Optimal Dynamic Matching

Mariagiovanna Baccara* SangMok Lee† Leeat Yariv‡§

August 9, 2015

Abstract

We study a dynamic matching environment where individuals arrive sequentially. There is a tradeoff between waiting for a thicker market, allowing for higher quality matches, and minimizing agents' waiting costs. The optimal mechanism cumulates a stock of incongruent pairs up to a threshold and matches all others in an assortative fashion instantaneously. In decentralized settings, a similar protocol ensues in equilibrium, but expected queues are inefficiently long. We quantify the welfare gain from centralization, which can be substantial, even for low waiting costs. We also evaluate welfare improvements generated by transfer schemes and by matching individuals in fixed time intervals.

Keywords: Dynamic Matching, Mechanism Design, Organ Donation, Market Design.

*Olin School of Business, Washington University in St. Louis. E-mail: mbaccara@wustl.edu

†Department of Economics, University of Pennsylvania. E-mail: sangmok@sas.upenn.edu

‡Division of the Humanities and Social Sciences, Caltech. E-mail: lyariv@hss.caltech.edu

§We thank Andy Postlewaite and Larry Samuelson for very helpful comments. We gratefully acknowledge financial support from the National Science Foundation (SES 0963583).

1 Introduction

1.1 Overview

Many matching processes are inherently dynamic, with participants arriving and matches being created over time. For instance, in the child-adoption process, parents and children arrive steadily – data from one US adoption facilitator who links adoptive parents and children indicates a rate of about 11 new potential adoptive parents and 13 new children entering the facilitator's operation each month.[1] While the overall statistics on the entry of parents and children into the US adoption process are not well-documented, adoption touches upon many lives; The Census 2000 indicates that about 1.6 million or 2.5 percent of all children have been adopted. Likewise, many labor markets entail unemployed workers and job openings that become available at different periods – the US Bureau of Labor Statistics reports approximately five million new job openings and slightly fewer than five million newly unemployed workers each month this year. A similar picture emerges when considering organ donation. According to the Organ Donation and Transplantation Statistics, a new patient is added to the kidney transplant list every 14 minutes and about 3000 patients are added to the kidney transplant list each month. A significant fraction of transplants are carried out using live donors – in 2013, about a third of nearly $17,000$ kidney transplants that took place in the US involved such donors.

Nonetheless, by and large, the extant matching literature has taken a static approach to market design – participants all enter at the same time and the market's operations are restricted in their horizon (see the literature review below for several important exceptions). In the current paper, we offer techniques for extending that approach to dynamic settings.

All of the examples we mentioned above share two important features. First, the quality of matches varies. In other words, agents care about the agents with whom they match. Second, waiting for a match is costly, be it for financial costs of keeping lawyers on retainer for potential adoptive parents, children's hardship from growing older in the care of social services, the lack of wages and needed employees in labor markets, or medical risks for organ patients and psychological waiting costs for donors. These two features introduce a crucial trade-off – on the one hand, a thick market can help generate a greater match surplus; on the other hand,

[1] This adoption facilitator is one of 25 registered in its state of operation. See Baccara, Collard-Wexler, Felli, and Yariv (2014) for details.

a thin market allows for quicker matching and cuts down on waiting costs. The goal of this paper is to characterize the resolution of this trade-off in both centralized and decentralized environments. Namely, we identify the optimal protocol by which a social planner would match agents over time. We also identify the conditions under which decentralized matching processes would especially benefit from centralized intervention using the optimal protocol.

Specifically, we consider a market that evolves dynamically. There are two classes of agents, which we refer to as squares and rounds. At each period, a pair consisting of a square and a round enters the market. Squares and rounds each have two types, one type more desirable to the other side than the other. For instance, if we think of squares and rounds as children relinquished for adoption and potential adoptive parents, types can stand for gender of children and wealth levels of potential adoptive parents, respectively (see Baccara, Collard-Wexler, Felli, and Yariv, 2014). Alternatively, if we think of the two classes of agents as workers and firms, types can represent skills of workers and benefit packages offered by firms. We assume that preferences are super-modular so that the (market-wide) assortative matching maximizes joint welfare. We also assume that, once agents arrive at the market, waiting before being matched comes at a per-period cost.

We start by analyzing the optimal matching mechanism in such settings. We show that the optimal mechanism takes a simple form. Whenever congruent pairs of agents – a square and round that are both of the more desirable type or the less appealing type – are present in the market, they are matched instantaneously. When only incongruent agents are present in the market, they are held in a queue. When the stock of incongruent pairs in the queue exceeds a certain threshold, they are matched in sequence, until the queue length falls within the threshold. Such thresholds induce a Markov process, where states correspond to the length of queues of incongruent pairs of agents. Any threshold yields a different steady-state distribution over possible queue lengths. We evaluate the expected welfare of the mechanism in the steady-state; The optimal mechanism utilizes the threshold that maximizes welfare. When waiting costs are vanishingly small, the welfare under the optimal mechanism approaches the maximum feasible, that generated by no matches of incongruent pairs. As time frictions, or waiting costs increase, the welfare generated by the optimal mechanism decreases.

This welfare decrease raises the question of the value of dynamic clearinghouses for non-trivial waiting costs in different environments, identified by type distributions and preferences. We therefore study the performance of a simple decentralized matching process in our setting.

As before, we consider agents arriving at the market in sequence. At each period in which they are on the market, agents declare their willingness to match with partners of either type. After these demands have been made, the maximal number of pairs of willing agents are matched in order of arrival.[2] Those who prefer to stay in the market, or have to stay for lack of willing partners, form a queue.[3] In our environment, individuals waiting in the market impose a negative externality on those who follow them in the queue, as they force them to wait a longer period and potentially miss desirable matches. On the other hand, waiting in the market generates a positive externality on agents on the other side of the market, as they are more likely to get a quicker match with a partner they prefer. As it turns out, the negative externality of waiting, which is not internalized by individuals, overwhelms this positive externality and leads to excessive waiting in the decentralized setting. In fact, the matching protocol induced by equilibrium in the decentralized matching process ends up resembling the protocol corresponding to the optimal mechanism, but with higher thresholds for the queues' lengths.

We evaluate the difference in welfare generated by a centralized and a decentralized market as a function of the underlying primitives of the environment, namely the agents' type distribution and the cost of waiting.

With respect to the type distribution, as the frequency of desirable types on either side of the market increases, the option value of waiting becomes higher and the wedge between the performance of the centralized and decentralized processes grows.

The comparative statics with respect to costs of waiting are more subtle. An increase in the cost of waiting has a direct and indirect effect. The direct effect is due to the longer expected queues in the decentralized setting. Fixing the expected queue lengths corresponding to the optimal and decentralized processes, an increase in per-period waiting costs on the welfare differential is effectively a multiplier effect – it is the difference between the expected time agents wait in queue under these two processes, multiplied by the change in costs. The indirect effect is that both the optimal threshold as well as the equilibrium threshold in the decentralized process decrease as a function of waiting costs. Since in the decentralized setting higher costs reduce the incentives to wait, this indirect effect works to mute the welfare

[2] This process is reminiscent of a double auction, as each agent submits a "demand function" specifying which types of agents she would be interested in matching with immediately.

[3] We provide a characterization of preferences that assure that the process is individually rational for all participants.

difference between the two processes. We show that the combination of these effects leads to a welfare wedge that is locally increasing in costs (formally, it is piece-wise increasing), but exhibits a general decreasing trend. Ultimately, when costs are prohibitively high, both processes lead to instantaneous matches and identical welfare levels.

Finally, we ask in which ways one can improve upon decentralization with interventions that are simpler than the full-fledged optimal mechanism. Indeed, centralization may be hard to implement for two main reasons: first, it requires the central planner to be able to force matches upon individuals and, second, it requires the central planner to monitor arrivals and to possibly create matches at every period, yielding potentially high administrative costs. To address the first issue, we consider a decentralized setting in which per-period taxes are introduced for the agents that decide to wait. Our characterization of the optimal mechanism allows us to identify a budget-balanced tax scheme that implements the optimal welfare levels. Such a tax scheme can be tailored so that it does not distort agents' incentives to enter the market to begin with. Nonetheless, even such a scheme may be viewed as cumbersome administratively since it requires continuous monitoring of agents' location in the queue. To address this problem, we also analyze the performance of a simple mechanism in which all matches occur at fixed time intervals. The length of these time intervals can be chosen to balance the costs of waiting and the quality of the resulting matches, which we characterize. We show that such a simple procedure, while still inferior to the fully optimal mechanism, can improve welfare substantially relative to full decentralization.

1.2 Related Literature

The interest surrounding dynamic matching is recent and the literature on this topic is still relatively limited. Much of this literature originally stemmed from the organ donation application. Zenios (1999) develops a queueing model to explain the differences between waiting times of different categories of patients anticipating a kidney transplant. In the context of kidney exchange, Ünver (2010) focuses on a market in which donors and recipients arrive stochastically, preferences are compatibility-based, and the goal of a central planner is to minimize total discounted waiting costs. Under some conditions, he shows that the efficient two-way matching mechanism always carries out compatible bilateral matches as soon as they become available. However, when multi-way matches are possible, some two-way matches

could be withheld in order to allow future multi-way matches.[4]

Akbarpour, Li, and Oveis Gharan (2014) is possibly the closest to our study as they also inspect the benefits of different mechanisms in a dynamic two-sided matching environment. In their setting, however, preferences are compatibility-based according to a network mapping the set of exchange possibilities. Agents in the system (thought of as patient-donor pairs) become "critical" at random dates, and perish immediately after if they remain unmatched. Therefore, when waiting costs are negligible, the goal of the planner is to minimize the number of perished agents. Market thickness is beneficial in that it guarantees the availability of immediate matches for agents who become critical. Left to their own devices, agents in that setting would match quickly and useful mechanisms induce agents to wait. In contrast, in our setting, agents in a decentralized process wait too long and useful mechanisms induce agents to wait shorter times on the market. In addition, while the welfare benchmark in Akbarpour, Li, and Oveis Gharan (2014) is that of an omniscient planner, our different focus allows us to characterize the optimal mechanism, which serves as the benchmark for welfare comparisons.

In a somewhat different realm, Leshno (2014) studies a one-sided market in which objects (say, public houses) need to be allocated to agents who wait in a queue. Welfare maximization always requires agents to be matched to their preferred objects. However, if agents' preferences are unknown to the planner and their preferred item is in short supply, agents may prefer a mismatched item earlier to avoid costly waiting. Leshno (2014) shows that the welfare loss from mismatches can be reduced substantially through a policy under which all agents who decline a mismatched item face the same expected wait for their preferred item.[5] Anderson, Ashlagi, Gamarnik, and Kanoria (2014) study an environment in which each agent is endowed with an item that can be exchanged with an item owned by someone else. Compatibility is stochastic, and three classes of feasible exchanges are considered: two-way exchanges only, two- and three-way cycles, and any kind of chain. They find that a policy that maximizes immediate exchanges without withholding them in the interest of market thickness performs

[4]Some recent models in inventory control have a similar flavor to the compatibility-based matching process considered by Ünver (2010), see e.g. Gurvich and Ward (2014).

[5]Bloch and Cantala (2014) also study dynamic allocations. In their setting, a mechanism is a probability distribution over all priority orders consistent with the waiting list. Given a priority order, whenever a new object becomes available, agents are proposed the object in sequence and can either accept or reject. They show the benefits of a strict seniority order.

nearly optimally.[6,7]

There is also a recent theoretical literature that studies decentralized matching processes that are dynamic, considering both informational and time frictions (see, e.g., Haeringer and Wooders, 2011, Niederle and Yariv, 2009, and Pais, 2008). In that literature, the number of agents on each side of the market is fixed at the outset and agents on one side can make directed offers to agents on the other side. The main goal is the identification of market features that guarantee that an equilibrium of the induced game generates a stable matching.

Another related strand of literature is the search and matching literature (e.g., Burdett and Coles, 1997, Eeckhout, 1999, and the survey by Rogerson, Shimer, and Wright, 2005). There, each period, workers and firms randomly encounter each other, observe the resulting match utilities, and decide jointly whether to pursue the match and leave the market or to separate and wait for future periods. As time frictions vanish, such markets generate outcomes that are close to a stable matching. A crucial difference with our setting is the stationarity of the market – the perceived distribution of potential partners does not change with time, and each side of the market solves an option value problem.[8]

Last, there is a rather large literature that considers dynamic matching of buyers and sellers and inspects protocols that increase efficiency or allow for Walrasian equilibrium outcomes to emerge as agents become increasingly patient (see, e.g., Satterthwaite and Shneyerov, 2007 and Taylor, 1995).[9]

[6] On the benefits of algorithms that create thicker pools in sparse dynamic allocation environments, see also Ashlagi, Jaillette, and Manshadi (2013).

[7] While most of this literature has been focusing on the design of algorithms to achieve socially desirable matchings Doval (2015) introduces a notion of stability in dynamic environments, and provides conditions under which dynamically stable allocations exist.

[8] In the context of marriage markets, Kocer (2014) considers learning over time and models the choice of temporary interactions with different potential partners as a multi-armed bandit problem. Choo (2015) develops a new model for empirically analyzing dynamic matching in the marriage market and applies that model to recent changes in the patterns of US marriages.

[9] Related, Budish, Cramton, and Shim (2015) consider financial exchanges and argue that high-frequency trading leads to inefficiencies, while frequent batch auctions, uniform-price double auctions that occur at fixed and small time intervals, can provide efficiency improvements. These results are reminiscent of our observations regarding the welfare improvements generated by matchings that occur at the end of each fixed window of time.

2 Setup

We study an infinite-horizon dynamic matching market. There are two kinds of agents: squares and rounds. Squares and rounds can stand for potential adoptive parents and children relinquished for adoption, workers and employers, patients and donors, etc.[10]

At each time t, one square and one round arrive at the market. Each square can be of either type A or B with probability p or $1-p$, respectively, and each round can be of type α or β with probability p or $1-p$, respectively. These types correspond to the attributes of participants – they can stand for the wealth of parents and race of children in the adoption application, level of education of employees and social benefits or promotion likelihoods for employers in labor markets[11], age or tissue types in the organ donation context[12], etc.

In our model, squares seek to match with rounds and vice versa. We denote by $U_x(y)$ the surplus for a type-x participant from matching with a type-y participant. We assume that preferences are assortative: A-squares are more desirable for all rounds and that α-rounds are more desirable for all squares. That is,

$$\begin{aligned} U_A(\alpha) &> U_A(\beta), \qquad U_B(\alpha) > U_B(\beta), \\ U_\alpha(A) &> U_\alpha(B), \qquad U_\beta(A) > U_\beta(B). \end{aligned}$$

It will be convenient to denote:

$$\begin{aligned} U_{A\alpha} &\equiv U_A(\alpha) + U_\alpha(A), \qquad U_{A\beta} \equiv U_A(\beta) + U_\beta(A), \\ U_{B\alpha} &\equiv U_B(\alpha) + U_\alpha(B), \qquad U_{B\beta} \equiv U_B(\beta) + U_\beta(B), \end{aligned}$$

as well as

$$U \equiv U_{A\alpha} + U_{B\beta} - U_{A\beta} - U_{B\alpha}.$$

[10] The organ donation application shares some features with our model when considering live donations from good samaritan donors, in which case the matching process is two-sided in nature.

[11] In some markets, wages differ across individual employees and can be thought of as transfers, which this paper does not handle. However, Hall and Kruger (2012) suggest that a large fraction of jobs have posted wages. Naturally, these wages may reflect general equilibrium wages tailored to the precise composition of the market. Nonetheless, the fact that these wages are fairly constant and do not fluctuate dramatically suggests they may not respond to particular characteristics of individual employees. Our model speaks to this segment of the market.

[12] This is a simplified representation that aims at capturing heterogeneity in types and the quality of different matches it implies. For organ donation, patients and donors are often classified into coarse categories based on age. However, more than two tissue types are often considered. For example, for kidney transplantation, the medical community currently looks at six tissue types, called major histo-compatibility complex or HLA antigens. For an extension to richer type sets, see the Online Appendix.

We will further assume that $U > 0$ so that the utilitarian efficient matching in a static market creates the maximal number of (A, α) and (B, β) pairs. The value of U captures the efficiency gain from such an assortative matching relative to the anti-assortative matching. Notice that if we assumed fully symmetric utilities, $U_x(y) = U_y(x)$, $U > 0$ would be tantamount to assuming super-modular assortative preferences (a-la Becker, 1974) and U can be thought of as the degree of super-modularity preferences exhibit.

We assume that each square and round suffer a cost $c > 0$ for each period they spend on the market waiting to be matched. We also assume that agents leave the market only by matching. In Section 7 we provide bounds on the utility of agents from remaining unmatched that assures this assumption is consistent with individual rationality in the processes we analyze.[13]

Several assumptions merit discussion. We assume that preferences are super-modular and that waiting costs are identical for squares and rounds for presentation simplicity. These assumptions are common in the literature and, as we describe in Section 5, lead to a conservative comparison of the optimal and decentralized matching protocols.

The assumption that the distribution of types of rounds mirrors that of squares also simplifies our analysis substantially as will be seen. It essentially implies that if we drew a large population of rounds and squares, the realized distributions of types would be approximately balanced with high probability. This may be a fairly reasonable assumption for certain applications, such as organ donation. Indeed, the distribution of tissue types of donors and patients is arguably similar. Furthermore, the age of a donor is known to have a strong impact on the expected survival of a graft (see, e.g., Gjertson, 2004 and Oien et al., 2007) and younger recipients have been suggested as the natural recipients of higher-quality organs (see Stein, 2011). Our assumptions then fit a world in which both patients and donors are classified as "young" or "mature" and patients' and donors' age distributions are similar. Our assumption also approximately holds for certain attributes in the online dating world (see Hitch, Hortacsu, and Ariely, 2010). Nonetheless, it might be a rather harsh assumption for other applications. As it turns out, the techniques we introduce can be used were we to relax that assumption. We replicate some of our analysis for general asymmetric settings in the Online Appendix.[14]

The other strong assumption we make is that a pair of agents arrives at the market in each

[13] We show that individual rationality is guaranteed when all agents are acceptable and when any β-round receives a utility lower than $U_\beta(B) - \frac{p}{1-p}[U_\alpha(A) - U_\alpha(B)]$ when leaving the market unmatched (analogously for B-squares).

[14] As mentioned, we also discuss in the Online Appendix an extension to richer type sets.

period. The analysis would remain virtually identical were we to assume that pairs arrive at random times following, say, a Poisson distribution. Moreover, our results extend directly if we assume that each period a fixed number of square-round pairs (possibly greater than one) enter the market. However, the assumption that participants arrive in pairs is important for the techniques we use. This assumption assures that the market is balanced throughout the matching process. It is a reasonable assumption for some applications. For example, in the adoption process presumably potential adoptive parents and birth mothers make important decisions (whom to match with, whether to leave the process, etc.) at spaced-out intervals. Given the limited variability in the volume of entrants on a monthly basis, the assumption of balanced arrivals provides a decent approximation of reality. Allowing for different arrival rates on both sides of the market introduces new considerations as matching participants in thin markets, while still entailing low waiting costs, imposes now a loss in terms of both the quality of matches and the number of individuals matched. The Online Appendix offers a more thorough discussion of how the analysis might be extended to allow for different arrival rates of squares and rounds.

3 Optimal Dynamic Matching

3.1 The Matching Process

In this section we characterize the socially optimal dynamic mechanism. At each period, after a new square-round pair arrives, the social planner determines which matches to implement, if any, to maximize the aggregate social welfare.

At any time t, before a new square-round pair enters the market, a queue corresponds to a vector $(k_A, k_B, k_\alpha, k_\beta)$, where each entry is the stock of squares or rounds of a particular type waiting in line from the previous period.

An optimal mechanism would match an A-square and an α-round whenever both are available. Indeed, the only reason to maintain an (A, α) pair in the queue is in order to match the A-square to a β-round and the α-round to a B-square. However, because of our supermodularity assumption, such matches would be inferior to the (A, α) and (B, β) matches that would be feasible instead. Matching the (A, α) pair immediately therefore generates at least

as high ultimate match surplus and entails a lower waiting cost.[15] Similarly, when a B-square and a β-round are available, an optimal mechanism would match them immediately as well. Therefore, we have the following:

Lemma 1. *Any optimal mechanism requires (A, α) and (B, β) pairs to be matched as soon as they become available.*

The lemma implies that at any point in time an optimal dynamic mechanism entails queues of only A-squares and β-rounds, or only B-squares and α-rounds. That is, the queue can take the form of either $(k, 0, 0, k)$ or $(0, k, k, 0)$, for some $k \geq 0$.

The optimal dynamic mechanism is then identified by the maximal stock of A-squares (and β-rounds) and the maximal stock of α-rounds (and B-squares) that are kept waiting in queue. In the following proposition we characterize the structure of the optimal mechanism.

Proposition 1 (The Optimal Mechanism)**.** *An optimal dynamic mechanism is identified by a pair of thresholds $(\bar{k}_A, \bar{k}_\alpha) \in \mathbb{Z}_+$ such that*

1. *whenever more than $\bar{k}_A$ A-squares are present, $k_A - \bar{k}_A$ pairs of type (A, β) are matched immediately, and*

2. *whenever more than $\bar{k}_\alpha$ α-rounds are present, $k_\alpha - \bar{k}_\alpha$ pairs of type (B, α) are matched immediately.*

From the symmetry of our environment, an optimal mechanism corresponds to symmetric thresholds: $\bar{k} = \bar{k}_A = \bar{k}_\alpha$.[16]

A dynamic mechanism with symmetric thresholds $(\bar{k}, \bar{k})$ as defined in Proposition 1 is depicted in Figure 1, where $k_{A\alpha} = k_A - k_\alpha$ captures the difference between the length of the queue of A-squares and the length of the queue of α-rounds. We call $k_{A\alpha}$ the *(signed) length of the A-α queue.*

This process induces the following Markov chain. Let s_t denote the number of A-squares (or β-rounds) minus the number of α-rounds (or B-squares) – that is the value of $k_{A\alpha}$ – at the beginning of time t, before the arrival of a new square-round pair. If an (A, α) or a (B, β)

[15] As waiting costs are linear, the identity of those matched has no welfare implications.

[16] As mentioned, in the Online Appendix we work out the extension to asymmetric environments, where the two thresholds may differ.

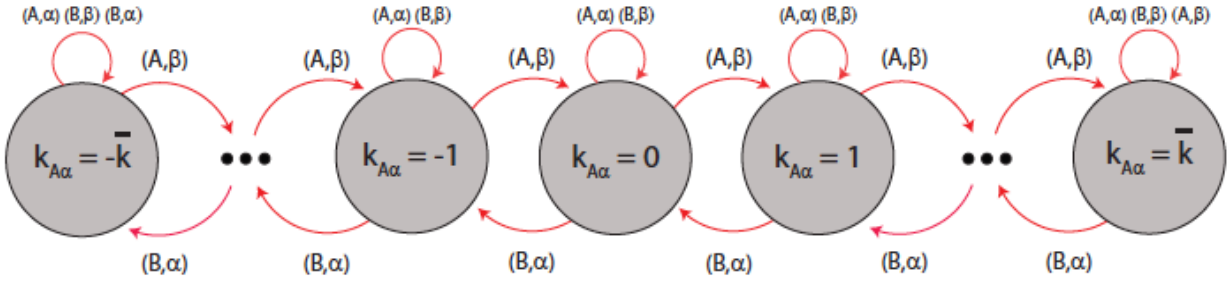

Figure 1: Structure of Optimal and Decentralized Matching Processes

pair arrive in period t, the mechanism matches an (A, α) or a (B, β) pair immediately, and the state remains the same: $s_t = s_{t+1}$. Suppose an (A, β) pair arrives in period t. As long as $0 \leq s_t < \bar{k}$, the mechanism creates no matches and s_{t+1} becomes $s_t + 1$. If $s_t < 0$, the mechanism creates one (A, α) match and one (B, β) match, and s_{t+1} becomes $s_t + 1$. Finally, if $s_t = \bar{k}$, the mechanism creates one (A, β) pair, and s_{t+1} remains the same, $s_{t+1} = s_t = \bar{k}$. Analogous transitions occur with the arrival of a (B, α) pair.

Therefore, we can describe the probabilistic transition as follows. Denote by

$$\mathbf{x}^t \equiv (x^t_{\bar{k}}, x^t_{\bar{k}-1}, ..., x^t_{-\bar{k}+1}, x^t_{-\bar{k}})^{tr} \in \{0,1\}^{2\bar{k}+1}$$

the timed vector capturing the state, $x^t_i = \mathbf{1}(s_t = i)$ – that is, x^t_i is an indicator that takes the value of 1 if the state is s_t and 0 otherwise. Then,

$$\mathbf{x}^{t+1} = \mathbf{T}_{\bar{k}} \mathbf{x}^t,$$

where

$$\mathbf{T}_{\bar{k}} = \begin{pmatrix} 1-p(1-p) & p(1-p) & \dots & 0 & 0 \\ p(1-p) & 1-2p(1-p) & \dots & 0 & 0 \\ 0 & p(1-p) & \dots & & \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ & & \dots & p(1-p) & 0 \\ 0 & 0 & \dots & 1-2p(1-p) & p(1-p) \\ 0 & 0 & \dots & p(1-p) & 1-p(1-p) \end{pmatrix}. \tag{1}$$

The above Markov chain is ergodic (i.e., irreducible, aperiodic, and positively recurrent). Therefore, an optimal mechanism corresponds to a matching process that reaches a steady state with a unique stationary distribution. For $\mathbf{T}_{\bar{k}}$, the steady-state distribution is uniform so that each state $s = \bar{k}, \bar{k}-1, \dots, -\bar{k}$ occurs with an equal probability of $\frac{1}{2\bar{k}+1}$.

3.2 Optimal Thresholds

In order to characterize the optimal threshold, we first evaluate the welfare corresponding to any arbitrary symmetric threshold.

First, we compute the average total waiting costs incurred by agents waiting in line for one period of time. During the transition from time t culminating at state s to time $t+1$, $2|s|$ agents wait in line. So the total costs of waiting incurred during this one time period is $2|s|c$. Thus, a mechanism with threshold $\bar{k}$ results in expected total costs of waiting equal to

$$\frac{1}{2\bar{k}+1}\left(\sum_{s=-\bar{k}}^{\bar{k}} 2|s|\right)c = \frac{2\bar{k}(\bar{k}+1)c}{2\bar{k}+1}.$$

Next we compute the average total surplus generated during one time period, tracking the Markov process described above. A new arrived pair is of type (A,α) with probability p^2, in which case the optimal mechanism generates a surplus equal to $U_{A\alpha}$. Similarly, when a new pair of type (B,β) arrives, which occurs with probability $(1-p)^2$, the optimal mechanism generates a surplus equal to $U_{B\beta}$. Suppose an (A,β) pair arrives. If $s_t < 0$, the mechanism creates one (A,α) and one (B,β) pair, generating a surplus equal to $U_{A\alpha}+U_{B\beta}$. If $0 \leq s_t < \bar{k}$, the mechanism creates no matches (and no additional surplus), and if $s_t = \bar{k}$, the mechanism creates one (A,β) pair and generates surplus equal to $U_{A\beta}$. Analogous conclusions pertain to the case in which a (B,α) pair arrives.

Thus, a mechanism with threshold $\bar{k}$ generates an expected total surplus equal to

$$\begin{aligned} & p^2U_{A\alpha} + (1-p)^2U_{B\beta} + \frac{2p(1-p)}{2\bar{k}+1}\left[\bar{k}\left(U_{A\alpha}+U_{B\beta}\right) + \frac{U_{A\beta}+U_{B\alpha}}{2}\right] \\ = \;& pU_{A\alpha} + (1-p)U_{B\beta} - \frac{p(1-p)U}{2\bar{k}+1}. \end{aligned}$$

Therefore, the net expected total welfare per period, accounting for waiting costs, is:

$$pU_{A\alpha} + (1-p)U_{B\beta} - \frac{p(1-p)U}{2\bar{k}+1} - \frac{2\bar{k}(\bar{k}+1)c}{2\bar{k}+1}. \tag{2}$$

When the cost of waiting c is high, waiting for potentially higher-quality matches does not justify any waiting and the mechanism matches agents instantaneously as they arrive. For sufficiently low costs, however, the optimal mechanism exhibits non-trivial waiting. The optimal threshold $\bar{k}^{opt}$ maximizes the welfare as given in (2).

The following proposition summarizes our discussion and provides the full characterization of the optimal mechanism.

Proposition 2 (Optimal Thresholds)**.** *The threshold*

$$\bar{k}^{opt} = \left\lfloor \sqrt{\frac{p(1-p)U}{2c}} \right\rfloor$$

identifies an optimal dynamic mechanism. In this optimal mechanism, all available (A, α) *and* (B, β) *pairs, and any number of* (A, β) *or* (B, α) *pairs exceeding* $\bar{k}^{opt}$*, are matched immediately. Furthermore, the optimal mechanism is generically unique.*[17]

The optimal threshold increases with the probability of any incongruent pair, $p(1-p)$, and with the degree of super-modularity U, which reflects the value of generating assortative matches. It decreases with waiting costs. In fact, when waiting costs are prohibitively high, namely when $c > \frac{p(1-p)U}{2}$, the maximal queue length is $\bar{k}^{opt} = 0$ and all matches are instantaneous.

3.3 Welfare

We now turn to the expected per-period welfare in the steady state under the optimal mechanism. Were we to consider no costs of waiting, the optimal mechanism would naturally entail long waits to get the maximal possible match surplus asymptotically by matching A-squares with α-rounds and B-squares with β-rounds. We denote the resulting welfare by:

$$S_\infty \equiv pU_{A\alpha} + (1-p)U_{B\beta}.$$

The optimal threshold identified in Proposition 2 allows us to characterize the welfare achieved by the optimal mechanism through equation (2) and to get the following corollary.[18]

Corollary 1 (Optimal Welfare)**.** *The welfare under the optimal mechanism is given by* $W^{opt}(c) = S_\infty - \Theta(c)$*, where* $\Theta(c)$ *is continuous, increasing, and concave in* c*,* $\lim_{c\to 0}\Theta(c) = 0$*, and* $\Theta(c) = p(1-p)U$ *for all* $c \geq \frac{p(1-p)U}{2}$*.*

[17] When $\sqrt{\frac{p(1-p)U}{2c}} \in \mathbb{Z}_+$, there are two optimal thresholds: $\sqrt{\frac{p(1-p)U}{2c}}$ and $\sqrt{\frac{p(1-p)U}{2c}} - 1$.

[18] In the Appendix, we provide the analytical formula for $\Theta(c)$ in terms of the fundamental parameters of our setting.

As waiting costs approach 0, the welfare induced by the optimal mechanism approaches S_∞. For costs large enough, the optimal mechanism matches all square-round pairs instantaneously as they arrive and the resulting welfare is $S_\infty - p(1-p)U$. For intermediate costs, the optimal mechanism generates welfare that is naturally in between these two values.[19,20] The observation that welfare under the optimal mechanism decreases as c increases is rather intuitive. Indeed, suppose $c_1 > c_2$. Were we to implement the optimal mechanism with waiting cost c_1 when the waiting cost is c_2, the distribution of matches would remain identical, while waiting costs would go down, thereby leading to greater welfare overall. This implies that the welfare under the optimal mechanism with waiting cost c_2, which would be at least weakly higher, is greater than that corresponding to waiting cost c_1. The amount by which welfare decreases when waiting costs increase depends on the number of agents expected to wait in line in the steady state. The higher the waiting costs, the lower the number of agents waiting in line on average. Therefore, the impact of an increase in costs by a fixed increment is greater at smaller costs, which leads to the concavity of $\Theta(c)$.

4 Decentralized Dynamic Matching

4.1 The Decentralized Process

Many dynamic matching processes are in essence decentralized: child adoption in the US and abroad, job searches in many industries, etc. It is therefore important to understand the implications of decentralized dynamics, particularly when considering centralized interventions.

In our decentralized matching process, we assume individuals join the market in sequence and decide when to match with a potential partner immediately and when to stay in the market and wait for a potentially superior match.

Formally, we assume that at each period t, there are three stages. First, a random square and round enter the market with random attributes as before: with probability p the square is an A-square and with probability p the round is an α-round. Second, each square and

[19] Notice that the value of S_∞ is effectively the analogue of the value generated by an "omnsicient" planner in our setting, which is used as one benchmark in Akbarpour, Li, and Oveis Gharan (2014). Corollary 1 suggests that the omniscient planner's value is a valid feasible benchmark when waiting costs vanish.

[20] In fact, simple algebraic manipulations imply that:

$$S_\infty - \sqrt{2p(1-p)Uc} - c \leq W^{opt}(c) \leq S_\infty - \sqrt{2p(1-p)Uc} + c.$$

round declare their demand – whether a square will match only with an α-round, only with a β-round, or any round and whether a round will match only with an A-square, only with a B-square, or with any square. In the third stage, given participants' demands, individuals are matched in order of arrival. That is, the process follows a first-in-first-out protocol.[21] Any remaining participants proceed to period $t+1$ (at an additional cost of c).[22] As before, the market effectively entails four possible queues of rounds and squares of all types and we denote by k_x the length of the queue corresponding to an x type of agent, where $x = A, B, \alpha, \beta$.

For simplicity, we will assume from now on a symmetric setting (results pertaining to asymmetric settings appear in the Online Appendix). That is, we assume:

$$U_A(\alpha) - U_A(\beta) = U_\alpha(A) - U_\alpha(B) \text{ and } U_B(\alpha) - U_B(\beta) = U_\beta(A) - U_\beta(B).$$

Last, we assume that the environment is *regular* in that

$$p\left(U_A(\alpha) - U_A(\beta)\right) \neq kc$$

for all natural numbers $k \in \mathbb{N}$. Regularity assures that neither squares nor rounds are ever indifferent between waiting in queue and matching immediately with an available partner.[23]

We focus on trembling-hand equilibria of the process. We do so to rule out situations in which a square and a round prefer to match with one another immediately but both choose to reject one another since such mutual rejection is robust to a unilateral deviation.[24]

4.2 Equilibrium Characterization

Given our assumptions on utilities, squares and rounds are always willing to match with α-rounds and A-squares, respectively, as soon as possible. Therefore, an A-square can always match with a round immediately (indeed, an equal number of squares and rounds arrive

[21] The assumption that matches are made in order of seniority in the market simplifies our analysis substantially (though we discuss an alternative ordering, that of last-in-first-out, in the Online Appendix). There is some anecdotal evidence that order of arrivals affects the sequence of matching in several markets. For instance, in international adoption many countries follow a first-in-first-out criterion to match abandoned children to adoptive parents. For example, see the protocol adopted by the China Center of Children's Welfare and Adoption (CCCWA) here: http://www.aacadoption.com/programs/china-program.html.

[22] The outcomes of the process we describe would be identical had we allowed participants to report no demand at a period and ruled out weakly dominated actions.

[23] The assumption of regularity is not crucial and similar analysis follows without regularity for any arbitrary tie-breaking rule. However, the presentation is far simpler for regular environments.

[24] Such behavior cannot be ruled out by simply eliminating weakly dominated strategies without further assumptions on the process.

at the market and an equal number of them exit by matching, so there is always a round available). If only β-rounds are available, depending on the status of the queue, an A-square may decide to wait in line for the chance of a superior match with an α-round later on or to match immediately with a β-round. Consequently, a β-round may wait in the queue either by choice, hoping to match with a willing A-square at later periods, or by necessity, when all squares in the queue are unavailable as they refuse to match with β-rounds. A similar description holds for α-rounds and B-squares.

4.2.1 Agents' Decision to Wait in Line

At the beginning of a period, the queue cannot entail both A-squares and α-rounds, as that implies an ideal match was not formed and waiting costs have been incurred at no benefit to some participants in the previous period. As before, we denote the (signed) length of the A-α queue at the beginning of a period by

$$k_{A\alpha} \equiv k_A - k_\alpha.$$

We first consider the decisions of an A-square (analogous analysis holds for an α-round). When an A-square arrives at the market and an α-round is available (one that had either been waiting in the queue or that has just arrived at the market as well), an A-square is matched immediately to an α-round, the identities of whom are prescribed by the order of arrival. If the arriving A-square is the first to arrive in line, that square is matched to an α-round. If there are A-squares already in queue, this implies that our A-square arrived with an α-round and the round will be matched with the first A-square in the queue. In that case, the newly arrived A-square has a choice of whether to wait in line or match with a β-round. However, notice that this square's decision is equivalent to the last A-square who had arrived and decided to wait. In that case, the new A-square waits and queues remain as they were.

Suppose now that an α-round is not available when an A-square enters the market. This implies that there is at least one β-round available (that the square can match with). The square then decides to wait in the queue based on the number of A-squares in the queue. An immediate match with a β-round delivers $U_A(\beta)$, whereas waiting in line till matching with an α-round delivers $U_A(\alpha)$ at an uncertain cost of waiting.

It is worth noting that once an A-square decides to wait in the queue (rather than match immediately with a β-round), she will wait until matching with an α-round (rather than leave

the queue by matching with a β-round at a later point). Indeed, as matches form on a first-in-first-out basis, her position in the queue moves up over time, and the expected time until matching with an α-round becomes shorter. That is, if it is optimal for her to wait in the queue upon entry, it is optimal for her to wait at any later period. The expected waiting time till a match with an α-round is therefore solely determined by the number of other A-squares who precede the square in the queue.

The following lemma identifies bounds on the size of the $k_{A\alpha}$ queue:

Lemma 2. *In all periods,* $-\bar{k}^{dec} \leq k_{A\alpha} \leq \bar{k}^{dec}$ *where*

$$\begin{aligned} \bar{k}^{dec} &\equiv \max\left\{k \in Z_+ \mid \frac{kc}{p} < U_A(\alpha) - U_A(\beta)\right\} \\ &= \max\left\{k \in Z_+ \mid \frac{kc}{p} < U_\alpha(A) - U_\alpha(B)\right\}. \end{aligned}$$

Intuitively, the time till an α-round enters the market is distributed geometrically (with parameter p). The expected time till an α-round arrives at the market is therefore $\frac{1}{p}$. An A-square who is k-th in line in the queue will be matched when k α-rounds arrive, which is expected to occur within $\frac{k}{p}$ periods. The expected waiting costs are therefore $\frac{kc}{p}$, which generate an increase in match utility of $U_A(\alpha) - U_A(\beta)$ (relative to matching with a β-round immediately). An A-square will wait as long as the expected benefit of waiting exceeds its costs, which is the comparison underlying the maximal size of the queue described in the lemma. Our regularity assumption further guarantees that an A-square or an α-round are never indifferent between waiting in line and matching immediately.

The lemma implies that A-squares' and α-rounds' behavior is essentially captured by the maximal queue length $\bar{k}^{dec}$. Whenever there are fewer than $\bar{k}^{dec}$ A-squares in queue, a new A-square will wait in the market; Whenever there are $\bar{k}^{dec}$ or more A-squares in the queue, that square will prefer to match with a β-round immediately. An analogous description holds for α-rounds and our symmetry assumptions assure that the maximal queue length is identical for A-squares and α-rounds.

We now turn to the decisions of B-squares and β-rounds. A β-round (similarly, a B-square) may decide to wait voluntarily by hoping to match with an A-square who will become available when the line for A-squares exceeds $\bar{k}^{dec}$. In principle, there are two effects at work. The first is similar to that experienced by the A-squares waiting in line: the longer the queue

of β-rounds waiting ahead in line, the longer the new β-round has to wait. The second effect, however, is due to A-squares' behavior in equilibrium: the longer is the queue, the closer are A-squares to reach the threshold $\bar{k}^{dec}$ and be matched with β-rounds. The next lemma illustrates that the sum of these two effects assures that B-squares and β-rounds cannot be found waiting simultaneously in equilibrium.

Lemma 3. *There can never be both a B-square and a β-round waiting in the market.*

To understand the intuition of Lemma 3, consider the first β-round arriving at the market. There cannot be other A-squares waiting in the market since any such squares would have arrived with β-rounds, in contradiction to our β-round being the first in line. Suppose that the β-round arrives with a B-square. To match with an A-square, he would have to wait for the arrival of at least $\bar{k}^{dec}+1$ A-squares, since the first $\bar{k}^{dec}$ A-squares will wait in line if not matched with an α-round. This implies that the cost of waiting for the β-round is at least $\frac{(\bar{k}^{dec}+1)c}{p}$, which is strictly greater than the cost $\frac{\bar{k}^{dec}c}{p}$ faced by the last A-square willing to wait in the market for her desirable match. From our super-modularity assumption, the benefit from waiting is

$$U_\beta(A) - U_\beta(B) < U_\alpha(A) - U_\alpha(B) = U_A(\alpha) - U_A(\beta).$$

Therefore, since an A-square is not willing to wait for an α-round when facing a queue of $\bar{k}^{dec}+1$, nor would the β-round. Similar calculations follow for longer queues of β-rounds (see Appendix). Given our focus on trembling-hand perfect equilibria, it follows that we can never have both β-rounds and B-squares stay in the market at the same time.

Lemmas 2 and 3 together imply that, in equilibrium, the decentralized process follows a protocol similar to that implemented by the optimal mechanism and depicted in Figure 1, though the thresholds governing them may be different. Indeed, whenever an A-square and an α-round, or a B-square and a β-round, enter the market together, a match is generated immediately (of either an (A,α) pair or a (B,β) pair, respectively) and the lengths of the queues do not change in that period. The length of the queues can increase or decrease only upon the arrival of an (A,β) pair or a (B,α) pair. As long as the queue (of either A-squares or α-rounds) is strictly shorter than its maximum of $\bar{k}^{dec}$, the arrival of a new incongruent pair can either increase or decrease the length of the queue by precisely one.

4.2.2 Steady State of Decentralized Matching

As for the optimal mechanism, the length of the A-α queue $k_{A\alpha}$ is characterized by a Markov chain. The transition matrix follows the description in Section 3.1. Formally, denote by $\mathbf{x}^t \equiv (x^t_{\bar{k}^{dec}}, x^t_{\bar{k}^{dec}-1}, ..., x^t_{-\bar{k}^{dec}+1}, x^t_{-\bar{k}^{dec}})^{tr} \in \{0,1\}^{2\bar{k}^{dec}+1}$ the state of the market in period t – that is, x^t_i is an indicator that takes the value of 1 if the (signed) queue length is precisely i and 0 otherwise. Then,

$$\mathbf{x}^{t+1} = \mathbf{T}_{\bar{k}^{dec}}\mathbf{x}^t,$$

where $\mathbf{T}_{\bar{k}^{dec}}$ is that described in (1). Since this Markov chain is ergodic, it is characterized by a unique stationary distribution, which is uniform across the $2\bar{k}^{dec}+1$ states.

The following proposition provides the characterization of the equilibrium steady state of the decentralized process.

Proposition 3 (Decentralized Steady State)**.** *There is a unique trembling-hand equilibrium under the decentralized process. This equilibrium is associated with a unique steady state distribution over queue lengths, such that the length of the A-α queue $k_{A\alpha} = k_A - k_\alpha$ is uniformly distributed over $\{-\bar{k}^{dec}, -\bar{k}^{dec}+1, \ldots, \bar{k}^{dec}\}$, and in any period, the queues contain either*

1. *$k_{A\alpha}$ A-squares and β-rounds, and no α-rounds and B-squares, or*
2. *$|k_{A\alpha}|$ α-rounds and B-squares, and no A-squares and β-rounds.*

The threshold $\bar{k}^{dec}$ is determined by the decisions of A-squares and α-rounds to wait, as specified in Lemma 2. The crucial difference between the decentralized and optimal mechanism is the threshold placed on the maximal stock of A-squares or α-rounds in waiting. Notice that a decision to wait in the market by, say, a square imposes a negative externality on succeeding squares, as it potentially affects their waiting time, and possibly the quality of their matches.[25] On the other hand, a decision to wait may impose a positive externality on future desirable agents on the other side of the market, who would face a ready desirable agent upon arrival. As it turns out, on net, externalities are negative and agents wait "too much." Recalling that we

[25] From a welfare perspective, the externality on the quality of the match is of less importance. As long as a social planner views identical agents as interchangeable, an immediate mismatch or a later mismatch have similar welfare consequences.

denoted by $\bar{k}^{opt}$ the threshold characterizing the optimal mechanism (specified in Proposition 2), we have the following corollary:

Corollary 2 (Threshold Comparison)**.** *Maximal waiting queues are longer in the decentralized process than they are under the optimal mechanism. That is,* $\bar{k}^{opt} \leq \bar{k}^{dec}$.

4.3 Welfare

Since the protocols are similar except for the queues' thresholds, the expected per-period welfare in the steady state characterized in Proposition 3 can be found using an analogous derivation to that carried out for the optimal mechanism. This derivation leads to an expression mirroring equation (2), accounting for the decentralized process' threshold $\bar{k}^{dec}$. That is, the expected per-period net welfare is given by:

$$W^{dec}(c) = S_\infty - \frac{p(1-p)U}{2\bar{k}^{dec}+1} - \frac{2\bar{k}^{dec}(\bar{k}^{dec}+1)c}{2\bar{k}^{dec}+1},$$

where, from Lemma 2,

$$\bar{k}^{dec} = \left\lfloor \frac{p\left(U_A(\alpha) - U_A(\beta)\right)}{c} \right\rfloor = \left\lfloor \frac{p\left(U_\alpha(A) - U_\alpha(B)\right)}{c} \right\rfloor. \tag{3}$$

To summarize, we have the following corollary:

Corollary 3 (Decentralized Welfare)**.** *The welfare under the decentralized process is given by* $W^{dec}(c) = S_\infty - \Psi(c)$*, where* $\lim_{c\to 0}\Psi(c) = p\left(U_A(\alpha) - U_A(\beta)\right)$*, and* $\Psi(c) = p(1-p)U$ *for all* $c \geq p\left(U_A(\alpha) - U_A(\beta)\right)$.

Recall Corollary 1, which characterized the welfare under the optimal mechanism. By definition, the welfare generated under the optimal mechanism is higher than that generated by the decentralized process, so that $\Theta(c) \leq \Psi(c)$. While the optimal mechanism generates welfare that is decreasing in waiting costs, this is not necessarily the case under the decentralized process. Furthermore, while the welfare under the optimal mechanism approaches S_∞ as waiting costs diminish, this is not the case under the decentralized process. As waiting costs become very small, there is a race between two forces. For any given threshold, the overall waiting costs decline. However, in equilibrium, decentralized thresholds increase, leading to greater expected wait times. As it turns out, the balance between these two forces generates significant welfare losses, given by $p\left(U_A(\alpha) - U_A(\beta)\right)$, even for vanishingly small costs. The next section provides a detailed comparison of the two procedures in terms of welfare.

5 Welfare Comparisons

By construction, the optimal mechanism generates welfare that is at least as high as that generated by the decentralized process. In this section we inspect how the welfare wedge responds to the underlying parameters of the environment, suggesting the settings in which centralized intervention might be particularly useful.

The following proposition captures the effect on the welfare wedge $W^{opt}(c) - W^{dec}(c)$ of both waiting costs c, the frequency p of A-squares or α-rounds, and the utility benefit for an α-round from matching with an A-square rather than a B-square (equivalently, the utility benefit for an A-square from matching with an α-round rather than a β-round).

Proposition 4 (Welfare Wedge – Comparative Statics)**.**

1. *For any interval* $[\underline{c}, \overline{c})$*, where* $\underline{c} > 0$*, there is a partition* $\{[c_i, c_{i+1})\}_{i=1}^{M-1}$*, where* $\underline{c} = c_1 < c_2 < ... < c_M = \overline{c}$*, such that* $W^{opt}(c) - W^{dec}(c)$ *is continuous and increasing over* (c_i, c_{i+1}) *and*

$$W^{opt}(c_i) - W^{dec}(c_i) > W^{opt}(c_{i+1}) - W^{dec}(c_{i+1})$$

   for all $i = 1, ..., M-1$*.*

2. *As* c *becomes vanishingly small, the welfare gap* $W^{opt}(c) - W^{dec}(c)$ *converges to a value that is increasing in* $p \in [0, 1)$ *and in* $U_\alpha(A) - U_\alpha(B)$*.*

To see the intuition for the comparative statics corresponding to waiting costs, notice that an increase in costs has two effects on the welfare gap. Since the equilibrium threshold under the decentralized process is greater than the optimal threshold (Corollary 2), an increase in waiting costs has a direct effect of magnifying the welfare gap. Nonetheless, there is also an indirect effect of an increase in waiting costs that arises from the potential changes in the induced thresholds. Consider a slight increase in waiting costs such that the optimal threshold does not change, but the decentralized threshold decreases. The decentralized process is then "closer" to the optimal process – both the matching surplus and the waiting costs are closer and the welfare gap decreases. In fact, as costs become prohibitively high, both processes lead to instantaneous matches and identical welfare levels. As we show in the proof of Proposition 4, the indirect effect overwhelms the direct effect at precisely such transition points and acts to shrink the welfare gap. The construction of the partition is done as follows. Each

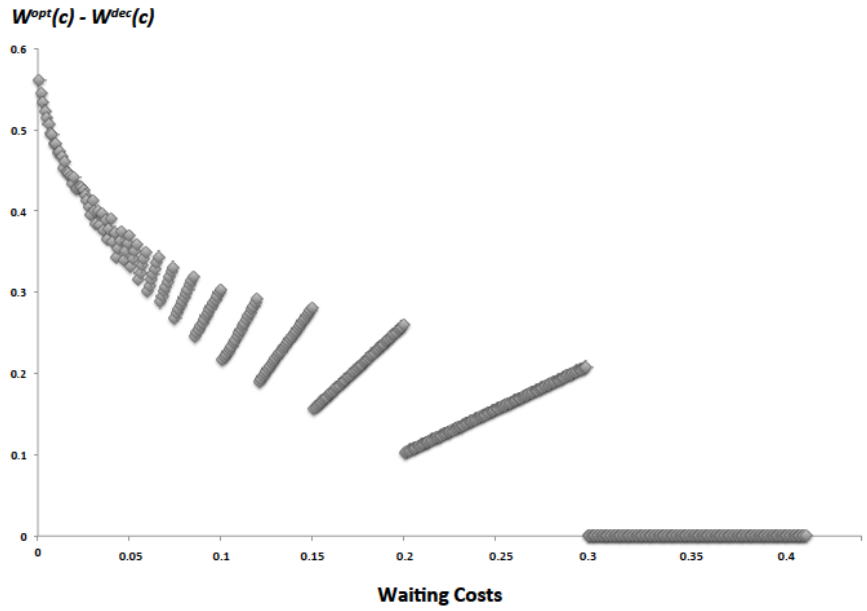

Figure 2: Welfare Gap between Optimal and Decentralized Matching as a Function of Costs

atom $[c_i, c_{i+1})$ of the partition corresponds to constant thresholds under the decentralized process. Over these intervals, the direct effect described above dominates and the welfare gap is increasing. Each of the endpoints $\{c_i\}_i$ corresponds to a decrease of the decentralized threshold by one. That is, the decentralized threshold is not constant over $[c_i, c_{i+2})$ for all i. Therefore, when comparing two such endpoints, the indirect effect kicks in and the decreasing trend of the welfare gap emerges. Figure 2 depicts the resulting pattern the welfare gap exhibits for $U_A(\alpha) = U_\alpha(A) = 3$, $U_A(\beta) = U_\alpha(B) = U_B(\alpha) = U_\beta(A) = 1$, $U_B(\beta) = U_\beta(B) = 0$, and $p = 0.3$. As suggested by the proposition, $W^{opt}(c) - W^{dec}(c)$ is piece-wise increasing in c. Nevertheless, overall, the gap has a decreasing trend.

To glean some intuition on the comparative statics the welfare gap displays with respect to p, consider two type distributions governed by p_1 and p_2 such that $p_1 < p_2 = mp_1$, $m > 1$. Suppose further that costs are so low that the match surplus in the decentralized process is very close to the optimum, S_∞. The individual incentives to wait are higher under p_2 than under p_1. In fact, in the decentralized setting, the distribution of steady state queue length is the uniform distribution where, from (3), under p_2, roughly $1 - 1/m$ of the probability mass is allocated to queue lengths larger than those realized under p_1. For each of these large steady state queue lengths, we have more pairs of agents waiting, i.e., increased per-period

waiting costs. The optimal mechanism internalizes the negative externalities, so the effect of the increased waiting costs is weaker. On the other hand, the benefit of this increase in queue length is a lower chance of producing mismatches. However, for sufficiently low c, the match surplus under p_1 is already close to its optimum of S_∞ and this effect is weak; in particular, the difference in terms of match surplus that the optimal and decentralized processes generate is similar under p_1 and p_2. Therefore, for sufficiently low c, the dominant effect is the one produced by the differential in expected waiting costs, which generates our comparative statics.[26] Notice that if $p = 0$ or $p = 1$, both the optimal mechanism and the decentralized processes generate the same welfare.

Going back to our assumption of super-modular preferences, note that the construction of the optimal mechanism would remain essentially identical were preferences sub-modular (with an appropriate relabeling of market participants). However, in the decentralized setting, sub-modular preferences would lead to a negative welfare effect compounding the negative externalities present in our setup. Namely, individual incentives would be misaligned with market-wide ones. In that respect, our comparison of optimal and decentralized processes assuming super-modular preferences is a rather conservative one.

Similarly, considering waiting costs that differ across the two sides of the market, would lead to a greater welfare wedge as well. Intuitively, suppose that squares experience a waiting cost of c_S and rounds experience a waiting cost of c_R, where $c_S > c_R$, with an average cost of $c = (c_S + c_R)/2$. The optimal mechanism with asymmetric costs would coincide with that corresponding to identical costs of c since per-pair costs are the same in both cases. In the decentralized process, A-squares would be willing to wait when the queue of A-squares is no longer than $\bar{k}_S^{dec}$ and α-rounds would be willing to wait when the queue of α-rounds is no longer than $\bar{k}_R^{dec}$, where

$$\bar{k}_S^{dec} = \left\lfloor \frac{p\left(U_A(\alpha) - U_A(\beta)\right)}{c_S} \right\rfloor \quad \text{and} \quad \bar{k}_R^{dec} = \left\lfloor \frac{p\left(U_\alpha(A) - U_\alpha(B)\right)}{c_R} \right\rfloor.$$

Suppose $\frac{p(U_A(\alpha)-U_A(\beta))}{c_x} \in \mathbb{N}$ for $x = S, R$ to avoid rounding issues. From convexity, it follows that the threshold $\bar{k}^{dec}$ corresponding to identical costs of c satisfies $\bar{k}^{dec} \leq (\bar{k}_S^{dec} + \bar{k}_R^{dec})/2$.

[26] In fact, we can show that for any $\Delta p > 0$, there exists $\delta > 0$ such that for every $c < \delta$ and $p \in [0, 1 - \Delta p]$, we have that the welfare wedge under $p + \Delta p$ and c is greater than under p and c. Furthermore, $\delta \to 0$ as $\Delta p \to 0$.

Therefore, the excessive waiting decentralized processes exhibit would be even more pronounced when costs are asymmetric across market sides.[27]

6 Transfers and Fixed Matching Windows

So far, we have shown that intervention in dynamic matching markets can have a substantial impact on welfare, at least when centralization is carried out using the optimal dynamic mechanism. However, the full-fledged optimal mechanism may be hard to implement. It requires that the formation of matches, even those of individuals who would prefer to wait in the market, be within the purview of the centralized planner. It also requires the central planner to monitor the market continuously to determine when matches should be formed, which may be administratively costly. In this section we show that improvements to decentralization can be achieved by mechanisms that relax each of these two requirements. We start by showing that a tax scheme imposed on those who wait in the market can yield the optimal waiting patterns, if tailored appropriately. We then analyze a simple centralized matching protocol that, while not optimal, can provide substantial welfare improvements over the decentralized matching process, and does not require the clearinghouse to monitor the market continuously. Namely, we consider matching processes in which the centralized clearinghouse matches all available agents every fixed number of periods.

6.1 Optimal Taxation

Certainly, a fixed per-period tax on waiting $\tau \geq 0$ can be set so that the resulting de-facto cost of waiting, $c + \tau$, is such that the decentralized process generates the optimal threshold (with costs c).[28] Namely, τ can be set so that

$$\frac{p\left(U_A(\alpha) - U_A(\beta)\right)}{c+\tau} = \sqrt{\frac{p(1-p)U}{2c}}. \tag{4}$$

In fact, if collected taxes in period t are given to new entrants in period $t+1$, they would have no effects on either overall welfare or individual incentives. Nonetheless, there is a risk

[27] In fact, in such a setting, β-rounds may wait in the market even when B-squares are present, leading to another channel of inefficient waiting. We show formally how our main results carry over to this more general environment in the Online Appendix.

[28] This is always possible from Corollary 2.

that such a policy would introduce a strong incentive for agents to enter the market only in order to gather the previous generation's taxes. As it turns out, there is a tax scheme that is budget balanced such that the expected tax (or subsidy) for an entering agent who is not privy to the pattern of queues in place is nil. Under such a scheme, no agent is tempted to enter the market only for the sake of reaping the benefits of taxes.

Indeed, consider a linear tax scheme – agents who are k-th in line pay $\tau^* k$ to the matching institution, regardless of whether they are a square or a round and regardless of their type. To achieve budget balance, the collected taxes are equally redistributed back to the existing agents in the market in each period.

When the length of queue is $\hat{k}$, the resulting net tax (that is added to the fixed cost c) for an agent who is k-th in line is

$$\tau^* k - \frac{2 \cdot \sum_{k=1}^{\hat{k}} \tau^* k}{2\hat{k}} = \tau^* \left(k - \frac{\hat{k}+1}{2} \right).$$

In particular, the net added tax on waiting for the last agent in the queue is $\frac{\tau(\hat{k}-1)}{2}$, which is increasing in the queue's length $\hat{k}$. Thus, using the definition of τ from equation (4), we want the tax levied on the last agents in the queue of length $\bar{k}^{opt}$ to satisfy

$$\frac{\tau^*(\bar{k}^{opt}-1)}{2} = \tau \quad \iff \quad \tau^* = \frac{2\tau}{\bar{k}^{opt}-1}.$$

Such a taxation policy might still be difficult to administer in terms of the financial activity it would entail – time-dependent taxes and redistribution of resources at each period. Without taxation, the optimal mechanism may be viewed as too complex administratively as well. Indeed, optimality dictates matchings that occur at points in time that are not perfectly predictable at the outset of the process and the clearinghouse needs to monitor the queues of individuals continuously. In the next subsection we offer the analysis of a simple mechanism that, while not optimal, can generate substantial welfare improvements relative to the decentralized process.

6.2 Matching with Fixed Windows

We now consider the class of mechanisms that are identified by a fixed-window size – every fixed number of periods, the efficient matching for the participants arriving at the market in that time window is formed and the market is cleared. Larger windows then allow for thicker

markets and potentially more efficient matchings. However, larger windows also correspond to longer waiting times for all participants. We identify the optimal fixed window. While the welfare it generates is still notably lower than that produced by the optimal mechanism, it can generate substantially greater welfare than that produced by the decentralized process.

The arrival of squares and rounds is the same as that analyzed so far. When fixed matching windows are used, a window size n governs the process. Namely, every n periods, the most efficient matching pertaining to the n squares and n rounds who arrived within that window is implemented.

Suppose that k_A and k_α are the numbers of A-squares and α-rounds who arrived during a window of n time periods, respectively. Given our assumptions on match utilities, efficient matchings correspond to a unique distribution of pair-types (number of (A,α) pairs, (A,β) pairs, etc.) and generate the maximal number of (A,α) and (B,β) pairs. The total matching surplus generated by a matching as such is

$$S(k_A,k_\alpha) \equiv \begin{cases} k_\alpha U_{A\alpha} + (k_A - k_\alpha)U_{A\beta} + (n-k_A)U_{B\beta} & \text{if } k_A \geq k_\alpha, \\ k_A U_{A\alpha} + (k_\alpha - k_A)U_{B\alpha} + (n-k_\alpha)U_{B\beta} & \text{otherwise.} \end{cases}$$

Consider now the expected waiting costs when the window size is n. The first square and round to arrive wait for $n-1$ periods, the second square and round to arrive wait for $n-2$ periods, etc. Thus, the total waiting cost is

$$2 \cdot c \cdot ((n-1) + (n-2) + \cdots + 0) = cn(n-1).$$

Therefore, the net expected welfare for each square-round pair generated by a window size n is

$$\begin{aligned} W_n \equiv\ & \frac{1}{n} \sum_{0 \leq k_\alpha, k_A \leq n} \binom{n}{k_A}\binom{n}{k_\alpha} p_A^{k_A}(1-p_A)^{n-k_A} p_\alpha^{k_\alpha}(1-p_\alpha)^{n-k_\alpha} S(k_A,k_\alpha) \\ & - c(n-1). \end{aligned}$$

Notice that for any window size n, the matching surplus per pair $S(k_A,k_\alpha)/n$ is at most $U_{A\alpha}$, while expected waiting costs per pair are $c(n-1)$. Denote by $n^{\max}$ the largest window size n such that $U_{A\alpha} \geq c(n-1)$. Every window of size $n > n^{\max}$ would then generate a lower welfare than that generated by a window of size 1, corresponding to instantaneously matching individuals. In particular, an optimal window size exists within the finite set $\{1,2,...,n^{\max}\}$.

6.2.1 Optimal Window Size

A characterization of the precise optimal window size is difficult to achieve. We now describe bounds on the optimal window size. These bounds land themselves to bounds on the maximal expected welfare generated by using fixed windows, which we can then compare to the welfare generated by the optimal mechanism as well as the decentralized process.

Consider the ex-ante marginal benefit produced by increasing the window size from n to $n+1$ that is incurred by the first n square-round pairs who arrive at the market. Suppose that an efficient matching corresponding to the first n pairs generates at least some mismatches (i.e., (A, β) or (B, α) pairs). The $(n+1)$-th pair could be beneficial for the first n pairs by correcting a mismatch. If there is no mismatch among the first n pairs, expanding the window size only leads to additional waiting costs for the first n pairs.

Denote the probability that any efficient matching with the first n squares and rounds has a mismatch (i.e., there is an unequal number of A-squares and α-rounds) by

$$Pr\left(|k_A - k_\alpha| \geq 1; n\right).$$

Since we assumed an identical distribution of types of squares and rounds, an efficient matching from n pairs has mismatches of type (A, β) or (B, α) with equal probability. A mismatch of type (A, β) is corrected by a new $(n+1)$-th pair of type (B, α), which occurs with probability $p(1-p)$, and the total benefit for the pair of originally mismatched square and round is

$$U_A(\alpha) + U_\beta(B) - U_A(\beta) - U_\beta(A).$$

A similar derivation follows for the benefit of "correcting" a mismatched pair of type (B, α). Conditional on any efficient matching entailing a mismatch, the expected benefit to the first n square-round pairs is then

$$\begin{aligned} & \frac{p(1-p)}{2}\left(U_A(\alpha) + U_\beta(B) - U_A(\beta) - U_\beta(A)\right) \\ + \quad & \frac{p(1-p)}{2}\left(U_\alpha(A) + U_B(\beta) - U_\alpha(B) - U_B(\beta)\right), \end{aligned}$$

which is equal to

$$\frac{p(1-p)\cdot U}{2}.$$

Thus, the ex-ante marginal welfare from expanding the window size from n to $n+1$ for each of the n first square-round pairs is:

$$\Delta_+ W_n \equiv \frac{p(1-p)\cdot U}{2n} \cdot Pr\left(k_A \neq k_\alpha; n\right) - 2c.$$

The optimal window size n^o is then determined by

$$\Delta_+ W_{n^o} \leq 0 \leq \Delta_+ W_{n^o-1}. \tag{5}$$

That is, it is (weakly) beneficial to increase the window size from $n^o - 1$ to n^o, and it is not beneficial to increase the window size from n^o to $n^o + 1$.

While it is difficult to accomplish a closed-form solution for the optimal window size, the following proposition utilizes inequality (5) to establish bounds on n^o.[29]

Proposition 5 (Optimal Window Size)**.**

1. *An upper bound for the optimal window size is given by*

$$n^o \leq \frac{p(1-p)U}{4c}.$$

2. *For any $\varepsilon > 0$, there exists c^* such that if $c < c^*$, a lower bound for the optimal window size is given by*

$$n^o \geq \frac{p(1-p)U}{(4+\varepsilon)c}.$$

Recall that U captures the extent to which preferences exhibit super-modularity, the welfare advantage of an assortative matching relative to an anti-assortative one. Intuitively, the bounds on the optimal window size increase with U and decrease with the cost of waiting. To understand the dependence on $p(1-p)$, let m be the number of pairs in the market of type (A, β) or (B, α). The designer's aim is to achieve a small number of such mismatches. Clearly, when m is odd, the probability of at least one mismatch is one. Suppose m is even, $m = 2l$. The conditional probability of having a mismatch is then $1 - \binom{2l}{l}\left(\frac{1}{2}\right)^{2l}$ which is increasing in

[29] The difficulty stems from the fact that inequality (5) depends crucially on

$$Pr\left(k_A \neq k_\alpha; n\right) = 1 - \sum_{l=0}^{n}\left(\binom{n}{l}p^l(1-p)^{n-l}\right)^2,$$

and state-of-the-art combinatorics has little to say about this function's behavior with changes in n.

l. If we focus on large n, the probability of an even or an odd m are approximately equal. Furthermore, when $p(1-p)$ increases, there is a greater probability of (A,β) or (B,α) pairs and the expected value of m is higher. Fixing all else, a greater $p(1-p)$ would then intuitively lead to a lower match surplus. Thus, the optimal window size grows with $p(1-p)$.

Notice that Proposition 5 implies that the optimal window size is smaller than the threshold induced by the decentralized process, at least for sufficiently low costs. That is, $n^o < \bar{k}^{dec}$ for low waiting costs. In particular, the bounds we identify in Section 7 on the value of leaving the market unmatched that guarantee individual rationality under the decentralized process, also assure individual rationality under the fixed-window mechanism. It also follows that the choice of the optimal window size allows for some alleviation of the excess waiting the decentralized process exhibits. Indeed, as we show in the next subsection, the welfare generated by the optimal fixed window dominates that produced by the decentralized process.

6.2.2 Welfare Bounds for Fixed Windows

If the waiting cost c is small, so that the optimal window size is large, we expect to have an approximate fraction p of A-squares and α-rounds, in which case per-pair surplus is close to S_∞. The bounds on the optimal size of the window allow us to provide bounds on how far the match surplus is from S_∞ and how costly the wait is. The following proposition illustrates the resulting bounds on the welfare generated by the optimal window size.

Proposition 6 (Fixed-Window Welfare)**.**

1. *A lower bound on the optimal welfare is given by*

$$W^{fix}(c) \geq S_\infty - \sqrt{\frac{2c}{U}}(U_{A\alpha} - U_{B\beta}) - \frac{p(1-p)U}{2} + 2c.$$

2. *For any $\varepsilon > 0$, there exists c^* such that if $c < c^*$, an upper bound on the optimal welfare is given by*

$$W^{fix}(c) \leq S_\infty + \sqrt{\frac{(2+\varepsilon)c}{U}}(U_{A\alpha} - U_{B\beta}) - \frac{p(1-p)U}{2+\varepsilon} + 2c.$$

Notice that as the cost of waiting becomes very small, some of the terms in the above bounds vanish. Furthermore, the bounds converge to one another and we get:

$$\lim_{c\to 0} W^{fix}(c) = S_\infty - \frac{p(1-p)U}{2}.$$

Intuitively, $p(1-p)$ reflects the probability of an incongruent pair entering the market. As that probability grows, welfare decreases. The value of U reflects the value of efficiently matching individuals. Mismatches are more costly when U is higher, and so U affects the expected welfare negatively as well.

Proposition 6 allows us to compare the performance of the optimal fixed-window mechanism with the performance of the optimal mechanism and the decentralized process.

Corollaries 1 and 3 illustrate the limit values of $W^{opt}(c)$ and $W^{dec}(c)$ as costs become very small. These limits combined with Proposition 6 lead to the following corollary.

Corollary 4 (Relative Performance of Fixed Window Mechanisms)**.**

$$\lim_{c\to 0}\frac{W^{opt}(c)-W^{fix}(c)}{W^{opt}(c)-W^{dec}(c)}=\frac{(1-p)U}{2\left(U_A(\alpha)-U_A(\beta)\right)}.$$

Notice that the right-hand side of the equality in Corollary 4 is lower than 1. Therefore, the corollary suggests that the optimal fixed-window mechanism potentially provides a substantial improvement over the decentralized process. Whenever $1-p, U>0$, the right hand side of the equality in Corollary 4 is strictly positive, which implies that the welfare generated by the optimal fixed-window mechanism is still significantly lower than that generated by the optimal mechanism. As p decreases, or U increases, the welfare generated by the fixed-window mechanism provides a greater improvement over the decentralized process.

7 Individual Rationality

Throughout the paper we assumed that agents leave the market only after being matched. Agents do not leave the market unmatched, regardless of their expected utility. Certainly, if remaining unmatched generates zero utility, our assumption violates individual rationality as some agents (in particular, B-squares or β-rounds) sometimes stay in the market simply for lack of available agents who will match with them. These agents could consequently earn a negative expected utility even when $U_x(y)>0$ for all x, y. Assume that any agent prefers to match with any other agent immediately over remaining unmatched. We now provide a bound on the value of remaining unmatched, or the value of an outside option agents have, that assures all the matching protocols we discuss are individually rational. Notice that since,

from Corollary 2, the decentralized threshold is higher than the optimal threshold, it suffices to find such a bound that guarantees that the decentralized process is individually rational.

In the decentralized process, A-squares or α-rounds always have the possibility of matching with β-rounds or B-squares instantaneously when they decide to wait in the market. Therefore, when deciding to wait they expect an even greater utility and individual rationality holds for them. Consider now β-rounds (analogously, B-squares). A β-round who enters as k-th in line can always declare all squares as acceptable in each period. Notice that, by construction, $k \leq \bar{k}^{dec}$. The time between arrivals of B-squares is distributed geometrically with probability $1-p$. Therefore, with such a strategy, the expected time for the β-round to match with a B-square is at most $k/(1-p)$, yielding a match utility of $U_\beta(B)$. The wait time till matching with a B-square could be even shorter if β-rounds who precede the β-round in question are willing to match only with A-squares. Furthermore, the β-round could end up matching with an A-square before k B-squares arrive at the market. It follows that such a strategy guarantees an expected utility of at least

$$U_\beta(B) - \frac{kc}{1-p} \geq U_\beta(B) - \frac{\bar{k}^{dec}c}{1-p} \geq U_\beta(B) - \frac{p}{1-p}\left[U_\alpha(A) - U_\alpha(B)\right] \equiv U^{\min}.$$

If β-rounds follow a different strategy in equilibrium, it must be that their utility is at least as high. Therefore, as long as the value of remaining unmatched is lower than $U^{\min}$, individual rationality holds under both the optimal and the decentralized processes (analogous calculations follow for B-squares and, under full symmetry of utilities, the bound corresponding to them is also $U^{\min}$). In addition, since the optimal fixed-window size is smaller than the decentralized equilibrium threshold, the construction above assures that this bound on remaining unmatched assures individual rationality for the fixed-window protocol as well.

8 Conclusions

In this paper we considered a dynamic matching setting and identified the optimal matching mechanism in an environment as such. The optimal mechanism always matches congruent pairs immediately and holds on to a stock of incongruent pairs up to a certain threshold. When matching follows a decentralized process, a similar matching protocol ensues in equilibrium, but the induced thresholds for waiting in the market are larger as individuals do not

internalize the net negative externalities they impose on those who follow. This difference generates a potentially significant welfare wedge between decentralized processes and centralized clearinghouses, even when waiting costs are vanishingly small. Our results provide guidance as to the features of the economy that could make centralized intervention more appealing.

We also offer some simple interventions to decentralized markets – transfer schemes that can induce optimal outcomes, and fixed-window mechanisms, which are arguably far less complex than the full-fledged optimal mechanism and can provide substantial welfare improvements relative to decentralization.

There are several natural extensions that we analyze and discuss in the Online Appendix: general asymmetric environments, a richer set of participant types, different arrival processes, and interventions that manipulate the queuing protocol governing the decentralized process.

9 Appendix

9.1 Proofs Regarding the Optimal Mechanism

Proof of Proposition 1: By Lemma 1, under the optimal mechanism, at any point in time, either only A-squares and β-rounds, or only B-squares and α-rounds are waiting in the market.

Consider a queue that contains only A-squares and β-rounds. From the stationarity of the process, and since waiting costs are linear, if it is optimal to retain k (A, β) pairs in the market at some period, it is also optimal to retain any $k' \leq k$ such pairs at any other period. Let $\bar{k}_A$ be the largest number of (A, β) pairs that might be allowed to wait in line. Such a number clearly exists as the optimal mechanism cannot allow an indefinite number of (A, β) pairs to wait in line for one more period. The value of $\bar{k}_A$ corresponds to the threshold that the optimal mechanism uses: i.e., whenever the queue of A-squares (or β-rounds) gets longer than $\bar{k}_A$, the mechanism creates (A, β) matches such that exactly $\bar{k}_A$ number of A-squares (or β-rounds) wait for one more period. Similar arguments follow for queues composed of (B, α) pairs. ■

Proof of Proposition 2: Define the marginal increase of welfare when increasing the threshold by one as:

$$MW(\bar{k}) \equiv W(\bar{k}+1) - W(\bar{k}) = \frac{2\left[p(1-p)U - 2c(\bar{k}+1)^2\right]}{(2\bar{k}+1)(2\bar{k}+3)}.$$

For a non-trivial (i.e., non-zero) optimal threshold, it is necessary that $MW(0) > 0$, or equivalently $c < \frac{p(1-p)U}{2}$.

Suppose c is small enough so that this is the case. Ignoring the integer constraint on $\bar{k}$,

$$\frac{d^2W(\bar{k})}{d\bar{k}^2} = \frac{1}{(2\bar{k}+1)^3}\left(\frac{2c}{p(1-p)} - 4U\right) < 0 \quad \text{as} \quad 2p(1-p)U > c.$$

Thus, accounting for the optimal threshold being an integer, there exists either a unique optimal threshold or two adjacent optimal thresholds.

An optimal threshold $\bar{k}^{opt}$ satisfies either

$$\begin{array}{lll} MW(\bar{k}^{opt}-1) > 0 & \text{and} & MW(\bar{k}^{opt}) \leq 0, \quad \text{or} \\ MW(\bar{k}^{opt}-1) \geq 0 & \text{and} & MW(\bar{k}^{opt}) < 0. \end{array}$$

We rewrite the two conditions as

$$\begin{array}{l} \sqrt{\frac{p(1-p)U}{2c}} - 1 \leq \bar{k}^{opt} < \sqrt{\frac{p(1-p)U}{2c}}, \quad \text{or} \\ \sqrt{\frac{p(1-p)U}{2c}} - 1 < \bar{k}^{opt} \leq \sqrt{\frac{p(1-p)U}{2c}}, \end{array}$$

and the Proposition's claim follows. ■

Proof of Corollary 1: Using the optimal thresholds from Proposition 2, we get that for $c \leq \frac{p(1-p)U}{2}$,

$$v(c) \equiv \frac{p(1-p)U}{2\bar{k}^{opt}+1} = \frac{p(1-p)U}{2\left\lfloor\sqrt{\frac{p(1-p)U}{2c}}\right\rfloor + 1}, \quad \text{and}$$

$$\begin{aligned} w(c) &\equiv \frac{2\bar{k}^{opt}(\bar{k}^{opt}+1)}{2\bar{k}^{opt}+1}c = \frac{(2\bar{k}^{opt}+1)c}{2} - \frac{c}{2(2\bar{k}^{opt}+1)} = \\ &= \left[\left(\left\lfloor\sqrt{\frac{p(1-p)U}{2c}}\right\rfloor + \frac{1}{2}\right) - \frac{1}{4\left\lfloor\sqrt{\frac{p(1-p)U}{2c}}\right\rfloor + 2}\right]c. \end{aligned}$$

We can then define $\Theta(c) = v(c) + w(c)$ to get the representation of $W^{opt}(c)$ in the corollary.

Take any $c < \frac{p(1-p)U}{2}$ for which $\bar{k}^{opt} \notin \mathbb{Z}_+$. There exists $\varepsilon > 0$ such that for every c' with $|c-c'| < \epsilon$, $\bar{k}^{opt}(c') = \bar{k}^{opt}(c)$.[30] Thus, W^{opt} is differentiable at c. Moreover, for any $c < \frac{p(1-p)U}{2}$ such that $\bar{k}^{opt} \in \mathbb{Z}_+$, $W^{opt}(c)$ is semi-differentiable. Hence, $W^{opt}(c)$ is continuous.

[30] We slightly abuse our notation and make the dependence of $\bar{k}^{opt}$ on the cost c explicit here.

At any differentiable point c (around which $\bar{k}^{opt}(c)$ is constant),

$$\frac{dW^{opt}}{dc} = \frac{\partial W(\bar{k}^{opt}(c), c)}{\partial c} = -\frac{2\bar{k}^{opt}(\bar{k}^{opt}+1)}{2\bar{k}^{opt}+1} < 0.$$

Furthermore, convexity of $W^{opt}(c)$ follows from the fact that at any semi-differentiable but not differentiable point c,

$$\frac{d_{-}W^{opt}}{dc} < \frac{d_{+}W^{opt}}{dc}.$$

■

9.2 Proofs Regarding Decentralized Processes

Proof of Lemma 2: The time till an α-round appears in the market is distributed geometrically, and so the expected time for arrival of any α-round is given by $\frac{1}{p}$. Consider an A-square who enters the market as the k-th in line when only β-rounds are available. The A-square will wait in the queue if and only if

$$\frac{kc}{p} < U_A(\alpha) - U_A(\beta).$$

From regularity, the A-square is never indifferent between waiting in the queue and matching immediately, and the lemma's conclusion follows. Analogous analysis holds for α-rounds. ■

Proof of Lemma 3:

Consider the first moment the market contains at least one β-round and at least one B-square and suppose the queue of β-rounds is of length k. We show that at least one β-round is willing to match with a B-square immediately. Toward a contradiction, suppose this is not the case. Since A-squares and α-rounds cannot be waiting in the market simultaneously, and not all squares are A-squares (as there is at least one B-square available), it follows that $k_{A\alpha} < k$. Notice that if a β-round is not willing to match with a B-square at a certain period, she will not be willing to match with a B-square in future periods as well. Thus, for the last β-round in queue to match with an A-square, all preceding $k-1$ β-rounds would need to match with A-squares first. From Lemma 2, at least $\bar{k}^{dec} - k_{A\alpha} + k \geq \bar{k}^{dec} + 1$ arrivals of A-squares would therefore be necessary for the last β-round in queue to match with an A-square. Hence, the expected waiting cost for that β-round is at least $\frac{(\bar{k}^{dec}+1)c}{p}$. However, by super-modularity, the marginal benefit for the β-round from matching with an A-square

instead of a B-square, $U_\beta(A) - U_\beta(B)$, is lower than $U_\alpha(A) - U_\alpha(B)$. Since an α-round is not willing to wait to match with an A-square when faced with a queue longer than $\bar{k}^{dec}$, the last β-round in the queue would prefer not to wait either, in contradiction.

Similar arguments follow for the B-squares in the market. Therefore, in any trembling-hand equilibrium, there cannot be both B-squares and β-rounds waiting in the market. ■

Proof of Proposition 3: First, the (signed) length of the A-α queue, denoted by $k_{A\alpha}$ constitutes an ergodic Markov chain. Following arguments in the body of the paper, the unique steady state distribution of $k_{A\alpha}$ is the uniform distribution over $\{-\bar{k}^{dec}, -\bar{k}^{dec}+1, \ldots, \bar{k}^{dec}\}$.

At any time t, suppose that $k_{A\alpha} > 0$. Clearly, the queue has no α-rounds. As equal numbers of squares and rounds enter and exit the market, it must be that $k_{A\alpha} + k_B = k_\beta$. Lemma 3 guarantees that $k_B = 0$, therefore $k_\beta = k_{A\alpha}$. Similarly for $k_{A\alpha} \leq 0$. ■

Proof of Corollary 2: Whenever $c > \frac{p(1-p)U}{2}$, the optimal mechanism matches arriving agents immediately, $\bar{k}^{opt} = 0$, and $\bar{k}^{opt} \leq \bar{k}^{dec}$.

Suppose, then, that $c < \frac{p(1-p)U}{2}$. We then have

$$\sqrt{\frac{p(1-p)U}{2c}} < \frac{p(1-p)U}{2c} \leq \frac{pU}{2c} \leq \frac{p(U_A(\alpha) - U_A(\beta))}{c}.$$

and the result follows from the definitions of $\bar{k}^{opt}$ and $\bar{k}^{dec}$. ■

9.3 Proof Regarding Welfare Comparisons

Proof of Proposition 4:

1. As in the proof of Corollary 1, $W^{opt}(c) - W^{dec}(c)$ is differentiable at any $c < \frac{p(1-p)U}{2}$ such that $\bar{k}^{opt}, \bar{k}^{dec} \notin \mathbb{Z}_+$. In a small neighborhood around any such c, the thresholds corresponding to both the optimal and decentralized thresholds are constant in c. Therefore,

$$\frac{d(W^{opt}(c) - W^{dec}(c))}{dc} = -\frac{2\bar{k}^{opt}(\bar{k}^{opt}+1)}{2\bar{k}^{opt}+1} + \frac{2\bar{k}^{dec}(\bar{k}^{dec}+1)}{2\bar{k}^{dec}+1} \geq 0,$$

where the inequality follows from Corollary 2, which implies that $\bar{k}^{dec} \geq \bar{k}^{opt}$. Furthermore, the proof of Corollary 1 implies that $W^{opt}(c)$ is continuous in c and so $W^{opt}(c) - W^{dec}(c)$ is increasing in any point c for which $\bar{k}^{dec} \notin \mathbb{Z}_+$.

Denote by:

$$\bar{k}^{dec}(c) \equiv \left\lfloor \frac{p\left(U_A(\alpha) - U_A(\beta)\right)}{c} \right\rfloor \quad \text{and} \quad \bar{k}^{opt}(c) \equiv \left\lfloor \sqrt{\frac{p(1-p)U}{2c}} \right\rfloor,$$

essentially explicitly expressing the dependence of both the decentralized and optimal thresholds on the waiting costs, respectively.

Let $\{d_k\}_{k=1}^{\infty}$ denote the decreasing sequence of costs such that $k = \frac{p(U_A(\alpha)-U_A(\beta))}{d_k}$. That is, cost d_k corresponds to the maximal cost such that the equilibrium threshold is k in the decentralized setting.

For an arbitrary m, we will show that

$$W^{opt}(d_{m+1}) - W^{dec}(d_{m+1}) > W^{opt}(d_m) - W^{dec}(d_m),$$

or equivalently that

$$W^{opt}(d_{m+1}) - W^{opt}(d_m) > W^{dec}(d_{m+1}) - W^{dec}(d_m). \tag{6a}$$

We first focus on $W^{opt}(c)$. When $\bar{k}^{opt}(c) > 0$, $\bar{k}^{opt}(c) \in \mathbb{Z}_+$, both $\bar{k}^{opt}(c)$ and $\bar{k}^{opt}(c) - 1$ generate the same welfare when waiting costs are c under the optimal mechanism (so, in particular, what follows holds even if d_m is a transition point corresponding to the thresholds under the optimal mechanism). Note that $W^{opt}(c)$ is piece-wise linear and continuous. It follows that:

$$W^{opt}(d_{m+1}) - W^{opt}(d_m) \geq \int_{d_{m+1}}^{d_m} \frac{2\bar{k}^{opt}(c)(\bar{k}^{opt}(c)+1)}{2\bar{k}^{opt}(c)+1} dc.$$

Let $k^0 \equiv \left\lfloor \sqrt{\frac{p(1-p)U}{2d_m}} \right\rfloor$. For any $c \in [d_{m+1}, d_m]$, $\bar{k}^{opt}(c) \geq k^0$ and

$$\frac{2\bar{k}^{opt}(c)(\bar{k}^{opt}(c)+1)}{2\bar{k}^{opt}(c)+1} = \frac{1}{2}\left((2\bar{k}^{opt}(c)+1) - \frac{1}{2\bar{k}^{opt}(c)+1}\right) \geq \frac{2k^0(k^0+1)}{2k^0+1}$$

Thus,

$$W^{opt}(d_{m+1}) - W^{opt}(d_m) \geq \int_{d_{m+1}}^{d_m} \frac{2k^0(k^0+1)}{2k^0+1} dc = \frac{2k^0(k^0+1)}{2k^0+1}(d_m - d_{m+1})$$

$$= \frac{2k^0(k^0+1)}{2k^0+1} p(U_A(\alpha) - U_A(\beta)) \left(\frac{1}{m} - \frac{1}{m+1}\right) = \frac{2k^0(k^0+1)}{2k^0+1} \frac{p(U_A(\alpha) - U_A(\beta))}{m(m+1)}.$$

Next, we consider $W^{dec}(c)$. Denote by

$$W(k,c) \equiv S_\infty - \frac{p(1-p)U}{2k+1} - \frac{2k(k+1)c}{2k+1}.$$

Note that

$$\begin{aligned} W^{dec}(d_{m+1}) - W^{dec}(d_m) &= W(m+1, d_{m+1}) - W(m, d_m) \\ &= W(m+1, d_{m+1}) - W(m+1, d_m) + W(m+1, d_m) - W(m, d_m). \end{aligned}$$

We use the following two results:

$$W(m+1,d_{m+1}) - W(m+1,d_m) = \frac{2(m+1)(m+2)}{2m+3}(d_m - d_{m+1})$$

$$= \frac{2(m+1)(m+2)}{2m+3}\frac{p(U_A(\alpha)-U_A(\beta))}{m(m+1)} = \frac{2(m+2)p(U_A(\alpha)-U_A(\beta))}{m(2m+3)},$$

and

$$W(m+1,d_m) - W(m,d_m) = \frac{p(1-p)U}{2m+1} + \frac{2m(m+1)d_m}{2m+1} - \frac{p(1-p)U}{2m+3} - \frac{2(m+1)(m+2)d_m}{2m+3}$$

$$= \frac{2p(1-p)U - 4(m+1)^2 d_m}{(2m+1)(2m+3)} = \frac{1}{(2m+1)(2m+3)}\left(2p(1-p)U - \frac{4(m+1)^2 p(U_A(\alpha)-U_A(\beta))}{m}\right).$$

Then,

$$\begin{aligned} & W^{dec}(d_{m+1}) - W^{dec}(d_m) \\ = \; & \frac{2(m+2)p(U_A(\alpha)-U_A(\beta))}{m(2m+3)} + \frac{2p(1-p)U}{(2m+1)(2m+3)} - \frac{4(m+1)^2 p(U_A(\alpha)-U_A(\beta))}{m(2m+1)(2m+3)}. \end{aligned}$$

We want to prove that (6a) holds. It suffices to show that

$$\begin{aligned} \frac{2k^0(k^0+1)}{2k^0+1}\frac{2m+3}{m+1} &> 2(m+2) + \frac{2m}{2m+1}\frac{(1-p)U}{U_A(\alpha)-U_A(\beta)} - \frac{4(m+1)^2}{2m+1} \qquad (7) \\ &= \frac{2m}{2m+1} + \frac{2m}{2m+1}\frac{(1-p)U}{U_A(\alpha)-U_A(\beta)}. \end{aligned}$$

To prove the above inequality, we consider the following two cases:

- **Case 1:** $k^0 \geq 2$**.** Note that

$$\frac{U}{U_A(\alpha)-U_A(\beta)} < \frac{(U_A(\alpha)-U_A(\beta)) + (U_\alpha(A)-U_\alpha(B))}{U_A(\alpha)-U_A(\beta)} = 2. \qquad (8)$$

  Since the left hand side of (7) is increasing in k^0, for (7) to hold, it suffices that

$$\frac{12}{5}\frac{2m+3}{m+1} > \frac{2m}{2m+1} + \frac{4m}{2m+1} = \frac{6m}{2m+1},$$

  which holds for all m.

- **Case 2:** $k^0 = 1$**.** One sufficient condition for (7) using (8) is

$$\frac{4(2m+3)}{3(m+1)} > \frac{6m}{2m+1},$$

which holds for $m = 1, 2$, or 3.

Since $k^0 = 1$,

$$\frac{p(1-p)U}{2d_m} = \frac{(1-p)Um}{2(U_A(\alpha) - U_A(\beta))} < 4.$$

Thus, another sufficient condition for (7) in this case is

$$\frac{4(2m+3)}{3(m+1)} > \frac{2m+16}{2m+1},$$

which holds for $m \geq 4$.

To construct the partition in the proposition, let $\overline{m} = \max\{i \mid d_i \leq \underline{c}\}$ and $\underline{m} = \min\{i \mid d_i < \overline{c}\}$. Now define $c_1 = \underline{c}, c_M = \overline{c}$. If $\underline{m} = \overline{m}$, set $M = 2$ and the partition has only one atom. Otherwise, if $\underline{m} < \overline{m}$, set $c_i = d_{\overline{m}-i+1}$ for $i = 2, ..., \overline{m} - \underline{m} + 1$ and $M = \overline{m} - \underline{m} + 2$.

2. Notice that

$$\lim_{c\to 0}\left(W^{opt}(c) - W^{dec}(c)\right) = \lim_{c\to 0}\Psi(c) - \lim_{c\to 0}\Theta(c) = p\left(U_A(\alpha) - U_A(\beta)\right).$$

In particular, for sufficiently small c, $W^{opt}(c) - W^{dec}(c)$ is increasing in both p and $U_A(\alpha) - U_A(\beta)$, as needed. ■

9.4 Proofs Regarding Fixed Matching Windows

Proof of Proposition 5:

1. Notice that

$$\Delta_+ W_n = \frac{p(1-p)\cdot U}{2n} \cdot Pr(k_A \neq k_\alpha; n) - 2c \leq \frac{p(1-p)U}{2n} - 2c.$$

As the upper bound of $\Delta_+ W_n$ decreases in n, let x be the unique solution of

$$\frac{p(1-p)U}{2x} - 2c = 0.$$

It follows that

$$n^o \leq x = \frac{p(1-p)U}{4c}.$$

2. We find an upper bound of $\Delta_+ W_n$ that will provide us with a lower bound on the optimal window size.

Fix a window size n, and define a multinomial random variable X_t for each time $t = 1, 2, \ldots, n$ such that

$$X_t = \begin{cases} 1 & \text{if } (A, \beta) \text{ arrives,} \\ -1 & \text{if } (B, \alpha) \text{ arrives,} \\ 0 & \text{otherwise.} \end{cases}$$

Thus, X_t takes the values of $1, -1$, or 0 with probabilities $p(1-p)$, $(1-p)p$, or $1-2p(1-p)$, respectively. Notice that

$$k_A - k_\alpha = \sum_{t=1}^{n} X_t.$$

Let $\Phi(.)$ be the cumulative distribution function of the standard normal distribution. We use the Berry-Esseen Theorem on the speed of convergence in the Central Limit Theorem (see Feller, 1972 and Tyurin, 2010).

Theorem 1. *(Berry-Esseen) Let $Y_1, Y_2, \ldots, Y_n$ be i.i.d random variables with mean 0 and variance σ^2, $Z_n = \frac{1}{n}(Y_1 + Y_2 + ... + Y_n)$, and $\hat{G}_n$ the cumulative distribution function of $\sqrt{n}Z_n/\sigma$. Then*

$$\sup_y \left| \hat{G}_n(y) - \Phi(y) \right| \leq \frac{E[|Y_1|^3]}{2\sigma^3\sqrt{n}}.$$

Now, note that

$$\mathbf{Var}[X_1] = \mathbf{E}[|X_1|^3] = 2p(1-p).$$

Let $\hat{F}_n$ be the empirical cumulative distribution function of $\frac{k_A - k_\alpha}{\sqrt{2np(1-p)}}$. By the Berry-Esseen Theorem,

$$\begin{aligned} Pr(k_A = k_\alpha \; ; \; n) &= \hat{F}_n\left(1/\sqrt{2np(1-p)}\right) - \hat{F}_n(0) \\ &\leq \Phi(1/\sqrt{2np(1-p)}) - \Phi(0) + \frac{1}{\sqrt{2np(1-p)}} \\ &\leq \frac{1}{2\sqrt{\pi np(1-p)}} + \frac{1}{\sqrt{2np(1-p)}}. \end{aligned}$$

Thus,

$$\begin{aligned} \Delta_+ W_n &\geq \frac{p(1-p)U}{2n}\left(1 - \left(\frac{1}{2\sqrt{\pi}} + \frac{1}{\sqrt{2}}\right)\frac{1}{\sqrt{np(1-p)}}\right) - 2c \qquad (9) \\ &= \frac{p(1-p)U}{2n}\left(1 - \frac{\sqrt{2\pi}+1}{2\sqrt{\pi np(1-p)}}\right) - 2c. \end{aligned}$$

We now show that there exists $c_1 > 0$ such that for all $c < c_1$, the optimal fixed size n^o is greater than 2 by illustrating that $\Delta_+ W_1 > 0$ and $\Delta_+ W_2 > 0$. Indeed,

$$\Delta_+ W_1 = \frac{p(1-p)U}{2} Pr(k_A \neq k_\alpha \; ; \; 1) - 2c = p^2(1-p)^2 U - 2c > 0,$$

and

$$\begin{aligned}
\Delta_+ W_2 &= \frac{p(1-p)U}{2} Pr(k_A \neq k_\alpha \; ; \; 2) - 2c \\
&= \frac{p(1-p)U}{4} \left(1 - \left(p^2 + (1-p)^2\right)^2 - 2p^2(1-p)^2\right) - 2c \\
&\geq \frac{p(1-p)U}{4}(1 - p^2 - (1-p)^2) - \frac{p^3(1-p)^3 U}{2} - 2c \\
&= \frac{p^2(1-p)^2 U}{2} - \frac{p^3(1-p)^3 U}{2} - 2c \geq \frac{3p^2(1-p)^2 U}{8} - 2c > 0
\end{aligned}$$

whenever $c < c_1 \equiv \frac{3p^2(1-p)^2 U}{16}$.

We now consider $n \geq 3$. From (9), $\Delta_+ W_n \geq h(r)$, where $r = n^{-1/2}$ and

$$h(r) \equiv \frac{p(1-p)Ur^2}{2} \left(1 - \frac{(\sqrt{2\pi}+1)r}{2\sqrt{\pi p(1-p)}}\right) - 2c.$$

We use the following observations, assuming that $c < c_2 \equiv \frac{4\pi p^2(1-p)^2 U}{27(\sqrt{2\pi}+1)^2} < c_1$:

1. $\Delta_+ W_3 \geq h\left(1/\sqrt{3}\right) = \frac{p(1-p)U}{6} \left(1 - \frac{\sqrt{2\pi}+1}{2\sqrt{3\pi p(1-p)}}\right) - 2c$. In order to show that $h(1/\sqrt{3}) > 0$, we show that

$$c < \frac{4\pi p^2(1-p)^2 U}{27(\sqrt{2\pi}+1)^2} < \frac{p(1-p)U}{12} \left(1 - \frac{\sqrt{2\pi}+1}{2\sqrt{3\pi p(1-p)}}\right).$$

   Indeed, the right inequality holds whenever

$$\frac{\pi}{27(\sqrt{2\pi}+1)^2} < \frac{1}{12} - \frac{\sqrt{2\pi+1}}{24\sqrt{3\pi}},$$

   which holds if and only if

$$\frac{12\pi}{27(\sqrt{2\pi}+1)^2} + \frac{\sqrt{2\pi+1}}{2\sqrt{3\pi}} = 0.4437422... < 1.$$

2. $h'(r) = 0$ at $r = 0$ and $\frac{4\sqrt{\pi p(1-p)}}{3(\sqrt{2\pi}+1)}$.

3. $h(0) < 0$ and $h\left(\frac{4\sqrt{\pi p(1-p)}}{3(\sqrt{2\pi}+1)}\right) = \frac{8\pi p^2(1-p)^2 U}{27(\sqrt{2\pi}+1)^2} - 2c > 0$.

4. From points 2 and 3 it follows that there exists a unique $r^* > 0$ such that $h(r^*) = 0$ and $h'(r^*) > 0$.

5. For any $\varepsilon > 0$, there exists c_3 such that for any $c < c_3$ we have $h\left(\sqrt{\frac{(4+\varepsilon)c}{p(1-p)U}}\right) > 0$. In order to show this, note that

$$h\left(\sqrt{\frac{(4+\varepsilon)\,c}{p(1-p)U}}\right) = \left(2+\frac{\varepsilon}{2}\right)c\left(1-\frac{(\sqrt{2\pi}+1)\sqrt{(4+\varepsilon)\,c}}{2\sqrt{\pi U}p(1-p)}\right) - 2c$$

and the claim holds for

$$c_3 = \left(1-\frac{4}{4+\varepsilon}\right)^2\left(\frac{2\sqrt{\pi U}p(1-p)}{(\sqrt{2\pi}+1)\sqrt{(4+\varepsilon)}}\right)^2.$$

6. From points 4 and 5, it follows that for $c < \min\{c_2, c_3\}$, $r^* < \sqrt{\frac{(4+\varepsilon)c}{p(1-p)U}}$.

To summarize, for any $c < c^* \equiv \min\{c_2, c_3\}$, for every window size n such that $\frac{1}{\sqrt{n}} \geq \sqrt{\frac{(4+\varepsilon)c}{p(1-p)U}}$, we have $h(1/\sqrt{n}) > 0$. On the other hand, $0 \geq \Delta_+ W_{n^o} \geq h(1/\sqrt{n^o})$. Therefore, $n^o \geq \frac{p(1-p)U}{(4+\varepsilon)c}$. ■

Proof of Proposition 6: We denote by S_n the ex-ante per-pair surplus when the window size is n. We find bounds on S_n starting from the following inequalities:

$$S_n \geq \mathbb{E}\left[\min\{\frac{k_A}{n}, \frac{k_\alpha}{n}\}\right]\cdot U_{A\alpha} + \left(1 - \mathbb{E}\left[\min\{\frac{k_A}{n}, \frac{k_\alpha}{n}\}\right]\right)\cdot U_{B\beta} \quad \text{and}$$
$$S_n \leq \mathbb{E}\left[\max\{\frac{k_A}{n}, \frac{k_\alpha}{n}\}\right]\cdot U_{A\alpha} + \left(1 - \mathbb{E}\left[\max\{\frac{k_A}{n}, \frac{k_\alpha}{n}\}\right]\right)\cdot U_{B\beta}.$$

Notice that

$$\begin{aligned}
\mathbb{E}\left[\min\left\{\frac{k_A}{n}, \frac{k_\alpha}{n}\right\}\right] &= \frac{\mathbb{E}\left[|k_A + k_\alpha|\right] - \mathbb{E}\left[|k_A - k_\alpha|\right]}{2n} = p - \frac{1}{2}\sqrt{\frac{\left(\mathbb{E}\left[|k_A - k_\alpha|\right]\right)^2}{n^2}} \\
&\geq p - \frac{1}{2}\sqrt{\frac{\mathbb{E}\left[(k_A - k_\alpha)^2\right]}{n^2}} \qquad \text{(Jensen's inequality)} \\
&= p - \frac{1}{2}\sqrt{\frac{\mathbb{E}\left[k_A^2\right] + \mathbb{E}\left[k_\alpha^2\right] - 2\mathbb{E}\left[k_A k_\alpha\right]}{n^2}} \\
&= p - \frac{1}{2}\sqrt{2\left(p^2 + \frac{p(1-p)}{n}\right) - 2p^2} \qquad \text{(since } E[k_A^2] = V(k_A) + (E[k_A])^2\text{)} \\
&= p - \sqrt{\frac{p(1-p)}{2n}}.
\end{aligned}$$

Similarly, we obtain

$$\mathbb{E}\left[\max\left\{\frac{k_A}{n},\frac{k_\alpha}{n}\right\}\right] \leq p+\sqrt{\frac{p(1-p)}{2n}}.$$

Therefore, for every window size n,

$$S_\infty - \sqrt{\frac{p(1-p)}{2n}}\left(U_{A\alpha}-U_{B\beta}\right) \leq S_n \leq S_\infty + \sqrt{\frac{p(1-p)}{2n}}\left(U_{A\alpha}-U_{B\beta}\right).$$

The per-pair welfare from a window size n is determined by:

$$W_n \equiv S_n - 2c(n-1).$$

The optimal ex-ante per-pair welfare from a static matching with a fixed-window size is defined as

$$W^o \equiv S_{n^o} - 2c(n^o-1).$$

We now turn to the parts of the proposition.

1. From part 1 of Proposition 5, $n^o \leq \frac{p(1-p)U}{4c}$. Therefore,

$$W_n \equiv S_n - 2c(n-1) \geq S_\infty - \sqrt{\frac{p(1-p)}{2n}}(U_{A\alpha}-U_{B\beta}) - 2c(n-1).$$

We can verify that the lower bound decreases with n by differentiating the right hand side of the last inequality (ignoring integer constraints). Thus,

$$W^{fix}(c) = W_{n^o} \geq S_\infty - \sqrt{\frac{2c}{U}}(U_{A\alpha}-U_{B\beta}) + 2c - \frac{p(1-p)Uc}{2}.$$

2. From the above inequalities,

$$W^{fix}(c) \leq S_\infty + \sqrt{\frac{p(1-p)}{2n^o}}(U_{A\alpha}-U_{B\beta}) - 2c(n^o-1).$$

From part 2 of Proposition 5, for any $\varepsilon > 0$, there exists c^* such that when $c < c^* \leq \frac{4\pi p^2(1-p)^2U}{27(\sqrt{2\pi}+1)^2}$, $n^o \geq \frac{p(1-p)U}{(4+2\varepsilon)c}$. Therefore,

$$W^{fix}(c) \leq S_\infty + \sqrt{\frac{2c}{U}}(U_{A\alpha}-U_{B\beta}) - \frac{p(1-p)U}{2+\varepsilon} + 2c.$$

■

10 References

Akbarpour, Mohammad, Shengwu Li, and Shayan Oveis Gharan. 2014. "Dynamic Matching Market Design," mimeo.

Anderson, Ross, Itai Ashlagi, David Gamarnik, and Yash Kanoria. 2014. "A Dynamic Model of Barter Exchange," mimeo.

Ashlagi, Itai, Patrick Jaillet, and Vahideh H. Manshadi. 2013. "Kidney Exchange in Dynamic Sparse Heterogenous Pools," mimeo.

Baccara, Mariagiovanna, Allan Collard-Wexler, Leonardo Felli, and Leeat Yariv. 2014. "Child-Adoption Matching: Preferences for Gender and Race," *American Economic Journal: Applied Economics*, **6(3)**, 133-158.

Becker, Gary S. 1974. "A Theory of Marriage: Part II," *The Journal of Political Economy*, **82(2)**, S11-S26.

Bloch, Francis and David Cantala. 2014. "Dynamic Allocation of Objects to Queuing Agents: The Discrete Model," mimeo.

Budish, Eric, Peter Cramton, and John Shim. 2015. "The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response," *The Quarterly Journal of Economics*, forthcoming.

Burdett, Ken and Melvyn G. Coles. 1997. "Marriage and Class," *The Quarterly Journal of Economics*, **112(1)**, 141-168.

Choo, Eugene. 2015. "Dynamic Marriage Matching: An Empirical Framework," forthcoming at *Econometrica*.

Doval, Laura. 2015. "A Theory of Stability in Dynamic Matching Markets," mimeo.

Eeckhout, Jan. 1999. "Bilateral Search and Vertical Heterogeneity," *International Economic Review*, **40(4)**, 869-887.

Feller, William. 1972. *An Introduction to Probability Theory and Its Applications, Volume II* (2nd ed.). New York: John Wiley & Sons.

Gjertson, David W. 2004. "Explainable Variation in Renal Transplant Outcomes: A Comparison of Standard and Expanded Criteria Donors," *Clinical Transplants*, **2004**, 303-314.

Gurvich, Itai and Amy Ward. 2014. "On the Dynamic Control of Matching Queues," *Stochastic Systems*, **4(2)**, 479-523.

Haeringer, Guillaume and Myrna Wooders. 2011. "Decentralized Job Matching," *International Journal of Game Theory*, **40**, 1-28.

Hall, Robert E. and Alan B. Krueger. 2012. "Evidence on the Incidence of Wage Posting, Wage Bargaining, and On-the-Job Search." *American Economic Journal: Macroeconomics*, **4(4)**, 56-67.

Hitch, Gunter J., Ali Hortacsu, and Dan Ariely. 2010. "Matching and Sorting in Online Dating," *The American Economic Review*, **100(1)**, 130-163.

Kocer, Yilmaz. 2014. "Dynamic Matching and Learning," working slides.

Leshno, Jacob. 2014. "Dynamic Matching in Overloaded Systems," mimeo.

Niederle, Muriel and Leeat Yariv. 2009. "Decentralized Matching with Aligned Preferences," mimeo.

Oien, Cecilia M., Anna V. Reisaeter, Torbjørn Leivestad, Friedo W. Dekker, and Pål-Dag Line. 2007. "Living Donor Kidney Transplantation: The Effects of Donor Age and Gender on Short- and Long-term Outcomes," *Transplantation*, **83(5)**, 600-606.

Pais, Joana V. 2008. "Incentives in decentralized random matching markets," *Games and Economic Behavior*, **64**, 632-649.

Rogerson, Richard, Robert Shimer, and Randall Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey," *Journal of Economic Literature*, **XLIII**, 959-988.

Satterthwaite, Mark and Artyom Shneyerov. 2007. "Dynamic Matching, Two-sided Incomplete Information, and Participation Costs: Existence and Convergence to Perfect Competition," *Econometrica*, **75(1)**, 155-200.

Stein, Rob. 2011. "Under Kidney Transplant Proposal, Younger Patients would Get the Best Organs," *The Washington Post*, February 24.

Taylor, Curtis. 1995. "The Long Side of the Market and the Short End of the Stick: Bargaining Power and Price Formation in Buyers', Sellers', and Balanced Markets," *The Quarterly Journal of Economics*, 110(3), 837-855.

Tyurin, Ilya S. 2010. "An improvement of upper estimates of the constants in the Lyapunov theorem," *Russian Mathematical Surveys*, **65(3(393))**, 201-202.

Ünver, Utku. 2010. "Dynamic Kidney Exchange," *Review of Economic Studies*, **77**, 372-414.

Zenios, Stefanos A. 1999. "Modeling the transplant waiting list: A queueing model with reneging," *Queueing Systems*, **31**, 239-251.