally possible increases in market thickness.

π to an allocation rule $\boldsymbol{Q}$ requires augmenting π with a queueing protocol, denoted by μ, that serves as a tie-breaking rule among agents of the same type.

Although the mechanism design problem is dynamic because the optimal policy varies with the state, the problem of incentivizing the agents to reveal their private information is essentially static as the agents' 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 agent and every history $\boldsymbol{h}_{t-1}$, regardless of the report of the other period t agent, assuming that all future agents report truthfully.[14] Formally, given a direct mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ let $q(\hat{\theta}, \vartheta, \boldsymbol{h}_{t-1})$ and $m(\hat{\theta}, \vartheta, \boldsymbol{h}_{t-1})$ denote the discounted probability of trade and expected discounted payment, respectively, for an agent that reports $\hat{\theta}$ at history $\boldsymbol{h}_{t-1}$ when the other period t agent reports ϑ.[15] For every history $\boldsymbol{h}_{t-1} \in \mathcal{H}_{t-1}$, $v \in \mathcal{V}$ and $c \in \mathcal{C}$, (P-IC) requires

$$\begin{aligned} v &= \arg\max_{\hat{\theta} \in \mathcal{V}} \left\{ v q(\hat{\theta}, c, \boldsymbol{h}_{t-1}) - m(\hat{\theta}, c, \boldsymbol{h}_{t-1}) \right\}, \\ c &= \arg\max_{\hat{\theta} \in \mathcal{C}} \left\{ m(\hat{\theta}, v, \boldsymbol{h}_{t-1}) - c q(\hat{\theta}, v, \boldsymbol{h}_{t-1}) \right\}, \end{aligned} \tag{P-IC}$$

while (P-IR) requires

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

As mentioned, we consider both the case where the designer maximizes expected discounted social surplus and the case where it maximizes expected discounted profit. In order to circumvent the indeterminacy that arises under expected discounted social surplus maximization we assume that the surplus-maximizing designer implements the efficient allocation with a mechanism that otherwise maximizes the designer's profits.[16] Under the mechanisms that are *optimal* given these objectives, 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,

[14] In the companion paper (Loertscher et al., 2020), we also consider weaker constraints. *Interim* constraints require that truthful reporting and participation is optimal for every period t agent and every history $\boldsymbol{h}_{t-1} \in \mathcal{H}_{t-1}$, assuming the other period t agent and all future agents report truthfully. *Bayesian* constraints require that truthful reporting and participation is optimal for every period t agent, assuming all other agents report truthfully.

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

[16] As is well-known, the payoff equivalence theorem does not hold for discrete type spaces. By restricting attention to the efficient mechanism that otherwise maximizes profit, we avoid this indeterminacy. This is also the appropriate assumption under which to analyze the possibility of efficient trade.

Elkind, 2007). These binding constraints yield

$$\begin{aligned} m(\underline{v}, c, \boldsymbol{h}_{t-1}) &= \underline{v} q(\underline{v}, c, \boldsymbol{h}_{t-1}), \quad m(\overline{c}, v, \boldsymbol{h}_{t-1}) = \overline{c} q(\overline{c}, v, \boldsymbol{h}_{t-1}) \\ m(\overline{v}, c, \boldsymbol{h}_{t-1}) &= \overline{v}(q(\overline{v}, c, \boldsymbol{h}_{t-1}) - q(\underline{v}, c, \boldsymbol{h}_{t-1})) + \underline{v} q(\underline{v}, c, \boldsymbol{h}_{t-1}), \\ m(\underline{c}, c, \boldsymbol{h}_{t-1}) &= \underline{c}(q(\underline{c}, v, \boldsymbol{h}_{t-1}) - q(\overline{c}, v, \boldsymbol{h}_{t-1})) + \overline{c} q(\overline{c}, v, \boldsymbol{h}_{t-1}). \end{aligned} \tag{5}$$

The incentive compatibility constraints for the inefficient types are satisfied and the allocation rule is *implementable* if and only if $q(\overline{v}, c, \boldsymbol{h}_{t-1}) \geq q(\underline{v}, c, \boldsymbol{h}_{t-1})$ and $q(\underline{c}, v, \boldsymbol{h}_{t-1}) \geq q(\overline{c}, v, \boldsymbol{h}_{t-1})$.[17] The virtual type functions are then given by

$$\Phi(\underline{v}) \coloneqq \underline{v} - \frac{w}{1-w}(\overline{v} - \underline{v}) \quad \text{and} \quad \Gamma(\overline{c}) \coloneqq \overline{c} + \frac{w}{1-w}(\overline{c} - \underline{c}), \tag{6}$$

with $\Phi(\overline{v}) = \overline{v}$ and $\Gamma(\underline{c}) = \underline{c}$.[18]

Next, for $i \leq t$, let $v^{B_i}(\boldsymbol{h}_t) \in \mathcal{V}$ and $c^{S_i}(\boldsymbol{h}_t) \in \mathcal{C}$ denote the types of buyer B_i and seller S_i given history $\boldsymbol{h}_t \in \mathcal{H}_t$, respectively. Expected discounted welfare under any direct, truthful mechanism that implements the allocation rule $\boldsymbol{Q}$ is then given by

$$W(\boldsymbol{Q}) := \sum_{t=1}^{\infty} \sum_{i=1}^{t} \sum_{\boldsymbol{h}_t \in \mathcal{H}_t} \delta^{t-1} \left(v^{B_i}(\boldsymbol{h}_t) Q_t^{B_i}(\boldsymbol{h}_t) - c^{S_i}(\boldsymbol{h}_t) Q_t^{S_i}(\boldsymbol{h}_t) \right) \mathbb{P}(\boldsymbol{H}_t = \boldsymbol{h}_t). \tag{7}$$

Following standard mechanism design arguments (see, for example, Loertscher et al., 2020) expected discounted profit under any direct mechanism that implements the allocation rule $\boldsymbol{Q}$ under the binding P-IC and P-IR constraints described above[19] is given by

$$R(\boldsymbol{Q}) := \sum_{t=1}^{\infty} \sum_{i=1}^{t} \sum_{\boldsymbol{h}_t \in \mathcal{H}_t} \delta^{t-1} \left(\Phi(v^{B_i}(\boldsymbol{h}_t)) Q_t^{B_i}(\boldsymbol{h}_t) - \Gamma(c^{S_i}(\boldsymbol{h}_t)) Q_t^{S_i}(\boldsymbol{h}_t) \right) \mathbb{P}(\boldsymbol{H}_t = \boldsymbol{h}_t). \tag{8}$$

The designer's problem is thus to determine allocation rules $\boldsymbol{Q}_0^*$ and $\boldsymbol{Q}_1^*$ that maximize (7) and (8) respectively subject to the feasibility[20], P-IC and P-IR constraints.

[17] 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.

[18] 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.

[19] In the companion paper (Loertscher et al., 2020), we show that the expected discounted profit of the designer is invariant under the weaker constraints outlined in Footnote 14. Thus, the designer has no incentive to conceal the history $\boldsymbol{h}_{t-1} \in \mathcal{H}_t$ from arriving agents.

[20] *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.

4.2 Optimal mechanisms

We now solve the designer's problem and derive the welfare- and profit-maximizing mechanisms. By construction, the optimal policy π_{τ^*} with $\Delta = \Delta_0$ derived in Section 3 maximizes expected discounted welfare subject to feasibility. We denote the optimal policy corresponding to welfare-maximization by π_0^* and the associated optimal threshold by τ_0^*.

Next, inspection of (7) and (8) shows the problems of welfare-maximization and profit-maximization are isomorphic insofar as the latter has the same structure as the former, with the virtual types replacing the true types. Thus, denoting by $\Delta_1 := \Phi(\overline{v}) - \Gamma(\overline{c}) = \Phi(\underline{v}) - \Gamma(\underline{c})$ the value of a suboptimal trade to the designer under profit maximization, it follows that the optimal policy π_{τ^*} with $\Delta = \Delta_1$ is optimal for the profit-maximizing designer. Note that among other things, this implies the analysis from Section 3 concerning market thickness T_τ also applies to a profit-maximizing designer. We denote the optimal policy corresponding to profit-maximization by π_1^* and the associated optimal threshold by τ_1^*.

Observe that

$$\Delta_1 = \Delta_0 - \frac{w}{1-w}(1 - \Delta_0) < \Delta_0. \tag{9}$$

Together with (8) this implies that expected revenue per period under periodic ex post efficiency (which induces trade in period t if and only if $v^{B_t} > c^{S_t}$) is

$$w^2 + 2w(1-w)\Delta_1 = w(2\Delta_0 - w). \tag{10}$$

This is negative if $w > 2\Delta_0$. Thus, under this parameter restriction, periodic ex post efficient trade is not possible without running a deficit. Moreover, we have $\Delta_1 \leq 0$ whenever $w \geq \Delta_0$ and in this case the optimal policy π_{τ^*} is such that $\tau^* = \infty$.[21]

The allocation rules that implement the optimal policies π_0^* and π_1^* are unique only up to the identities of agents that are cleared from the market when more than one agent of a given type is present. Thus, as stated in Section 4.1, mapping these policies to an optimal allocation rule requires that we also specify a queueing protocol μ in order to break ties. In implementing a threshold policy, the queueing protocol 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.[22] In expressing the designer's optimization problem as a Markov decision process with a simplified state space, we have thus shown that for a given policy the

[21]Intuitively, $\Delta_1 \leq 0$ implies the profit associated with executing a suboptimal trade is non-positive. Thus, the designer only executes efficient trades and is willing to store an unbounded number of efficient traders.

[22]For example, in period t a first-in-first-out (FIFO) 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 (LIFO) queueing protocol gives the highest priority to period t agents and clears any stored efficient types according to a reversed order of arrival.

designer's payoff does not vary with the treatment of individual agents. Combining this with the isomorphism between welfare- and profit-maximization we have the following lemma.[23]

Lemma 2. *Expected discounted welfare and profit under any optimal P-IC and P-IR mechanism that implements a policy π_τ are independent of the queueing protocol μ.*

Putting everything together, the welfare- and profit-maximizing policies can be implemented as follows. First, by Lemma 2, one can select any queueing protocol μ. Together with an optimal policy π_{τ^*} this gives rise to an optimal allocation rule $\boldsymbol{Q}^*$. Second, given $\boldsymbol{Q}^*$, one can compute the associated expected discounted allocation rule $\boldsymbol{q}^*$.[24] Provided the expected discounted allocation rule is implementable, the transfers under the optimal P-IC and P-IR mechanism can then be computed using (5) and Footnote 17. Formally, we have the following theorem.

Theorem 2. *Both the welfare-maximizing policy π_0^* and the profit-maximizing policy π_1^* can be implemented with P-IC and P-IR mechanisms.*

We conclude this section by discussing properties of these optimal mechanisms, with a particular focus on the profit-maximizing mechanism and its implications for indirect taxation and regulation of monopolist market-makers. Combining Proposition 3 with (9) implies the following corollary:

Corollary 1. *A profit-maximizing designer induces optimally more market thickness than a designer who maximizes expected discounted social surplus.*

Corollary 1 is reminiscent of Hotelling's (1931) finding that a monopolist extracts an exhaustible resource at a slower rate than a perfectly competitive industry.[25]

Indirect taxation For perfectly thick markets, as is well known, specific and ad valorem taxes are equivalent. We now briefly show that in optimally thick markets, ad valorem taxes

[23] We will shortly see that the invariance described in Lemma 2 does not hold if we restrict the flexibility with which the designer sets transfers and focus instead on posted-price mechanisms.

[24] Note that the proof of Theorem 2 contains an explicit calculation of an expected discounted allocation rule $\boldsymbol{q}$ given a threshold policy π_τ and a convenient queueing protocol μ.

[25] Inefficiently few matches also take place under profit maximization in the dynamic matching model of Fershtman and Pavan (2019). Corollary 1 does not necessarily extend to finite horizon models with richer type spaces. For example, Loertscher et al. (2017) consider a two-period version of Myerson and Satterthwaite (1983) where in every period a buyer-seller pair arrives, with a common discount factor applied to period two and with each agent drawing her type independently from a continuous distribution with compact support. Based on static mechanism design intuition, one might expect the designer to increase profit by restricting trade in each period. However, this leads to a decrease in the probability that period one agents trade in period two, which reduces the benefit of waiting in period one. Thus, in some cases it is optimal for the designer to increase period one trade to raise additional profit.

are superior to specific taxes, assuming that the designer maximizes expected discounted profits and that authorities can observe and, under an ad valorem tax, tax the designer's profit.[26] This is analogous to the standard assumption in oligopoly models of indirect taxation that firms' profits can be observed and taxed.[27] Under a specific tax $\sigma > 0$ per unit traded, the value of an efficient trade decreases from 1 to $1 - \sigma$ while the value of a suboptimal trade decreases from Δ_1 to $\Delta_1 - \sigma$. Given σ, the optimal policy of the designer is thus the same as for our original problem with Δ_1 replaced by $\Delta(\sigma) = (\Delta_1 - \sigma)/(1 - \sigma)$ and 1 replaced by $1 - \sigma$. Observe that $\Delta'(\sigma) < 0$ and $\Delta(0) = \Delta_1$. Corollary 1, with Δ_1 replaced by $\Delta(\sigma)$, thus implies that increasing σ will induce the designer to increase the threshold τ^*. Thus, a specific tax distorts the relative value of suboptimal trades, inducing the market maker to create an excessively thick market and further reducing the welfare gains of buyers and sellers. When $\sigma > \Delta_1$, the designer will become perfectly patient and never execute a suboptimal trade. In contrast, an ad valorem tax levied as a percentage on the designer's profit will not affect the relative value of a suboptimal trade. Thus, the policy will not change and an ad valorem tax can be levied without affecting social welfare gains. Consequently, ad valorem taxes are superior to specific taxes in markets whose thickness is endogenously determined by profit maximization.[28]

4.3 Posted-price mechanisms

The optimal direct mechanism asks agents to report types and makes payments and allocations that depend, in general, on the reports of the contemporaneously arriving agents. In what follows, we focus on one particular class of mechanisms: posted-price mechanisms. Posted-price mechanisms are simple and widely used in practice. It also allows us to introduce an efficient budget-balanced mechanism and to derive properties of equilibrium price distributions and dynamics.

Under a *posted-price mechanism*, the designer posts prices p^B for buyers and p^S for sellers at the start of each period. These prices depend, in general, on the state y but not on the reports of the agents who arrive in that period. Upon arrival, agents observe the state and the posted prices. Then they indicate the quantities $q^B \in \{0, 1/2, 1\}$ and $q^S \in \{0, 1/2, 1\}$ they are willing to trade at the current prices, where $q^i = 1/2$ indicates i is indifferent

[26] We focus on a profit-maximizing designer to simplify the analysis. Otherwise, we would have to derive the optimal policies and mechanisms anew and impose an assumption as to how much the designer cares for tax revenue relative to social surplus and her own profit.

[27] This analysis extends directly to uniform and fixed frequency market clearing, which are introduced and analyzed in Section 5 below.

[28] Observe that the distorting effects of specific taxes vanish as δ approaches 1 because in the limit suboptimal trades vanish.

between trading and not trading.

We say that agents *bid sincerely* if the buyer with value v bids $q^B(p^B) = 1$ if $v > p^B$, $q^B(p^B) = 1/2$ if $v = p^B$ and $q^B(p^B) = 0$ otherwise and the seller with cost c bids $q^S(p^S) = 1$ if $c < p^S$, $q^S(p^S) = 1/2$ if $c = p^S$ and $q^S(p^S) = 0$ otherwise. If all agents bid sincerely, and prices satisfies $p^B \in [\underline{v}, \overline{v}]$ and $p^S \in [\underline{c}, \overline{c}]$, the designer can infer the stored traders' types.

The designer chooses what trades to execute at the posted prices on the basis of the demand and supply schedules of all present agents and a policy π_τ. In the event of ties, a queueing protocol specifies how these are broken. Accordingly, a posted-price mechanism is characterized by the pricing rules p^B and p^S, policy π_τ and queueing protocol μ. In what follows, we denote by τ_L (τ_H) the state $y = \tau$ in which the stored agents are sellers (buyers).

Efficient posted-price mechanisms We begin with efficient posted-price mechanisms and first describe a *budget-balanced posted-price mechanism that implements the efficient policy*, provided $\tau_0^* \geq 1$. For any state $y \in \{0, \ldots, \tau_0^*\}$, this mechanism posts a uniform price for the buyers and sellers, that is, $p^B = p^S = p_0^{BB}$ as follows: $p_0^{BB}(y) = 1/2$ for $y < \tau_0^*$, $p_0^{BB}(\tau_L) = \Delta_0$ and $p_0^{BB}(\tau_H) = 1 - \Delta_0$.

Proposition 5. *Provided $\tau_0^* \geq 1$, the efficient policy π_0^* can be implemented using a budget-balanced P-IC and P-IR posted-price mechanism with the uniform pricing rule $p^B = p^S = p_0^{BB}$ and a last-in-first-out (LIFO) queueing protocol.*

Letting

$$\delta_0^* := \frac{\Delta_0}{w(1-w) + \Delta_0(1 - 2w(1-w))},$$

the following corollary provides a necessary and sufficient condition for $\tau_0^* \geq 1$ and a sufficient condition for the possibility of efficient trade based on the primitives of the model.

Corollary 2. *Since $\tau_0^* \geq 1$ is equivalent to $\delta \geq \delta_0^*$, the efficient policy can be implemented without running a deficit if $\delta \geq \delta_0^*$.*

Proposition 5 provides a reconciliation of sorts of the Coase Theorem, according to which the ultimate allocation is efficient if transaction costs are negligible, with the mechanism design literature and the insights of Vickrey (1961), Hurwicz (1972), Myerson and Satterthwaite (1983) and Gresik and Satterthwaite (1989) that, in static settings, private information can be an insurmountable transaction cost. In our dynamic setting, the impossibility of efficient trade—which, as noted, applies to the one-shot bilateral trade problem in our setting if $w > 2\Delta_0$—can be overcome as soon as it is efficient to store one suboptimal trade.

Proposition 5 also combines insights from Hagerty and Rogerson (1987) and McAfee (1992) into a new possibility result. These papers study static allocation problems with continuous distributions, where ex post efficient trade is not possible. The posted-price mechanism from the bilateral trade setting of Hagerty and Rogerson sacrifices an efficient trade if the buyer's value is above the seller's cost but the types are on the same side of the posted price. Building on this, McAfee considers a setting with multiple buyers and sellers and a posted-price mechanism that sacrifices one trade that would occur under efficiency—if the types of the marginal buyer and seller who would trade under efficiency are on the same side of the posted price. Like McAfee's mechanism, the posted-price mechanism in our dynamic setting makes prices a function of some agents' bids (namely, of those who have already arrived), and like both Hagerty and Rogerson's and McAfee's mechanisms, it does not execute trades when both the buyer's and the seller's types are on the same side of the posted price. In contrast to static settings, not executing such a trade is efficient, which is what gives us the possibility result.

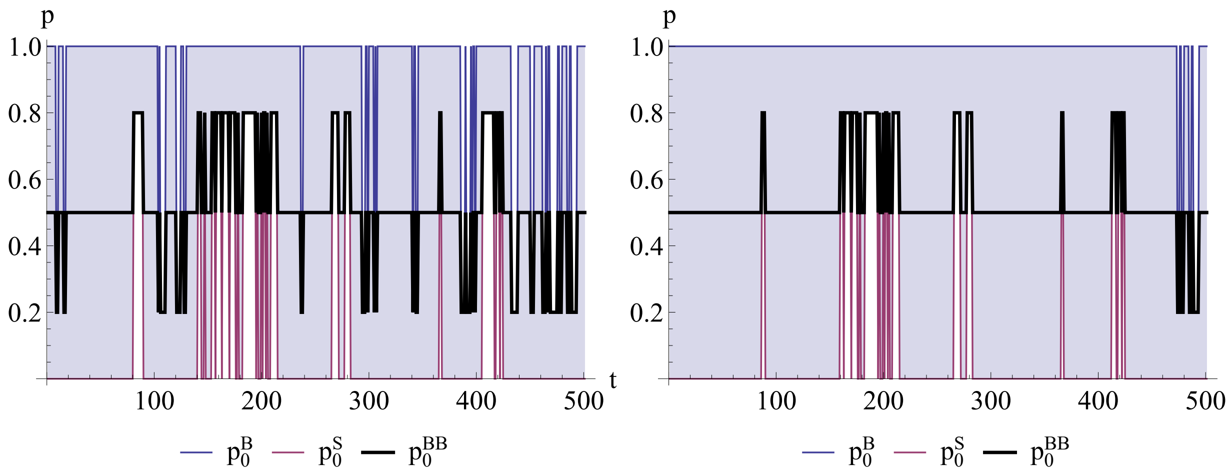
The implementation of the efficient policy (provided $\tau_0^* \geq 1$) via the budget-balanced posted-price mechanism also enables us to characterize the stationary price distribution and derive price dynamics. Let P_t denote the price posted in period t. By Lemma 1, under the stationary distribution we have $\mathbb{P}(P_t = \underline{v}) = \mathbb{P}(P_t = \overline{c}) = \frac{1}{2\tau_0^*+1}$ and $\mathbb{P}(P_t = 1/2) = \frac{2\tau_0^*-1}{2\tau_0^*+1}$. It follows that the variance of the posted prices under the stationary distribution is $\mathrm{Var}(P_t) = \frac{(1-2\Delta_0)^2}{2(2\tau_0^*+1)}$. Thus, we have the following corollary to Proposition 5:

Corollary 3. *The probability* $\mathbb{P}(P_t = 1/2)$ *increases in* δ *and* $w(1-w)$ *and decreases in* Δ_0*. The variance* $\mathrm{Var}(P_t)$ *decreases in* δ *and* $w(1-w)$*. Moreover,* $\lim_{\delta\to 1} \mathbb{P}(P_t = 1/2) = 1$ *and* $\lim_{\delta\to 1} \mathrm{Var}(P_t) = 0$.

These comparative statics results largely mirror the comparative statics of τ_0^*. The exception is the effect of the value of a suboptimal trade, Δ_0, on the variance, which cannot be signed in general.[29] 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 δ, provided only that $\delta \geq \delta_0^*$.

The budget-balanced posted-price mechanism also allows us to connect our measures of market thickness with the concept of an agent's likely price impact, denoted $p_{im}(\tau_0^*)$, which we define as the probability that the agent's arrival changes the posted price from the Walrasian price of $1/2$ to one of the extremal prices Δ_0 or $1-\Delta_0$ under the stationary

[29] An increase in Δ_0 both shifts probability mass away from $1/2$ and narrows the gap between the lowest price Δ_0 and the highest price $1-\Delta_0$ in the support.



(a) $\tau^* = 3$

(b) $\tau^* = 6$

Figure 4: For the same realizations of the arrival process with $w = 0.5$ and $\Delta = 0.2$ and the efficient allocation, we plot the budget-balanced posted price p_0^{BB} (black), as well as the buyers' price p_0^B (blue) and the sellers' price p_0^S (purple) under the profit-maximizing posted price mechanism. We also shade the bid-ask spread $p_0^B - p_0^S$ under the profit-maximizing posted-price mechanism. Panel (a) illustrates the market clearing policy $\tau_0^* = 3$ (which, from Figure 3, is optimal under the discount factor range $\delta \in [0.935, 0.962]$) and Panel (b) illustrates $\tau_0^* = 6$ (which is optimal under the discount factor range $\delta \in [0.984, 0.987]$).

distribution. That is,

$$p_{im}(\tau_0^*) := \mathbb{P}(P_t = \underline{v} | p_{t-1} = 1/2) = \mathbb{P}(P_t = \overline{c} | p_{t-1} = 1/2) = \frac{w(1-w)}{2\tau_0^* + 1} = \frac{1}{2}\mathbb{P}_S(\tau_0^*).$$

Thus, knowledge of $p_{im}(0)$ and $p_{im}(\tau_0^*)$ allows one to construct our measure of market thickness T^*, using the equivalence (4), to obtain $T^* = \rho_{\tau_0^*} = 1 - p_{im}(\tau_0^*)/p_{im}(0)$.

With small adjustments, we obtain a *profit-maximizing posted-price mechanism that implements the efficient policy* provided $\tau_0^* \geq 1$. The pricing rule $p_0^B(y)$ and $p_0^S(y)$ is such that $p_0^B(y) = 1$ and $p_0^S(y) = 0$ for $y < \tau_0^*$, $p_0^B(\tau_L) = \Delta_0$ and $p_0^S(\tau_L) = 0$ and $p_0^B(\tau_H) = 1$ and $p_0^S(\tau_H) = 1 - \Delta_0$.

Proposition 6. *The posted-price mechanism with pricing rule $p_0^B(y) = 1$ and $p_0^S(y) = 0$ for $y < \tau_0^*$, $p_0^B(\tau_L) = \Delta_0$ and $p_0^S(\tau_L) = 0$ and $p_0^B(\tau_H) = 1$ and $p_0^S(\tau_H) = 1 - \Delta_0$ and a LIFO queueing protocol provides a P-IC and P-IR implementation of the efficient market clearing policy π_0^*, provided $\tau_0^* \geq 1$. Moreover, this mechanism maximizes the market maker's profit within the class of efficient posted-price mechanisms.*

Figure 4 shows sample paths of posted-price mechanisms. Both panels display the paths of the budget-balanced price $p_0^{BB}(y)$ and the profit-maximizing bid and ask prices $p_0^S(y)$ and

$p_0^B(y)$ that implement the efficiency policy for the same realizations of values and costs. Panel (a) has a storage threshold of 3 and panel (b) a threshold of 6. As the storage threshold increases, the volatility of prices decreases, indicating that the comparative statics of the stationary distribution are borne out in the "comparative" dynamics.

The intuition behind Proposition 6 is as follows. If the designer increases any of the posted prices in any state, the mechanism will fail to implement the efficient allocation. So we have found the profit-maximizing posted-price implementation, provided the appropriate incentive constraints hold. Checking the P-IR constraints and the P-IC constraints for worst-off types is completely routine, while the P-IC constraints hold for the efficient types by virtue of the LIFO queueing protocol, which ensures that efficient types that do not bid sincerely eventually leave the market without trading.

The posted-price mechanism in Proposition 6 yields less profit for the designer than the optimal efficient mechanism due to the loss in profit associated with executing efficient trades in state $y = \tau^*$. Accounting for this loss, the profit associated with executing a suboptimal trade in any state $y \geq 1$ is given by $\Delta_1^{PP} := \Delta_0 - (1-\Delta_0)/(1-w)$, where $\Delta_1^{PP} < \min\{\Delta_1, 0\}$. Using Δ_1^{PP} we construct "posted-price virtual" type functions

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

with $\Phi^{PP}(\overline{v}) = \overline{v}$ and $\Gamma^{PP}(\underline{c}) = \underline{c}$, that account for how restriction to the class of posted-price mechanisms impacts the profit of designer and the informational rents of the agents in any state $y \geq 1$. Formally, we have the following proposition.

Proposition 7. *The profit-maximizing efficient posted-price mechanism yields strictly less profit than the optimal efficient mechanism, and can be derived from $R(\boldsymbol{Q})$ as defined in* (8) *with the virtual types replaced by their posted-price counterparts.*

Profit-maximizing posted-price mechanism We now specify the profit-maximizing posted-price mechanism. The condition

$$\frac{w(1-\Delta_0)}{1-\delta} - \frac{w(1-w)}{\sqrt{(1-\delta)(1-\delta+4\delta w(1-w))}} \geq 0 \tag{11}$$

plays an important role in characterizing the profit-maximizing posted-price mechanism.[30]

Proposition 8. *If $w \geq \Delta_0$ or if $\Delta_0 > w$ and* (11) *holds, then the profit-maximizing posted-price mechanism involves the pricing rule $p^B(y) = 1$ and $p^S(y) = 0$ in every state. Otherwise,*

[30]Note that the left-hand side of (11) increases monotonically in δ, that (11) is violated when $\delta = 0$ and that (11) always holds in the limit as $\delta \to 1$.

the profit-maximizing posted-price mechanism has the pricing rule $p^B(y) = 1$ and $p^S(y) = 1 - \Delta_0$ (or, equivalently, $p^B(y) = \Delta_0$ and $p^S(y) = 0$) in every state. These mechanisms satisfy P-IC and P-IR.

When $w \geq \Delta_0$ the profit-maximizing posted-price mechanism coincides with the profit-maximizing mechanism since $\Delta_1 \leq 0$, which implies that the profit-maximizing mechanism only executes efficient trades. For $w < \Delta_0$, the profit-maximizing mechanism generates more profit than the profit-maximizing posted-price mechanism. Unlike the posted-price mechanisms that implement the efficient policy, the profit-maximizing posted-price mechanism involves posting the same bid-ask spread in every state. Since $\Delta_1^{PP} < 0$ for all states $y \geq 1$, this mechanism must post a bid-ask spread of 1 for all states $y \geq 1$. Thus, the underlying policy must be to either clear a given type of suboptimal trade whenever one becomes available or clear an efficient trade whenever one becomes available.

4.4 Accounting for second-best mechanisms

The derivation of optimal market thickness in Section 3 was based on the assumption that if no trades are ever stored, the suboptimal trades are always executed when available. Of course, as noted above, periodic ex post efficiency is not possible without running a deficit if $w > 2\Delta_0$. We now briefly show that accounting for this incentive problem does not invalidate the conclusions derived above. If anything, the relative gains from increasing market thickness are even larger. In particular, for $w > 2\Delta_0$, the static second-best mechanism always executes efficient trades and executes suboptimal trades with probability $q^{SB} := \frac{w}{2(w-\Delta_0)}$, generating an expected surplus $S_0^{SB}(w)$. Denote by τ^{SB} the optimal threshold policy accounting for the second-best mechanism in the static problem and by $T_\tau^{SB} = \frac{S_\tau(w)-S_0^{SB}}{S_\infty(w)-S_0^{SB}(w)}$ the relative gains from increasing market thickness when accounting for the second-best mechanism in the static problem. Finally, denote by δ^{SB} the value of the discount factor such that $\tau^{SB} = 0$ if and only if $\delta \leq \delta^{SB}$, which is the analogue for second-best mechanisms to δ_0^* defined above Corollary 2.

Proposition 9. *Assume $w > 2\Delta_0$. For any $\tau \geq 1$, we have $T_\tau^{SB} > T_\tau$. Further, $\delta^{SB} < \delta_0^*$ holds, and we have $\tau_0^{SB} = 1$ for any $\delta \in (\delta^{SB}, \delta_0^*]$ and $\tau_0^{SB} = \tau_0^*$ for any $\delta > \delta_0^*$.*

Accounting for second-best mechanisms in the static problem increases the returns to increasing market thickness for two reasons. First, for a given threshold we have $\tau \geq 1$, $T_\tau^{SB} > T_\tau$ since the gains from static, second-best trade are smaller. Second, because the static second-best mechanism effectively reduces the value of a suboptimal trade, the designer is more inclined to store one suboptimal trade once one accounts for second-best mechanisms in the static bilateral trade problem. This is reflected in fact that $\delta^{SB} < \delta_0^*$.

We now show that an analogous result also applies to Proposition 2. Specifically, assuming that buyers and sellers draw their types independently from continuous distributions F and G that satisfy the assumption stated in Section 2, we show that there exists a P-IC and P-IR mechanism with two classes, called *two-class threshold mechanism*, that in expectation balances the budget in every period and implements the threshold policy π_τ.

Denoting by S_0^{SB} the expected per-period welfare in the one-shot second-best mechanism when the buyer and seller draw their types from the prior distributions, S_τ^{TC} the per-period expected welfare in the two-class threshold mechanism with threshold τ and $T_\tau^{TC} = \frac{S_\tau^{TC} - S_0^{SB}}{S_\infty - S_0^{SB}}$ the relative increase in market thickness in the two-class threshold mechanism, we have:

Proposition 10. *The two-class threshold mechanism satisfies P-IC and P-IR, balances the budget in expectation in every period and implements any threshold policy π_τ. The probability that trade occurs at the Walrasian price p goes to 1 as τ goes to infinity. Moreover, for any $\tau \geq 1$, we have $T_\tau^{TC} > T_\tau$.*

Proposition 10 mirrors many of the features of the posted-price mechanisms for the binary type setup. The two-class threshold mechanism is almost a posted-price mechanism. Only in the instance when an efficient trader of the type stored arrives in state $y = \tau$ will one need to "take a closer look" and run a second-best mechanism, which is different from posted prices. Because these instances become rarer as τ increases, it is also the case that the variance of the price distribution will decrease as τ increases. Thus, the comparative statics of the price distribution in Corollary 3 are not so much driven by the binary type assumption as by the fact that a threshold mechanism is used. Moreover, even though a threshold policy with $\tau = 1$ does not eliminate all incentive problems with continuous types, storing traders alleviates the incentives just like it does with binary types. As shown in the proof, the second-best mechanism with a stored trader outperforms the second-best mechanism in the one-shot bilateral trade problem, and as a consequence, market thickness increases even faster when accounting for incentives, that is, $T_\tau^{TC} > T_\tau$.

5 Approximately optimal market clearing

The optimal policy π_{τ^*} derived in Section 3 discriminates among different traders present at the same time. For example, if an efficient pair arrives, an efficient trade is always immediately executed, regardless of the state, which implies that not all trades are cleared at once. The optimal policy π_{τ^*} thus involves *discriminatory* market clearing. In reality, market clearing often takes a cruder form.[31] Under what we call *uniform market clearing*,

[31]For example, in foreign exchange spot markets such as Thomson Reuters, ParFX, and EBS, clearing is uniform in that all compatible orders are cleared at once. However, the time intervals that elapse between

the entire market is cleared at the time of clearing, and we speak of *fixed frequency market clearing* if, in addition to market clearing being uniform, the market is cleared at fixed intervals. We now analyze social surplus under these alternative market clearing regimes and show, among other things, that as $\delta \to 1$, the specific form of dynamic market clearing does not matter from a social surplus perspective. We also show that, for $\delta \to 1$, expected discounted profit-maximization generates more welfare than static ex post efficient trade. We defer the derivation and analysis of the class of optimal mechanisms under uniform and fixed frequency market clearing to Appendix B and C.

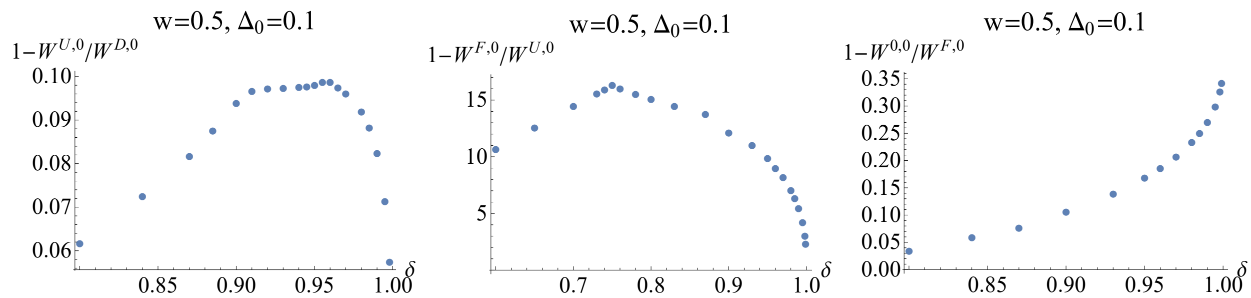
For $\alpha \in \{0, 1\}$, we let W_α^D denote expected discounted social surplus (starting from an empty market at $t = 0$) under optimal *discriminatory* market clearing. That is, W_0^D and W_1^D respectively denote expected discounted social surplus under the policies $\pi_{\tau_0^*}$ and $\pi_{\tau_1^*}$. Similarly, W_α^U and W_α^F respectively denote expected discounted social surplus for optimal uniform and fixed frequency market clearing. Expected discounted social surplus under periodic ex post efficient trade is denoted W_0^0 and given by $W_0^0 = S_0(w)/(1-\delta)$.[32] Finally, expected discounted profit W_1^0 under periodic profit-maximizing bilateral trade is $W_1^0 = (w^2 + 2w(1-w)\max\{\Delta_1, 0\})/(1-\delta)$. Discriminatory market clearing is the least restricted—and hence most *sophisticated*—market clearing policy among those considered here and periodic ex post efficient trade the most restricted—and hence least sophisticated. For $\alpha \in \{0, 1\}$, the following are then natural measures of the relative gains from additional sophistication: $G_\alpha^{D,U}(\delta) := 1 - W_\alpha^U(\delta)/W_\alpha^D(\delta)$, $G_\alpha^{U,F}(\delta) := 1 - W_\alpha^F(\delta)/W_\alpha^U(\delta)$ and $G_\alpha^{F,0}(\delta) := 1 - W_\alpha^0(\delta)/W_\alpha^F(\delta)$. Note that the definition of $W_0^0(\delta)$ does not account for incentives, so if anything it overestimates welfare under periodic ex post efficient trade. In contrast, as shown in Theorems B1 and C2 in the Online Appendices, both uniform and fixed frequency market clearing permit P-IC and P-IR implementation. Hence, $G_0^{F,0}(\delta)$ provides a lower bound for the gains from fixed frequency market clearing relative to static ex post efficiency.

Theorem 3. $\lim_{\delta\to 1} G_\alpha^{D,U}(\delta) = \lim_{\delta\to 1} G_\alpha^{U,F}(\delta) = 0 < (1-w)(1-2\Delta_\alpha) = \lim_{\delta\to 1} G_\alpha^{F,0}(\delta)$.

According to Theorem 3, the relative gains from additional sophistication vanish while the relative gains from any degree of sophistication relative to instantaneous trade remain

clearings depend on the orders received. In other trading venues, such as the Global Dairy Trade in New Zealand, market clearing is both uniform and occurs at a fixed frequency.

[32]Under continuous-time double auction mechanisms feasible trades are also executed immediately. Thus, the outcome of periodic ex efficient trade is the same as the outcome that would result under a continuous-time double auction with truthful bidding. However, continuous-time double auctions are not incentive compatible as the bid of a given trader affects both the probability of trade and, in the event that trade occurs, the market price. Under strategic bidding one would expect efficient types to bid shade in order to avoid trading with an inefficient type so that they receive a higher expected payoff. Although the equilibrium behavior of a continuous-time double auction is difficult to characterize (see for example, Satterthwaite and Williams, 2002), the outcome under the first-best mechanism provides an efficiency benchmark for evaluating the outcome of a continuous-time double auction.



(a) Discrim. vs. uniform (b) Uniform vs. fixed freq. (c) Fixed freq. vs. decent.

Figure 5: The relative gains from additional sophistication.

strictly positive as δ approaches 1. This phenomenon is illustrated numerically in Figure 5. The outcome under periodic ex post efficient trade coincides with the efficient outcome when δ is so close to 0 that storing is not efficient.

Next, we compare social surplus under the *profit-maximizing mechanisms* to those under periodic ex post efficient trade. Note that a simple argument based on less constrained optimization cannot be used: As one goes from periodic ex post efficient trade to profit-maximizing discriminatory market clearing, one not only eliminates constraints but also alters the objective that is maximized.

Theorem 4. *For all $k \in \{D, U, F\}$ there exists $\delta_k \in [0, 1)$ such that $W_1^k(\delta_k) = W_0^0(\delta_k)$ and $W_1^k(\delta_k) > W_0^0(\delta_k)$ for all $\delta > \delta_k$.*

By Theorem 4, for sufficiently large discount factors, profit-maximization generates more expected discounted social than periodic ex post efficient trade. This is so because efficient types trade with relatively high probability under the profit-maximizing mechanism, which is efficient for a sufficiently large discount factor.[33]

6 Related literature

With its focus on optimal market thickness, our paper relates to the literature on double auctions, such as Gresik and Satterthwaite (1989), Satterthwaite and Williams (1989, 2002), McAfee (1992), Rustichini et al. (1994), and Cripps and Swinkels (2006), whose origins go back to the impossibility theorem of Myerson and Satterthwaite (1983). Apart from

[33]The gist of this result does, again, not depend on the distributions being binary. For example, if buyers and sellers draw their types independently from the uniform distribution on $[0, 1]$, expected social surplus under ex post efficiency in the bilateral trade setting is $1/6$. Under profit-maximization in the large market limit, per trader-pair social surplus is $3/16$. This also shows that our result is not driven by the fact that in our setting monopoly in the large market limit is not inefficient. The superior sorting in the large market limit outweighs the efficiency loss due to rent extraction by the monopoly in the uniform-uniform case.

the dynamics inherent to its setting, our paper differs from the focus and approach of this literature in two ways. The driving force for increasing market thickness in our approach is to improve sorting, and we derive the optimal degree of market thickness in the binary type setting.

As we study a dynamic version with binary types of the classical bilateral trade problem of Myerson and Satterthwaite (1983), our paper also relates to the literature on the (im)possibility of efficient trade with private information, initiated by Vickrey (1961), Hurwicz (1972) and Myerson and Satterthwaite (1983), and the Coase Theorem (Coase, 1960).[34] In particular, we show that with binary types efficient trade is possible as soon as the efficient policy in the dynamic setting differs from periodic ex post efficiency and can be implemented with a budget-balanced price posting mechanism. This mechanism combines insights from the static mechanisms of Hagerty and Rogerson (1987) and McAfee (1992), in which a posted price sometimes sacrifices an efficient trade to avoid a deficit, into the dynamic setting, in which trades are not sacrificed but merely stored, which is efficient.[35] In turn, the posted-price implementation also allows us to relate the equilibrium price distribution with the literature on the microfoundations of Walrasian equilibrium such as Satterthwaite and Shneyerov (2007) and Lauermann (2013), which establishes convergence to efficiency and to the large market limit Walrasian price as frictions vanish. In our setting, the equilibrium price distribution also converges to a mass point at this price as the discount factor goes to one, but in contrast to that literature, the equilibrium is efficient for any discount factor such that storing is optimal.

The paper also relates to the vast literature on mechanism and market design, which is one of the major achievements of economics over the past half century or so (see Börgers, 2015, for an introduction and overview of this literature). With a few notable exceptions that we discuss next, the main bulk of this body of research has confined attention to static, one-shot allocation problems in which all the relevant agents and their information, albeit privately held, are present at the outset. In particular, we apply the techniques developed by Myerson (1981) to a dynamic setting with discrete types and two-sided private information static versions of which have previously been studied by, among others, Myerson and Satterthwaite (1983). We use the notions of periodic ex post incentive compatibility and individual rationality that were introduced by Bergemann and Välimäki (2010). Much

[34] Milgrom (2017) provides persuasive arguments that complexity may be an additional source of transaction costs impeding efficient (re)allocation.

[35] There is an important difference between our possibility result and the recent literature on the (im)possibility of efficient bilateral trade in repeated settings that started with Athey and Miller (2007), where the efficient policy does not vary with the discount factor and more patience merely slackens the individual rationality constraints. In contrast, in our setting, it is precisely the change in the efficient policy resulting from increases in the discount factor that renders efficient trade without a deficit possible.

of the recent literature on dynamic mechanism design, including Athey and Miller (2007), Bergemann and Välimäki (2010), Athey and Segal (2013), Pavan et al. (2014) and Skrzypacz and Toikka (2015), studies settings in which a static population of agents receives private information over time. In contrast, our paper considers a dynamic population of agents with persistent types. In such setups, the current allocation decision determines the set of feasible allocations in future periods and the designer faces the optimal timing problem of deciding when to run a static mechanism. Recent contributions to this strand of literature include Parkes and Singh (2003), Gershkov and Moldovanu (2010) and Board and Skrzypacz (2016). However, none of the aforementioned papers explicitly address the optimal timing problem, consider varying degrees of sophistication of the mechanisms or compare welfare and profit maximization.[36]

Our paper is methodologically related to the recent literature on dynamic matching where monetary transfers cannot be used to incentivize agents. For example, building on Ünver (2010) and Anderson et al. (2017), Akbarpour et al. (2020) study efficiency in a dynamic matching model in which exchange possibilities have a network structure. Our paper draws inspiration from the work of Baccara et al. (forthcoming), which, motivated by the problem of matching children and parents in an adoption "market," considers a dynamic, two-sided matching problem. Their efficient algorithm is similar to the optimal market clearing policy in our paper. There are, however, crucial differences. We adhere to the standard assumptions in the dynamic mechanism design literature of geometric discounting and quasilinear payoffs.[37] This permits the use of monetary transfers to incentivize agents and allows us to study a broad range of questions that cannot be addressed in a setting without transfers. At the heart of optimal market thickness is the tradeoff between the benefits of better trading opportunities accruing from accumulating traders and the cost of delay. This tradeoff has been studied in different but related environments by, among other others, Vayanos (1999), Rostek and Weretka (2015) and Du and Zhu (2017). In particular, Du and Zhu (2017) derive the optimal trading frequency when new information arrives over time. Our paper complements their analysis by deriving the optimal trading mechanism.

Our companion paper (Loertscher et al., 2020) shows that the Markov decision process methodology is flexible and can be used to analyze a variety of extensions. It constructs the Markov decision process for finite discrete type spaces and proves appropriately generalized versions of Proposition 3, Theorems 2, 3 and 4 and Corollaries 1 and 2. It also considers several generalizations of the arrival process, including unpaired arrivals, continuous-time

[36]Recent papers that address the optimal timing problem in one-sided settings include Pai and Vohra (2013) and Mierendorff (2016) and references therein.

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

arrivals, group arrivals and multi-unit traders.

7 Conclusions

Optimally thick markets efficiently balance the gains from improved sorting against the cost of delay. We show that, measured by the maximum numbers of traders present at any point in time, optimally thick markets are surprisingly thin, with the maximum number of traders often being in the single digits. At the same time, when evaluated by how much of the maximum achievable per trader-pair welfare increase is reaped, even seemingly thin markets are surprisingly thick: storing one suboptimal trade rather than none achieves two-thirds of the maximum increase that is obtained by going from the perfectly thin bilateral trade setup to the perfectly thick large market limit. We also show that this number does not depend on distributional assumptions. The paper thus provides an explanation for why typical real-world markets have few traders and so differ, observationally, from the textbook model of competitive markets and for why this model still provides a reasonably good approximation for the behaviour of these markets. Our results pertaining to continuous distributions also show that first-order gains can be achieved by simple, coarse mechanisms.

There are many avenues for future research. That larger markets improve sorting seems worth exploring and understanding more generally and systematically. For example, it would be useful to know for, say, continuous distributions under which conditions the benefits of increasing market thickness are largest. Alternatively, one could augment the model we study to allow for strategic arrival and the possibility of arbitrage. Applying the methodology and insights from this paper to other environments, where, for example, agents' efficient trading positions are endogenous, also seems promising. Furthermore, a point of much interest in the finance literature has been that large institutional traders optimally reduce their price impact by breaking up their trades when a fixed, suboptimal mechanism is used to clear the market. In light of this, an interesting extension of our model would be to accommodate large traders and to analyze the impact of traders' size on the optimal market clearing mechanism. Last but not least, as in dynamic settings like ours the efficient policy is not a prior-free concept, the development of incentive compatible, prior-free dynamic mechanisms whose allocation rules converge to an efficient one would be valuable and interesting.

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 (forthcoming): "Optimal Dynamic Matching," *Theoretical Economics*.

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.

Borovkov, K. (2014): *Elements of Stochastic Modelling*, World Scientific Publishing Company, 2 ed.

Coase, R. H. (1960): "The Problem of Social Cost," *Journal of Law and Economics*, 3, 1–44.

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.

Hagerty, K. M. and W. P. Rogerson (1987): "Robust Trading Mechanisms," *Journal of Economic Theory*, 42, 94–107.

Hendricks, K. and A. Sorensen (2018): "Dynamics and Efficiency in Decentralized Online Auction Markets," Working paper.

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.

Latouche, G. and V. Ramaswami (1999): *Introduction to Matrix Analytic Methods in Stochastic Modeling*, Philadelphia, Pennsylvania: Society for Industrial and Applied

Mathematics.
Lauermann, S. (2013): "Dynamic Matching and Bargaining Games: A General Approach," *American Economic Review*, 103, 663–689.
Loertscher, S., E. V. Muir, and P. G. Taylor (2017): "Dynamic Market Making," Working paper.
——— (2020): "The Dynamics of Optimal Market Clearing: A General Approach," Working paper.
McAfee, P. (2002): "Coarse Matching," *Econometrica*, 70, 2025–34.
McAfee, R. P. (1992): "A Dominant Strategy Double Auction," *Journal of Economic Theory*, 56, 434–450.
Mierendorff, K. (2016): "Optimal Dynamic Mechanism Design with Deadlines," *Journal of Economic Theory*, 161, 190–222.
Milgrom, P. (2017): *Discovering Prices*, New York: Columbia University Press.
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.
Pai, M. M. and R. Vohra (2013): "Optimal Dynamic Auctions and Simple Index Rules," *Mathematics of Operations Research*, 38, 682–697.
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.

Appendix

A Proofs

A.1 Proof of Theorem 1

Proof. By Theorem 6.2.10 of Puterman (1994) there exists a deterministic stationary optimal policy of the Markov decision process. Furthermore, the optimal policy must immediately clear any efficient pairs that are available in each state $\boldsymbol{x}$. These pairs yield the maximal gains from trade and hence there is no benefit associated with storing such pairs. Next, note that sample paths of the Markov decision process are such that if x_S efficient traders are stored in a given period then $x_S - 1$ efficient traders must have been stored in some previous period. Thus, if x_S efficient traders are stored under the stationary optimal policy, it must also be optimal to retain $x_S - 1$ efficient trades. Finally, an unbounded number of efficient traders cannot be stored under the optimal policy since 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. Thus, the expected discounted benefit of storing an additional suboptimal pair converges to zero, while the benefit of immediately clearing a suboptimal pair is always $\Delta > 0$. Putting all this together, there exists a maximum number τ^* of suboptimal trades which can be optimally stored and the optimal policy π^* is a threshold policy. □

A.2 Proof of Lemma 1

Proof. The transition matrix $\boldsymbol{P}$ of the Markov chain $\{Y_t\}_{t\in\mathbb{N}}$ is given by

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

where $\lambda = w(1-w)$. So we are dealing with a simple birth-and-death process (see, for example, pages 184–189 in Borovkov, 2014). For such processes, it is well-known that the stationary distribution of the Markov chain $\{Y_t\}_{t\in\mathbb{N}}$ is given by $\boldsymbol{\kappa}$. □

A.3 Proof of Proposition 2

Proof. In the static bilateral trade setup, 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 $\tilde{\Delta} = (\Delta_H + \Delta_L)/2$ is the expected surplus of a suboptimal trade.

Consider now the dynamic mechanism that separates buyers and sellers into two classes: efficient traders, meaning $v \geq p$ and $c \leq p$, and inefficient traders, meaning $v < p$ and $c > p$. The mechanism 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 τ. Note that this mechanism does not distinguish the nature of these traders (i.e. whether they arrived as part of a suboptimal pair that creates surplus Δ_H or Δ_L). However, since each type of suboptimal pair arrives with an equal probability, a suboptimal trade creates an expected surplus of $\tilde{\Delta}$. Setting $\Delta_0 = \tilde{\Delta}/(\overline{v}-\underline{c})$, our analysis from the binary type setting applies and we have, in particular, that the relative increase in market thickness from storing τ efficient traders is T_τ. □

A.4 Proof of Proposition 3

Proof. Since $T^* = \frac{2\tau^*}{2\tau^*+1}$ is increasing in τ^*, it suffices to prove each of the comparative statics results for τ^*. That is, we show that τ^* is increasing in δ, Δ and $w(1-w)$. To accomplish this we derive and solve the Bellman equations that characterize τ^*. Taking any $y \in \{0, 1, \ldots, \tau\}$ we let $V_\tau^D(y)$ denote the expected present value of having y identical efficient traders stored at the end of a period under the threshold policy π_τ.

First, we characterize τ^* for $\tau^* \geq 2$. Let $\lambda = w(1-w)$ and take any threshold policy π_τ with $\tau \geq 1$. For $y \in \{1, \ldots, \tau-1\}$, this policy is characterized by the Bellman equation

$$V_\tau^D(y) = \delta \left[w^2(1 + V_\tau^D(y)) + \lambda(1 + V_\tau^D(y-1) + V_\tau^D(y+1)) + (1-w)^2 V_\tau^D(y) \right], \qquad (12)$$

with boundary conditions

$$V_\tau^D(0) = \delta\left[w^2(1+V_\tau^D(0)) + 2\lambda V_\tau^D(1) + (1-w)^2 V_\tau^D(0)\right],$$
$$V_\tau^D(\tau) = \delta\left[w^2(1+V_\tau^D(\tau)) + \lambda(1+V_\tau^D(\tau-1)+\Delta_\alpha+V_\tau^D(\tau)) + (1-w)^2 V_\tau^D(\tau)\right].$$

Since Δ is the instances reward from clearing a single suboptimal pair, the optimal threshold τ^* is characterized by the stopping condition

$$V_{\tau^*}^D(\tau^*) - V_{\tau^*}^D(\tau^*-1) > \Delta \quad \text{and} \quad V_{\tau^*+1}^D(\tau^*+1) - V_{\tau^*+1}^D(\tau^*) \leq \Delta. \tag{13}$$

Next, set $\tilde{V}_\tau^D(y) = V_\tau^D(y) - V_\tau^D(y-1)$. Then for $\tau \geq 2$ and $y \in \{2,\ldots,\tau-1\}$, $\tilde{V}_\tau^D(y)$ satisfies

$$\tilde{V}_\tau^D(y) = \delta\left[(1-2\lambda)\tilde{V}_\tau^D(y) + \lambda\tilde{V}_\tau^D(y-1) + \lambda\tilde{V}_\tau^D(y+1))\right],$$

with boundary conditions

$$\tilde{V}_\tau^D(1) = \delta\left[\lambda + (1-3\lambda)\tilde{V}_\tau^D(1) + \lambda\tilde{V}_\tau^D(2)\right],$$
$$\tilde{V}_\tau^D(\tau) = \delta\left[\lambda\Delta + (1-2\lambda)\tilde{V}_\tau^D(\tau) + \lambda\tilde{V}_\tau^D(\tau-1)\right].$$

Solving this last recursion in MATHEMATICA without the boundary condition at $y=\tau$ yields

$$\tilde{V}_\tau^D(y) = \left(\frac{z_-}{2}\right)^n k_0 + \left(\frac{z_+}{2}\right)^n \left(\frac{2-(a+2)k_0}{b+2}\right),$$

where

$$z_\pm = 2 \pm \frac{\sqrt{(1-\delta)(1-\delta+4\delta w(1-w))} \pm (1-\delta)}{\delta w(1-w)} \tag{14}$$

and the constant k_0 is pinned down by the boundary condition at $y=\tau$. Elementary calculations (which we omit here for the sake of brevity) show that $z_+ > 2$ and $z_+ \in (0,2)$. Imposing the boundary condition at $y=\tau$ in MATHEMATICA 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

$$\tilde{V}_\tau^D(\tau) = \frac{(z_+ - 2)\left(\frac{z_+ z_-}{2}\right)^\tau + \Delta z_+ \left(2z_+^{\tau-1} - z_-^\tau\right)}{z_+^{\tau+1} - 2z_-^\tau}$$
$$\Rightarrow \tilde{V}_\tau^D(\tau) + \Delta = \frac{(z_+ - 2)\left[\left(\frac{z_+ z_-}{2}\right)^\tau - \Delta\left(z_+^\tau + z_-^\tau\right)\right]}{z_+^{\tau+1} - 2z_-^\tau}.$$

Thus, by (13), $\tau^* \geq \tau$ for $\tau \geq 2$ if and only if

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

Second, We characterize when $\tau^* \geq 1$. In particular, letting $\lambda = w(1-w)$, $V_1(0)$ and $V_1(1)$ satisfy the Bellman equation

$$V_1(0) = \delta\left[w^2(1 + V_1(0)) + 2\lambda V_1(1) + (1-w)^2 V_1(0)\right],$$
$$V_1(1) = \delta\left[w^2(1 + V_1(1)) + \lambda(1 + V_1(0) + \Delta + V_1(1)) + (1-w)^2 V_1(1)\right].$$

Solving this recursion in Mathematica yields

$$V_1(0) = \frac{\delta\left(2\delta\Delta\lambda^2 + w^2(\delta(\lambda-1)+1) + 2\delta\lambda w\right)}{(1-\delta)(\delta(3\lambda-1)+1)}, \tag{15}$$
$$V_1(1) = \frac{\delta\left(\Delta\lambda(\delta(2\lambda-1)+1) + \delta\lambda w^2 + w(\delta(2\lambda-1)+1)\right)}{(1-\delta)(\delta(3\lambda-1)+1)}.$$

Since we also have $V_0(0) = \delta(w^2 + 2\lambda\Delta)/(1-\delta)$, $\tau^* \geq 1$ provided

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

Now, since $\delta \in [0,1)$ and $\lambda > 0$ we have $1 - \delta > 0$ and $-1 + \delta - 3\delta\lambda < 0$. Hence, this last inequality is equivalent to $2\lambda(\Delta + \delta(\Delta(2\lambda-1)+\lambda)) \leq 0$. Rearranging, we obtain that $\tau^* \geq 1$ if and only if

$$\delta - \frac{\Delta}{\lambda + \Delta(1-2\lambda)} \geq 0. \tag{16}$$

We are now in a position to prove the desired comparative statics results. First, since every term in the recursion that characterizes $\tilde{V}_\tau^D(y)$ for $\tau \geq 2$ is increasing in δ, it follows from (13) that τ^* is increasing in δ for $\tau^* \geq 2$. Furthermore, since the right-hand side of (16) is increasing in δ, putting everything together we have that τ^* is increasing in δ.

Second, for $\tau \geq 2$ we have

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

where the inequality follows from $z_+ > 2$ and $z_- \in (0,2)$. It then follows from (13) that τ^* is decreasing in Δ for $\tau^* \geq 2$. Furthermore, since

$$\frac{d}{d\,\Delta}\left(\delta - \frac{\Delta}{\lambda + \Delta(1-2\lambda)}\right) = \frac{\lambda}{(\lambda+\Delta(1-2\lambda))^2} < 0$$

we have that the right-hand side of (16) is decreasing in Δ. Putting everything together we have that τ^* is decreasing in Δ.

Finally, using Mathematica for $\tau \geq 2$ we have

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

where the inequality follows from $z_+ > 2$ and $z_- \in (0,2)$. It then follows from (13) that τ^* is increasing in λ for $\tau^* \geq 2$. Furthermore, since

$$\frac{d}{d\,\lambda}\left(\delta - \frac{\Delta}{\lambda + \Delta(1-2\lambda)}\right) = \frac{\Delta(1-2\Delta)}{(\lambda+\Delta(1-2\lambda))^2} > 0$$

we have that the right-hand side of (16) is increasing in λ. Putting everything together we have that τ^* is increasing in λ. Thus, τ^* is increasing in $\lambda = w(1-w)$ as required and, given values of δ and Δ, τ^* is maximized when $w = 1/2$. □

The proof of this proposition has the following useful corollary.

Corollary A1. $\tau^* \geq 1$ *if and only if* $\delta \geq \Delta/(w(1-w) + \Delta(1-2w(1-w)))$ *and, for* $\tau \geq 2$, $\tau^* \geq \tau$ *if and only if* $\left(\frac{z_+ z_-}{2}\right)^\tau \geq \Delta\left(z_+^\tau + z_-^\tau\right)$, *where* $z_\pm$ *is defined in* (14).

A.5 Proof of Proposition 4

Proof. Consider the decision to store the jth efficient trader. The instantaneous reward from clearing this trader immediately is Δ, while an upper bound on the payoff associated with storing this trader is $\delta^j(1-\Delta)$.[38] For storing the jth trader to be profitable, we must have $\delta^j(1-\Delta) > \Delta$. Hence, τ^* must be such that $\tau^* < \log\left(\frac{\Delta}{1-\Delta}\right)/\log(\delta)$. In the limit as $\delta \to 1$ (i.e. taking a Laurent series expansion of $1/\log(\delta)$ about $\delta = 1$), we have an upper bound on τ^* of

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

□

A.6 Proof of Lemma 2

Proof. Given any policy π, we can compute expected discounted social surplus under π using the reward function $r(x_E, x_S) = x_E + \Delta_0 x_S$. Thus, expected discounted social welfare is invariant to the queueing protocol μ. Furthermore, assuming that the policy π is implemented using an optimal mechanism, we can computed expected discounted profit under π by using the reward function $r(x_E, x_S) = x_E + \Delta_1 x_S$. It immediately follows that the expected discounted profit of the designer is also invariant to the queueing protocol μ. □

A.7 Proof of Theorem 2

Proof. It suffices to show that a generic threshold policy π_τ can be implemented using an optimal P-IC and P-IR mechanism. By Lemma 2 we can augment π_τ with any queueing protocol μ. Thus, without loss of generality we can assume that when multiple identical efficient agents are present in period t, the agents that arrived in period t have absolute priority and 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 and a stored efficient type trades in a given period if and only if they are first in the queue and an appropriate suboptimal pair arrives. Thus, the waiting time $Z(n)$ for the nth agent in the queue follows a negative binomial distribution that counts the number of periods until n appropriate suboptimal pairs have arrived. Setting

[38] Note that this is an upper bound because the value associated with storing the jth 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 designer. However, if the jth efficient trader was not stored then the efficient trader from the jth appropriate suboptimal pair would instead be stored and subsequently generate an expected discounted payoff of at least Δ (otherwise it would be optimal to store even one efficient trader, let alone j of them).

$\lambda = w(1-w)$, for $i \geq n$, the probability mass function of $Z(n)$ is

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

Therefore, using MATHEMATICA we can compute the discounted probability of trade for the nth agent in the queue. We have

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

The policy π_τ and chosen queueing protocol uniquely pin down the allocation rule $\boldsymbol{Q}$ and we can now compute the expected discounted allocation rule $\boldsymbol{q}$. Recall that $q(\hat{\theta}, \vartheta, \boldsymbol{h}_{t-1})$ denotes the discounted probability of trade for an agent that reports $\hat{\theta}$ at history $\boldsymbol{h}_{t-1}$ when the other period t agent reports ϑ. Note that given the policy π_τ and chosen queueing protocol we can simplify the state space. In particular, rather than using the complete period $t-1$ history $\boldsymbol{h}_{t-1}$, sufficient statistics for the state are $y_H, y_L \in \{0, \ldots, \tau\}$, where y_H denotes the number of stored efficient buyers and y_L denotes the number of stored efficient sellers.

We first compute $\boldsymbol{q}$ for arriving buyers. Specifically, for $y_H = \tau$ and any $c \in \mathcal{C}$ we have

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

for $y_H \in \{0, \ldots, \tau - 1\}$ (which also subsumes the case $y_L = 0$), we have

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

for $y_L \in \{1, \ldots, \tau - 1\}$ and any $c \in \mathcal{C}$ we have

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

and finally for $y_L = \tau$ we have

$$q(\overline{v}, \underline{c}, y_L) = 1 = q(\underline{v}, \underline{c}, y_L) = 1 \quad \text{and} \quad q(\overline{v}, \overline{c}, y_H) = 1 > q(\underline{v}, \overline{c}, y_H) = 0.$$

Since $q(\overline{v}, c, y_H) \geq q(\underline{v}, c, y_H)$ for all $c \in \mathcal{C}$ and $y_H \in \{0, \ldots, \tau\}$ and $q(\overline{v}, c, y_L) \geq q(\underline{v}, c, y_L)$ for all $c \in \mathcal{C}$ and $y_L \in \{1, \ldots, \tau\}$, the allocation rule is implementable for the buyers using a P-IC and P-IR mechanism. Analogous calculations and an analogous argument apply to the sellers.

An optimal P-IC and P-IR mechanism $\langle \boldsymbol{Q}, \boldsymbol{M} \rangle$ that implements π_τ is then completely specified once we define a payment rule. For this, we make use of (5) and Footnote 17 to compute $\boldsymbol{M}_t$ for the buyers. Specifically, for $y_H = \tau$ and any $c \in \mathcal{C}$ we have

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

for $y_H \in \{0, \dots, \tau - 1\}$ (which, again, subsumes the case $y_L = 0$) we have

$$M_t^{B_t}(\overline{v}, \underline{c}, y_H) = \overline{v}, \quad M_t^{B_t}(\underline{v}, \underline{c}, y_H) = 0, \quad M_t^{B_t}(\overline{v}, \overline{c}, y_H) = \overline{v}\left(\frac{\delta\lambda}{1 - \lambda + \delta\lambda}\right)^n,$$
$$M_t^{B_t}(\underline{v}, \overline{c}, y_H) = 0,$$

for $y_L \in \{1, \dots, \tau - 1\}$ and any $c \in \mathcal{C}$ we have

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

and for $y_L = \tau$ we have

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

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. □

A.8 Proof of Proposition 5

Proof. By construction the posted-price mechanism does not run a deficit. Under truthful reporting the mechanism immediately executes efficient trades and does not execute any suboptimal trades when less than τ_0^* efficient traders are stored. Once τ_0^* efficient buyers are stored, the designer posts period t prices of $p_B = p_S = 1 - \Delta_0$, so that any efficient trades created in period t can still be executed and if a buyer of type $\overline{v}$ arrives with a seller of type $\underline{v}$, then a suboptimal $(\overline{v}, \underline{c})$ trade can also be executed. Similarly logic applies to the case in which τ_0^* efficient sellers are stored. It only remains to check the P-IC and P-IR constraints. Note that inefficient traders always receive a payoff of zero if they report truthfully and a non-positive payoff if they do not. Thus, the P-IC and P-IR constraints are always satisfied for these types and we only need to check the P-IC constraints of efficient traders. Consider the P-IC constraints of efficient buyers. If such a buyer arrives to a market with $p_B = \Delta_0$ then, regardless of the report of the arriving seller, truthful reporting yields the buyer the highest possible expected discounted payoff of $1 - \Delta$ and the P-IC constraint is satisfied.

If an efficient buyer arrives to a market with $p_B = 1/2$ then the LIFO queueing protocol ensures that the buyer will never trade in a period with $p_B = \Delta_0$. Thus, while truthful reporting guarantees the buyer a positive payoff, reporting to be of type $\underline{v}$ guarantees the buyer a payoff of zero. Since this holds regardless of the report of the arriving seller, the P-IC constraints for efficient buyers are satisfied. An analogous argument shows that the P-IC constraints are also satisfied for the efficient sellers. □

A.9 Proof of Corollary 2

Proof. This result follows from setting $\Delta = \Delta_0$ in Corollary A1. □

A.10 Proof of Proposition 6

Proof. Proving that the posted-price mechanism provides a P-IC and P-IR implementation of π_0^* proceeds in precisely the same manner as the proof of Proposition 5. Noting that the efficient allocation cannot be implemented by a P-IC and P-IR posted-price mechanism with a larger bid-ask spread completes the proof. □

A.11 Proof of Proposition 7

Proof. The change in the designer's expected per-period payoff when the state changes from $y = \tau_0^* - 1$ to $y = \tau_0^*$ is given by

$$w(\Delta_0 - 1) + w(1-w)\Delta_0 = w(1-w)\left[\Delta_0 - \frac{1-\Delta_0}{1-w}\right] = w(1-w)\Delta_1^{PP}.$$

In contrast, under the optimal efficient mechanism, this change is given by

$$w(1-w)\left[\Delta - \frac{w(1-\Delta)}{1-w}\right] = w(1-w)\Delta_1.$$

Thus, we can compute profit under the profit-maximizing posted-price implementation of efficiency using Δ_1^{PP}, provided $\tau_0^* \geq 1$. Since $\Delta_1 > \Delta_1^{PP}$ this shows that profit is strictly higher under the optimal efficient mechanism. Routine calculations show that $\Delta_0 < 1/2$ and $w < 1/2$ together imply $\Delta_1^{PP} < 0$. □

A.12 Proof of Proposition 8

Proof. As argued in the proof of Proposition 6, the designer will only post a bid-ask spread of 1 (to execute efficient trades) or Δ_0 (to execute efficient trades and one type of suboptimal trade) under the profit-maximizing post-price mechanism. Furthermore, as was shown in the proof of Proposition 6, $\Delta_1^{PP} < 0$ for any state $y \geq 1$. Therefore, the designer will optimally post a bid-ask spread of 1 in any state with $y \geq 1$. Thus, all that remains is to determine the optimal bid-ask spread for the state $y = 0$.

If the designer posts a bid-ask spread of Δ_0 in state $y = 0$, the market will always remain in the state $y = 0$ and designer's expected per period payoff is $w\Delta_0$. If the designer posts a bid-ask spread of 1 in state $y = 0$, any state $y \in \mathbb{N}$ is feasible and the designer earns a per period payoff of w^2 in state $y = 0$ and w in states $y \geq 1$. It immediately follows that if $w \geq \Delta_0$, then the profit-maximizing posted-price mechanism consists of posting a bid-ask spread of 1 in every period (or, equivalently, a pricing rule of $p^B = 1$ and $p^S = 0$ regardless of the state).

From this point onward, we assume that $\Delta_0 > w$ and determine when it is optimal to post a bid-ask spread of Δ_0 in every period and when it is optimal to post a bid-ask spread of 1 in every period. First, we characterize expected discounted profit under posted-price mechanism that posts a bid-ask spread of 1 in every period. Let $V_1^{PP}(y)$ denote the expected present value of having y identical suboptimal pairs present at the end of any period under this posted price mechanism. Letting $\lambda = w(1-w)$, these variables satisfy the infinite recursion

$$V_1^{PP}(y) = \delta[w + (1-2\lambda)V_1^{PP}(y) + \lambda V_1^{PP}(y-1) + \lambda V_1^{PP}(y+1)],$$

with boundary conditions

$$V_1^{PP}(0) = \delta[w^2 + (1-2\lambda)V_1^{PP}(0) + 2\lambda V_1^{PP}(1)] \quad \text{and} \quad \lim_{y\to\infty} V_1^{PP}(y) = \frac{\delta w}{1-\delta}.$$

Solving this recursion in MATHEMATICA with the boundary condition at $y = 0$ yields

$$V_1^{PP}(y) = \frac{\delta w}{1-\delta} + k_0\left(\frac{z_-}{2}\right)^y + \left(k_0 + \frac{\delta\lambda}{\sqrt{(1-\delta)(1-\delta+4\delta\lambda)}}\right)\left(\frac{z_+}{2}\right)^y,$$

where z_+ and z_- are defined in 14 and k_0 is a constant that will be pinned down by the boundary condition that must hold in the limit as $y \to \infty$. Elementary calculations (omitted for the sake of brevity) show that $z_+ > 2$ and $z_- \in (0, 2)$. Therefore, by setting $k_0 = -\delta\lambda/\sqrt{(1-\delta)(1-\delta+4\delta\lambda)}$, we satisfy the boundary condition in the limit as $y \to \infty$ and we finally have

$$V_1^{PP}(y) = \frac{\delta w}{1-\delta} - \frac{\delta \lambda}{\sqrt{(1-\delta)(1-\delta+4\delta\lambda)}} \left(\frac{z_-}{2}\right)^y.$$

Expected discounted profit in the state $y = 0$ under the posted-price mechanism that posts a bid-ask spread of 1 in every period is thus given by

$$\frac{w}{1-\delta} - \frac{\lambda}{\sqrt{(1-\delta)(1-\delta+4\delta\lambda)}},$$

while under the posted-price mechanism that posts a bid-ask spread of Δ_0 in every period this quantity is given by $\frac{w\Delta_0}{1-\delta}$. Therefore, under the assumption that $\Delta_0 > w$, the profit-maximizing posted-price mechanism involves posting a bid-ask spread of 1 in every period if and only if

$$\frac{w(1-\Delta_0)}{1-\delta} - \frac{\lambda}{\sqrt{(1-\delta)(1-\delta+4\delta\lambda)}} \geq 0.$$

Thus, making the substitution $\lambda = w(1-w)$ we obtain 11 as required. □

A.13 Proof of Proposition 9

Proof. We start by assuming that $w > 2\Delta_0$ 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 $R(q^{SB}) = w^2 + 2w(1-w)q^{SB}\Delta_1 = w(w + 2q^{SB}(\Delta_0 - w))$. 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 $q^{SB} = w/(2(w-\Delta_0))$. Noting that $2(w-\Delta_0) = w + w - \Delta_0 > w + \Delta_0 > w$, where the first inequality follows from $2\Delta_0 < w$ and the second from the assumption $\Delta_0 > 0$, 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 $S_0^{SB}(w) = w^2 + 2w(1-w)q^{SB}\Delta_0$. Hence, the relative gains from increasing market thickness once one accounts for the second-best mechanism in the static problem, denoted T_τ^{SB}, satisfy for any $\tau \geq 1$,

$$T_\tau^{SB} = \frac{S_\tau(w) - S_0^{SB}(w)}{S_\infty(w) - S_0^{SB}(w)} = \frac{2}{2\tau+1}\left(\tau + \frac{(1-q^{SB})\Delta_0}{1-2q^{SB}\Delta_0}\right) > \frac{2\tau}{2\tau+1} = T_\tau,$$

where the inequality follows because $2\Delta_0 < 1$.

Next, we derive δ^{SB}. Expected discounted welfare in the state $y = 0$ under the threshold policy $\tau = 1$ was denoted by $V_1(0)/\delta$ in the proof of Proposition 3 and an analytic expression for this quantity is given by (15). Under the threshold policy $\tau = 0$, assuming the second-best mechanism is used in the static bilateral trade problem, expected discounted welfare is $S_0^{SB}(w)/(1-\delta)$. Therefore, $\tau^{SB} \geq 1$ provided

$$\frac{(2\delta\Delta_0\lambda^2 + w^2(\delta(\lambda-1)+1) + 2\delta\lambda w)}{(1-\delta)(\delta(3\lambda-1)+1)} \geq \frac{w^2 + 2\lambda q^{SB}\Delta_0}{1-\delta}$$

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 that $\tau^{SB} \geq 1$ if and only if

$$\delta \geq \delta^{SB} = \frac{\Delta_0}{\lambda + \Delta_0(1-2\lambda) + (1-w)(w-2\Delta_0)(1+\Delta)}.$$

Note that $(1-w)(w-2\Delta_0)(1+\Delta) > 0$ since $w > 2\Delta > 0$ and $w < 1$. Thus, comparing δ^{SB} to the expression for δ_0^* from Corollary 2, we have that $\delta^{SB} < \delta_0^*$ as required.

Finally, that $\tau_0^{SB} = 1$ for any $\delta \in (\delta^{SB}, \delta^*]$ and $\tau_0^{SB} = \tau_0^*$ for any $\delta > \delta^*$ follows immediately from Proposition 5. That is, since the first-best mechanism can be implemented without running a deficit whenever $\tau_0^* \geq 1$, using the second-best mechanism in the static bilateral trade problem only affects the decision to store the first suboptimal pair. $\square$

A.14 Proof of Proposition 10

Proof. We begin with the definition of the mechanism. Identical efficient traders (sellers willing to sell at p or buyers willing to buy at p) are stored up to a threshold $\tau \geq 1$. We let let τ_L (τ_H) be the state in which τ sellers (buyers) are stored. The two-class threshold mechanism then commences each period by posting the large market Walrasian price p and what occurs following this depends on the state. First, consider states $y < \tau$. Agents who arrive observe the price and y and indicate their demands and supplies as in the posted-price mechanisms analyzed in Section 4.3. If there is a pair willing to trade at p, an (efficient) trade is executed using a LIFO queueing protocol (in the event of excess demand or excess supply).

Second, in states τ_L and τ_H, the price p is posted and a sequential game is played as follows. In state τ_L, the arriving seller first indicates whether they are willing to trade at the price p.[39] 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 and if the buyer is unwilling to trade at this price they are 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) second-best mechanism for the case where the seller's cost distributions is $G(c)/G(p)$ for $c \in [0, p]$ and the buyer's distribution is $F(v)$ for $v \in [0, 1]$. Thus, the agents that arrive to the market in state τ_L always end up leaving the market during that period, regardless of any trading outcomes. Analogously, in state τ_H, the arriving buyer first indicates whether they are 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 $F(v)/(1 - F(p))$ for $v \in [p, 1]$. Whether or not they trade, both arriving agents are cleared from the market.

We now prove the asserted properties. The P-IC and P-IR properties follow along the same lines as for the posted-price mechanisms. This works because traders all face a price of p, except for those that participate in the second-best mechanism (which will be addressed temporarily). The LIFO queueing protocol ensures that no efficient type can misreport and subsequently participate in the second-best mechanism. If an inefficient type misreports and ends up participating in the second-best mechanism, this guarantees that agent a non-positive payoff. Consider then the agents who participate in the second-best mechanism. Notice that no agent 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 agents, this is a one-shot game and the dominant-strategy properties of these mechanisms imply (P-IC) and (P-IR).

Given this equilibrium behaviour, the dynamics of the Markov decision problem are the same as those derived in Section 3 and the stationary distribution is given by Lemma 1. The only state in which trade does not always occur at the price p is the state τ. The probability of not being in that state is $\frac{2\tau-1}{2\tau+1}$, which goes to 1 as τ goes to infinity. We are thus only left

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

to establish $T_\tau^{TC} > T_\tau$ for $\tau \geq 1$.

Let f and g denote the respective densities of the probability distributions F and G. 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.[40] 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 agents 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 α_0 this value of α, where $\alpha_0 > 0$ 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 α_L such that

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

Notice that $\int_0^1\int_0^p(\Phi(v)-\Gamma(c))Q^{\alpha_0}(v,c)dG(c)dF(v)/G(p) > 0$ 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

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

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.

Second-best welfare in the one-shot bilateral trade problem 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

$$S_0^{SB} = wS_\infty + w(1-w)(\Delta_L^0+\Delta_H^0) - w^2\varepsilon^0. \tag{17}$$

[40] 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.

Denoting by $s^{TC}(y)$ the expected surplus under the two-class mechanism in state y, we have $s^{TC}(0) = wS_\infty$ and $s^{TC}(y) = S_\infty$ for any $y \in \{1\ldots,\tau-1\}$. Let $s_L^{TC}(\tau)$ and $s_H^{TC}(\tau)$ denote expected surplus in state τ 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_τ^{TC} denote the expected per-trader surplus for the two-class mechanism under the stationary distribution we thus have

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

which yields

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

To show that $T_\tau^{TC} > T_\tau = 1 - \kappa_0$ it thus suffices to show that the second term on the right-hand side of this last equation is positive. Using (17) and regrouping terms we have

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

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 $T_\tau^{TC} > T_\tau$ as required. □

A.15 Proof of Theorem 3

Proof. Under discriminatory, uniform and fixed frequency market clearing per period welfare converges to w as $\delta \to 1$ and under instantaneous market clearing is always given by $w^2 + 2w(1-w)\Delta_0$. Hence, $\lim_{\delta\to 1} W_\alpha^U(\delta)/W_\alpha^D(\delta) = \lim_{\delta\to 1} W_\alpha^F(\delta)/W_\alpha^U(\delta) = 1$ and $\lim_{\delta\to 1} W_\alpha^0(\delta)/W_\alpha^F(\delta) = w + 2(1-w)\Delta_0$ and the desired result immediately follows. □

A.16 Proof of Theorem 4

For ease of exposition, we formally prove the theorem for the $k = D$ case. The other cases are similar. However, before proceeding to the proof of the main result, we start by stating and proving the following lemma.[41]

Lemma A1. *Let the optimal threshold τ^* under discriminatory market clearing and integer $i \geq 0$ be given. Then $V_\tau^D(i)$ increases in τ for $\tau \in \{0, 1, \ldots, \tau^* - 1\}$, decreases in τ for $\tau \geq \tau^* + 1$ and attains a global maximum at $\tau = \tau^*$.*

Proof. First, take $\tau = \tau^* + 1$. Then $V_{\tau-1}^D(i) \geq V_\tau^D(i)$ by the principle of optimality of dynamic programming. Next, consider $\tau = \tau^* + 2$. To prove that $V_{\tau-1}^D(i) \geq V_\tau^D(i)$ it suffices to show that this holds for $i = \tau$ (since this is the only state at which the policies π_τ and $\pi_{\tau-1}$ diverge). In this case we have $V_{\tau-1}^D(\tau) = r(0,1) + V_{\tau-1}^D(\tau - 1) \geq V_\tau^D(\tau)$ since clearing two suboptimal pairs in state τ is optimal under τ^* and there are no complementaries associated with clearing multiple pairs (i.e. $r(0,2) = r(0,1) + r(0,1)$). Thus, clearing one suboptimal pair in state τ must yield a higher payoff than taking no action. Iterating, we have that $V_\tau^D(i)$ decreases in τ for $\tau \geq \tau^*$.

Second, set $\tau = \tau^* - 1$. Then $V_{\tau+1}^D(i) \geq V_\tau^D(i)$ by the principle of optimality of dynamic programming. Next, setting $\tau = \tau^* - 2$ we have $V_{\tau+1}^D(\tau+1) \geq r(0,1) + V_\tau^D(\tau) = V_\tau^D(\tau+1)$ since storing in state $\tau + 1$ is optimal under τ^*. Iterating, we have that $V_\tau^D(i)$ increases in τ for $\tau \in \{0, 1, \ldots, \tau^*\}$. Putting all of this together implies that we have a global maximum at $\tau = \tau^*$. □

We now prove Theorem 4 for $k = D$.

Proof. First, let $W^\infty(\delta)$ denote welfare under the market clearing policy with $\tau^* = \infty$ (that is, the mechanism under which all suboptimal trades are stored indefinitely and only efficient trades are executed). Further, let $\tau_0^D(\delta)$ and $\tau_1^D(\delta)$ denote the optimal thresholds under welfare-maximizing and profit-maximizing discriminatory market clearing, respectively. Recall that these functions are both increasing in δ. Further, $W^\infty(\delta)$ and $W_0^0(\delta)$ are continuous,

[41] This lemma applies specifically to discriminatory market clearing, a version that applies to uniform market clearing (and also holds for uniform market clearing) can be found in Lemma B2.

increasing functions of δ and there exists $\tilde{\delta}$ such that $W^{\infty}(\tilde{\delta}) > W_0^0(\tilde{\delta})$ because $\tau_0^D(\delta) \to \infty$ as $\delta \to 1$. Since $\tau_0^D(\delta) \leq \tau_1^D(\delta) < \infty$ for $\delta \in [0, 1)$ by Corollary 1, Lemma A1 implies that $W_1^D(\delta) \geq W^{\infty}(\delta)$ for all $\delta \in [0, 1)$. We also have that $W_1^D(\delta)$ decreases discontinuously at points at which $\tau_1^D(\delta)$ increases (again, by Lemma A1) and increases continuously in δ at all points at which $\tau^{D,1}(\delta)$ is constant (that is, the underlying market clearing policy does not vary). It immediately follows that there exists $\delta_D \leq \tilde{\delta}$ such that $W_1^D(\delta_D) = W_0^0(\delta_D)$ and $W_1^D(\delta) > W_0^0(\delta)$ for $\delta > \delta_D$. □

Online Appendices

B Uniform market clearing

Under uniform market clearing, the state space, transition probabilities and reward function of the associated Markov decision process are the same as those of the Markov decision process derived in Section 3 for discriminatory market clearing. Uniform market clearing only affects the set of actions available to the designer in a given state. Let $\mathcal{A}'_{\boldsymbol{x}}$ denote the set of actions available to the designer in state $\boldsymbol{x}$ under uniform market clearing. Under discriminatory market clearing we had $\mathcal{A}_{\boldsymbol{x}} = \{(a_E, a_S) : a_E, a_S \in \mathbb{Z}_{\geq 0}, a_S \leq x_E, a_S \leq x_S\}$. However, for the uniform market clearing case the designer can elect only to wait or clear the entire market, implying that $\mathcal{A}'_{\boldsymbol{x}} = \{(x_E, x_S), (0, 0)\}$. Setting $\mathcal{A}' = \cup_{\boldsymbol{x}\in\mathcal{X}}\mathcal{A}'_{\boldsymbol{x}}$, we need to determine the optimal policy of the Markov decision process $\langle \mathcal{X}, \mathcal{A}', P, r, \delta\rangle$. Recall that P specifies the transition probabilities of the Markov chain, $r(\boldsymbol{x})$ specifies the reward earned by the designer when action $\boldsymbol{a} = \boldsymbol{x}$ is implemented and δ is the discount factor. Here, P, r and δ are the same for both discriminatory and uniform market clearing.

In general, we will use the term *threshold policies* to describe any class of policies which can be summarized by one-dimensional sufficient statistics, the thresholds τ. As was the case with discriminatory market clearing, under uniform market clearing we can restrict attention to a class of threshold polices. We then use the structure that threshold policies impose to prove that the optimal policy is a threshold policy.

Definition B1. *Given a threshold $\tau \in \mathbb{N}$, the associated* threshold policy π_τ *of the Markov decision process $\langle \mathcal{X}, \mathcal{A}', P, r, \delta\rangle$ is such that*

$$\pi_\tau(\boldsymbol{x}) = \boldsymbol{0} \quad \text{if} \quad r(\boldsymbol{x}) \leq \tau \quad \text{and} \quad \pi_\tau(\boldsymbol{x}) = \boldsymbol{x} \quad \text{if} \quad r(\boldsymbol{x}) > \tau.$$

Under a threshold policy the market maker stores both efficient and suboptimal pairs up to a threshold value of τ. We now describe the associated structure of the Markov chain $\{\boldsymbol{Y}_t\}_{t\in\mathbb{N}}$, as illustrated in Figure 6. One can think of the number of stored efficient pairs as the *level* of the Markov chain and the number of stored suboptimal trades as the *phase* of the Markov chain within that level. We include an additional level for the state $\boldsymbol{0}$, denoted by level $\emptyset$. Under the threshold policy τ, $\overline{y}_E = \lfloor \tau \rfloor$ is the maximum number of efficient pairs that can be stored. For $i \in \{0, 1, \ldots, \overline{y}_E\}$, the maximum number of suboptimal pairs stored at level i is $\overline{k}_i = \lfloor (\tau - i)/\Delta_\alpha \rfloor$, where $\alpha \in \{0, 1\}$. Therefore, the Markov chain is a level-dependent quasi-birth-death process (see, for example, Latouche and Ramaswami, 1999). Similarly to the case of discriminatory market clearing, we can exploit its structure to show that the optimal market clearing policy is a threshold policy.

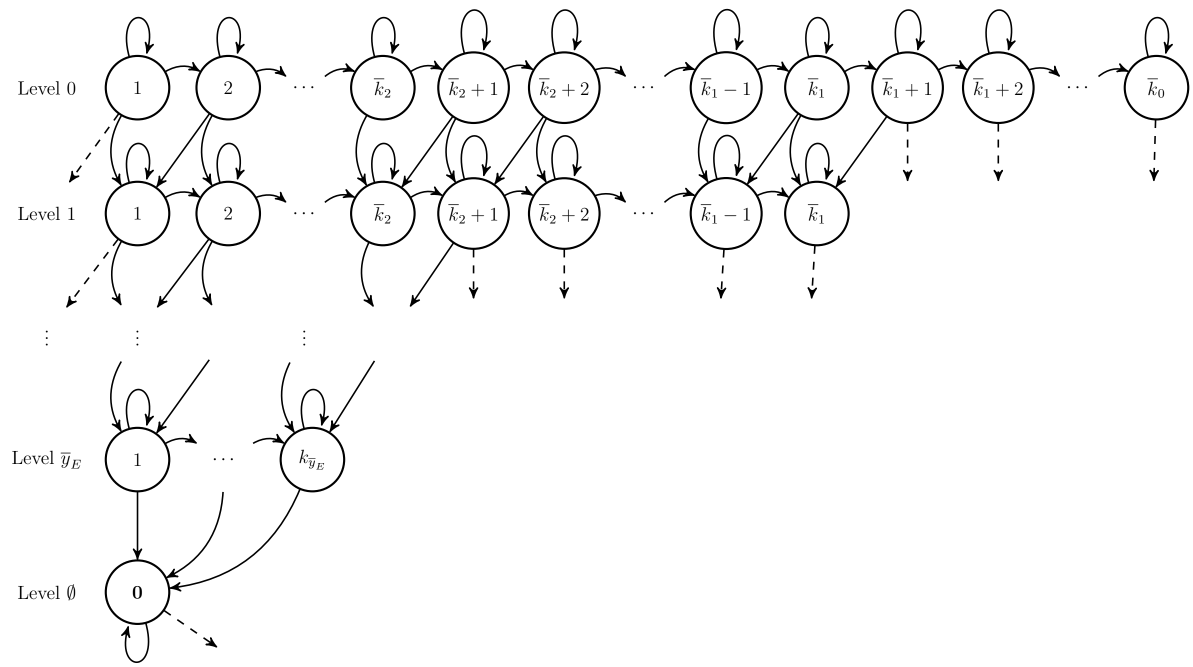


Figure 6: The structure of the quasi-birth-death under the threshold policy with threshold τ. Dashed arrows are used to denote some transitions to and from the state $\mathbf{0}$.

Theorem B1. *Under uniform market clearing, the optimal market clearing policy is a threshold policy. It can be implemented using a P-IC and P-IR mechanism.*

The proof of Theorem B1 proceeds in a similar manner to the proof of Theorem 2, using a dynamic programming characterization of the optimal threshold τ^*. Algorithm D1 in Online Appendix D uses the optimal stopping condition derived from the Bellman equation to compute τ^*.

Proof. Since the state space $\mathcal{X}$ is countable, the feasible action sets $\mathcal{A}'_{\boldsymbol{x}}$ are finite for all states $\boldsymbol{x}$ and the reward function is deterministic, a stationary deterministic optimal policy exists (see, for example, Theorem 6.2.10 of Puterman, 1994). Let π^* denote any optimal policy of the Markov decision process $\langle \mathcal{X}, \mathcal{A}', P, r, \delta \rangle$. The optimal policy must clear the market whenever it is in a state of the form $(x_E, 0)$, with $x_E \in \mathbb{N}$. Furthermore, given any fixed number of stored efficient pairs, as the number of stored suboptimal pairs diverges to infinity, the expected time until each additional stored pair is rematched diverges to infinity. Therefore, the benefit of storing each additional suboptimal pair converges to zero, while the immediate reward for clearing a suboptimal pair from the market is fixed at Δ_α, where $\alpha \in \{0, 1\}$. Thus, for a given number of stored efficient pairs, the optimal policy cannot allow an unbounded number of identical suboptimal pairs to accumulate.

It follows that for every $x_E^* \in \mathbb{Z}_{\geq 0}$ there exists a state $\boldsymbol{x}^* = (x_E^*, x_S^*)$ such that $\pi^*(\boldsymbol{x}^*) = \mathbf{0}$ and $\pi^*(x_E^*, x_S^* + 1) = (x_E^*, x_S^* + 1)$. We call such states *cutoff* states. Denote the expected present value of being in the cutoff state $\boldsymbol{x}^*$ under the optimal policy by $V_{\pi^*}^U(\boldsymbol{x}^*)$, the total expected discounted reward earned by the designer in the subsequent period. It is finite because an unbounded number of pairs cannot accumulate under π^* and we are considering a discounted process. For any state $\boldsymbol{x}$, the benefit of waiting to clear the market is increasing in x_S and the benefit of clearing is increasing in $r(\boldsymbol{x})$. Since $r(x_E^* + 1, x_S^*) > r(\boldsymbol{x}^*)$ and $r(x_E^* + 1, x_S^* - 1) > r(\boldsymbol{x}^*)$ it follows that if $\pi^*(x_E^*, x_S^* + 1) = (x_E^*, x_S^* + 1)$, we must also have $\pi^*(x_E^* + 1, x_S^*) = (x_E^* + 1, x_S^*)$ and $\pi^*(x_E^* + 1, x_S^* - 1) = (x_E^* + 1, x_S^* - 1)$. Finally, let $V_{\pi^*}^U(\mathbf{0})$ denote the expected present value of being in the state $\mathbf{0}$ under the optimal policy. The Bellman equation which characterizes $V_{\pi^*}^U(\boldsymbol{x}^*)$ is, for $\alpha \in \{0, 1\}$, then given by

$$V_{\pi^*}^U(\boldsymbol{x}^*) = \delta \left[w^2(r(\boldsymbol{x}^*) + 1 + V_{\pi^*}^U(\mathbf{0})) + w(1-w)(r(\boldsymbol{x}^*) + \Delta_\alpha + V_{\pi^*}^U(\mathbf{0})) \right.$$
$$\left. + w(1-w)(r(\boldsymbol{x}^*) + 1 - \Delta_\alpha + V_{\pi^*}^U(\mathbf{0})) + (1-w)^2 V_{\pi^*}^U(\boldsymbol{x}^*) \right]. \tag{18}$$

If the market is cleared in state $\boldsymbol{x}^*$, the payoff is the immediate reward $r(\boldsymbol{x}^*)$ plus the expected present value of being in the state $\mathbf{0}$. By the principle of the optimality of dynamic programming,

$$V_{\pi^*}^U(\boldsymbol{x}^*) \geq r(\boldsymbol{x}^*) + V_{\pi^*}^U(\mathbf{0}). \tag{19}$$

Notice that the right-hand sides of (18) and (19) depend directly on $\boldsymbol{x}^*$ only through $r(\boldsymbol{x}^*)$. Replace $r(\boldsymbol{x}^*)$ with τ^* in (18) and (19) and suppose (19) holds with equality. Then, for every cutoff state $\boldsymbol{x}^*$, $r(\boldsymbol{x}^*) \leq \tau^*$. Using the definition of τ^*, substituting (19) into (18) and rearranging, it can be shown that τ^* satisfies

$$\tau^* + V_{\pi^*}^U(\mathbf{0}) = \frac{\delta w}{1 - \delta}. \tag{20}$$

Thus, for any state $\boldsymbol{x} \in \mathcal{X} \setminus \{(x_E, 0) : x_E \in \mathbb{N}\}$, the market should be cleared if and only if $x_E^* + \Delta_\alpha x_S^* > \tau^*$. Therefore, the optimal policy π^* is a threshold policy, where the threshold $\tau^* \in \mathbb{R}_{\geq 0}$ is characterized by (20).

We now show that the optimal threshold policy can be implemented with a P-IC and P-IR mechanism. Start by constructing a direct allocation rule from the optimal market clearing policy. Let $\hat{h} \in \{\overline{v}, \underline{v}\}^{\mathbb{N}} \times \{\underline{c}, \overline{c}\}^{\mathbb{N}}$ be a realization of the report process and $\hat{h}_t$ denote $\hat{h}$ restricted to its first $2t$ components. Let $\{\tau_j^{\hat{h}}\}_{j \in \mathbb{N}}$ denote the subset of periods such that the designer optimally chooses to clear the market under π^*, given $\hat{h}$ and set $\tau_0^{\hat{h}} = 0$ for convenience. For all $i \in \mathbb{N}$ there exists $j \in \mathbb{N}$ such that $\tau_{j-1}^{\hat{h}} < i \leq \tau_j^{\hat{h}}$. The period $\tau_j^{\hat{h}}$

history of reports can be mapped to $X_{\tau_j^{\hat{h}}}$, the state of $\langle \mathcal{X}, \mathcal{A}', P, r, \delta \rangle$ in period $\tau_j^{\hat{h}}$. Then, fixed a queueing protocol μ that is used to break ties, if buyer i is part of an efficient or a suboptimal pair in period $\tau_j^{\hat{h}}$ we simply set $Q^{B_i}_{\tau_j^{\hat{h}}}(\hat{h}_{\tau_j^{\hat{h}}}) = 1$ and, for all $k \in \mathbb{N} \setminus \{\tau_j^{\hat{h}}\}$, $Q_k^{B_i}(\hat{h}_k) = 0$. Otherwise, we set $Q_k^{B_i}(\hat{h}_k) = 0$ for all $k \in \mathbb{N}$. Proceed analogously for seller i.

To show that this allocation rule can be implemented using a P-IC and P-IR mechanism, it suffices to verify that $q^{B_i}(\overline{v}, \hat{h}_{i-1}) \geq q^{B_i}(\underline{v}, \hat{h}_{i-1})$ and $q^{S_i}(\underline{c}, \hat{h}_{i-1}) \geq q^{S_i}(\overline{c}, \hat{h}_{i-1})$. These constraints hold under π^* since the arrival of a $\overline{v}$ or $\underline{c}$ agent cannot increase the expected number of periods until the next market clearing event (since the Markov chain transitions to a state with fewer expected transitions between it and the $\mathbf{0}$ state) and $\overline{v}$ and $\underline{c}$ agents are more likely to trade as part of any given market clearing event (since these agents have rematching priority over $\underline{v}$ and $\overline{c}$ agents). □

An analogous result to Proposition 3, specifically that τ^* is decreasing in Δ and increasing in δ, immediately follows from the dynamic programming characterization.[42] Similarly, our measure T_τ of market thickness is still valid under uniform market clearing and we have that optimal market thickness T^* is decreasing in Δ and increasing in δ. Finally, we prove the following useful lemma.

Lemma B2. *Let the optimal policy π^* be given and consider the associated set $\mathcal{Y}^*$ of positive recurrent states of the Markov chain. Then for any policies π and π' with respective positive recurrent sets $\mathcal{Y}$ and $\mathcal{Y}'$ such that $\mathcal{Y} \subset \mathcal{Y}' \subset \mathcal{Y}^*$ we have $V^U_{\pi^*}(\boldsymbol{x}) \geq V^U_{\pi'}(\boldsymbol{x}) \geq V^U_{\pi}(\boldsymbol{x})$ for all states $\boldsymbol{x} \in \mathcal{X}$.*

Proof. That $V^U_{\pi^*}(\boldsymbol{x}) \geq V^U_{\pi'}(\boldsymbol{x})$ and $V^{DU}_{\pi^*}(\boldsymbol{x}) \geq V^D_{\pi}(\boldsymbol{x})$ for all states $\boldsymbol{x} \in \mathcal{X}$ follows from the principle of optimality of dynamic programming. That $V^U_{\pi'}(\boldsymbol{x}) \geq V^U_{\pi}(\boldsymbol{x})$ follows from the fact that $\mathcal{Y}' \setminus \mathcal{Y} \subset \mathcal{Y}^*$, so storing in the states $\mathcal{Y}' \setminus \mathcal{Y}$ is optimal under π^*. □

C Fixed frequency market clearing

Under fixed frequency market clearing, the state space of the Markov chain is given by $\{(y_E, y_S) : 0 \leq y_E + y_S \leq \tau, y_E, y_S \in \mathbb{Z}^2_{\geq 0}\}$. If the market is cleared after τ periods, then both the number of buyers of type $\overline{v}$ present and the number of sellers of type $\overline{c}$ present follow a binomial distribution with τ trials and probability of success p, and likewise for the numbers of buyers of type $\underline{v}$ and sellers of type $\underline{c}$ present. Thus, the market maker's expected

[42]Note that τ^* is no longer increasing in $w(1-w)$ since now both efficient and suboptimal pairs are stored.

discounted payoff is given by

$$V_\tau^F = \frac{\delta^{\tau-1}}{1-\delta^\tau} \sum_{j=0}^{\tau} \sum_{k=0}^{\tau} \binom{\tau}{j} \binom{\tau}{k} (\min\{j,k\} + |j-k|\,\Delta_\alpha)\, w^{j+k}(1-w)^{2\tau-j-k}. \tag{21}$$

Here, the market maker can only determine the frequency at which markets are cleared. Thus, the optimal market clearing policy is trivially a threshold policy, where the market is cleared every τ^* periods. Algorithm D2, which can be found in Online Appendix D, uses this formula to compute the optimal market clearing threshold τ^*.

Once again, we have that τ^* is increasing in δ and decreasing in Δ_0. Our measure T_τ of market thickness continues to be valid for fixed frequency market clearing and the comparative statics that apply to τ^* holds for T^*. We also immediately have a version of Lemma B2 that applies to fixed frequency market clearing. Finally, we have the following result.

Corollary C2. *Under fixed frequency market clearing, the optimal market clearing policy is a threshold policy. It can be implemented using a P-IC and P-IR mechanism.*

Proof. Under fixed frequency market clearing, threshold policies are trivially optimal. We can repeat the procedure from the proof of Theorem B1 in order to construct a direct allocation rule for fixed frequency market clearing. However, in this case the set of optimal market clearing times is deterministic and given by $\{i\tau^*\}_{i\in\mathbb{N}}$. The constraints $q^{B_i}(\overline{v}) \geq q^{B_i}(\underline{v})$ and $q^{S_i}(\underline{c}) \geq q^{S_i}(\overline{c})$ must then hold since $\overline{v}$ and $\underline{c}$ agents have rematching priority over $\underline{v}$ and $\overline{c}$ agents. □

D Algorithms

D.1 Uniform Market Clearing

We next define a similar algorithm that applies to uniform market clearing. However, first we must derive the Bellman equation that characterizes the optimal threshold τ^*. We start by introducing the notation $Z = \{(0,0),(0,1),(1,0),(1,-1)\}$, which captures the set of possible changes to the state $\boldsymbol{y} = (y_E, y_S)$ following the next arrival. Introducing this notation is convenient because it allows us to sum over all possible transitions of the Markov chain. Define the function $P_Z : Z \to [0,1]$ by

$$P_Z(1,0) = w^2, \quad P_Z(0,1) = w(1-w), \quad P_Z(1,-1) = w(1-w) \quad \text{and} \quad P_Z(0,0) = (1-w)^2,$$

which gives the probability of each of the changes captured in Z. For example, $(1,-1)$ corresponds to the arrival of a suboptimal pair that results in a stored suboptimal being rematched to create an efficient pair. This occurs with probability $w(1-w)$, provided $y_S > 0$.

Let $V_\tau^U(y_E, y_S)$ denote the expected discounted present value of being in state (y_E, y_S) under the threshold policy with threshold τ. If the state of the market is $(y_E, 0)$ for some $y_E > 0$, the market maker will immediately clear and earn a reward of y_E plus the expected present vale of being in state $\mathbf{0}$. Therefore, we have

$$V_\tau^U(y_E, 0) = y_E + V_\tau^U(\mathbf{0}). \tag{22}$$

Next suppose the market is in any state $\boldsymbol{y} = (y_E, y_S)$ such that $y_S > 0$ and $r(\boldsymbol{y}) < \tau$, where $r(y_E, y_S) = y_E + y_S\Delta_\alpha$ denotes the immediate reward from clearing the market. Under the threshold policy τ, the market maker will earn an immediate reward only when the market reaches a state $\boldsymbol{y}'$ such that $r(\boldsymbol{y}') \geq \tau$. Consequently,

$$\begin{aligned} V_\tau^U(\boldsymbol{y}) = \delta \sum_{\boldsymbol{z}\in Z} P_Z(\boldsymbol{z}) \big[& V_\tau^U(\boldsymbol{y}+\boldsymbol{z})\mathbb{1}(r(\boldsymbol{y}+\boldsymbol{z}) < \tau) \\ & + (r(\boldsymbol{y}+\boldsymbol{z}) + V_\tau^U(\mathbf{0}))\mathbb{1}(r(\boldsymbol{y}+\boldsymbol{z}) \geq \tau)\big]. \end{aligned} \tag{23}$$

Any threshold policy is characterized by this linear system. As with discriminatory market clearing, this Bellman equation can be used to to derive a stopping condition satisfied by τ^*.

By the proof of Theorem B1, the optimal threshold τ^* is such that for any $x_E^* > 0$ there exists a cutoff state $\boldsymbol{x}^* = (x_E^*, x_S^*)$ with

$$V_{\tau^*}^U(\boldsymbol{x}^*) > r(\boldsymbol{x}^*) + V_{\tau^*}^U(\mathbf{0}) \quad \text{and} \quad V_{\tau^*}^U(x_E^*, x_S^*+1) \leq r(x_E^*, x_S^*+1) + V_{\tau^*}^U(\mathbf{0}).$$

That is, a cutoff state is such that the market is optimally cleared if an additional identical suboptimal pair arrives. In the proof of Theorem 2, we show that this implies that the market is then also optimally cleared if an efficient or a non-identical suboptimal pair arrives. Since τ^* applies to all cutoff states, to compute τ^* it suffices to find a single cutoff state. Algorithm D1 determines τ^* by computing the cutoff state $(0, x_S^*)$ using the aforementioned stopping condition.

Algorithm D1. *Begin with the threshold policy characterized by $\tau = \Delta_\alpha$, where Δ_α is the value of a single suboptimal trade. Solve the linear system defined in (22) and (23). If $V_\tau^U(0,1) \geq \Delta_\alpha + V_\tau^U(\mathbf{0})$, proceed to step 2. Otherwise, return $\tau^* = 0$. At step i,*

1. *Solve (22) and (23) with $\tau = i\Delta_\alpha$ to determine $V_\tau^U(0,i)$ and $V_\tau^U(\mathbf{0})$.*

2. *If* $V_\tau^U(0,i) \geq i\Delta_\alpha + V_\tau^U(\mathbf{0})$*, proceed to step* $i+1$*. Otherwise, set* $\tau' = (i-1)\Delta_\alpha$*.*

If $\tau' + \Delta_\alpha < 1$*, return* $\tau^* = \tau'$*. Otherwise, for all* $j \in \mathbb{N}$ *such that* $\tau' + \Delta_\alpha < j$*,*

1. *Set* $k = \lfloor(\tau' + \Delta_\alpha - j)/\Delta_\alpha\rfloor$ *and solve* (22) *and* (23) *with* $\tau = j + k\Delta_\alpha$ *to determine* $V_\tau^U(j,k)$ *and* $V_\tau^U(\mathbf{0})$*.*

2. *If* $V_\tau^U(k,j) \geq j + k\Delta_\alpha + V_\tau^U(\mathbf{0})$ *update* $\tau' = j + k\Delta_\alpha$*.*

Return $\tau^* = \tau'$*.*

Note that depending on the value of δ, Algorithm D1 may not be the most economical algorithm. For example, for larger value of δ, a computationally more efficient algorithm could proceed by initially increasing the candidate threshold by increments of 1 and then increasing the candidate threshold by increments of Δ_α.

D.2 Fixed Frequency Market Clearing

Finally, we have an algorithm to compute τ^* under fixed frequency market clearing.

Algorithm D2. *Begin with the threshold policy characterized by* $\tau = 2$ *and compute* $W^F(2)$ *and* $W^F(1)$ *using* (21)*. If* $W^F(2) \geq W^F(1)$ *proceed to step 2. Otherwise, return* $\tau^* = 1$*. At step* i*,*

1. *Compute* $W^F(i)$ *using* (21)*.*

2. *If* $W^F(i) \geq W^F(i-1)$*, proceed to step* $i+1$*. Otherwise, return* $\tau^* = i - 1$*.*