Rigidity of Transfers and Unraveling in Matching Markets*

Songzi Du[†] Yair Livne[‡]

January 10, 2016

Abstract

We study the role of transfers in the timing of matching. In our model, agents have the option of matching early and exiting in period 1, before new agents arrive in period 2; in period 2 after all agents arrive a stable matching is implemented for those who are present. We prove that without flexible transfers, when the number of agents in period 1 is large and the number of new arrivals is small, on average at least one quarter of all agents in period 1 have strict incentives to match early, if they anticipate others are participating in the stable matching of period 2. We define a minimal notion of sequential stability and prove that without flexible transfers the probability of sequential stability tends to 0 as the number of period-one agents gets large. We show that flexible transfers eliminate these timing problems.

*We are grateful to Mike Ostrovsky, Andy Skrzypacz and Bob Wilson for guidance during research and extensive discussion. And for helpful comments we thank Yossi Feinberg, Alex Frankel, Ben Golub, John Hatfield, Fuhito Kojima, Scott Kominers, Vijay Krishna, Eli Livne, Paul Milgrom, Muriel Niederle, and Michael Peters. A previous version of this paper was circulated under the title "Unraveling and Chaos in Matching Markets."

[†]Simon Fraser University, Department of Economics. Email: songzid@sfu.ca

[‡]Quora, Inc., Mountain View, CA. Email: yair@quora.com

1 Introduction

In this paper we analyze a dynamic matching model with the possibility of early matching before all agents (or all information) arrive to the market. We show that flexible transfers are crucial for preventing early matching. Moreover, we show that without flexible transfers there may not even exist an intertemporal matching scheme that is stable (even though the market would be stable if it was static!). Surprisingly, this problem gets particularly acute if the market is large, counter to our intuitions that thick markets always operate better than thin ones.

Many two-sided labor matching markets suffer from unraveling and intertemporal instabilities, with matching decisions being taken before the normal timeline or market participants failing to coordinate on a timeline. Some well-known examples of unraveling include the markets for federal judicial clerks, clinical psychology interns, and medical residents[1] (Roth and Xing, 1994). In particular, in the judicial clerk market there have been six failed attempts between 1978 and 1998 to coordinate the timing of hiring (Avery, Jolls, Posner, and Roth, 2007).[2] Consistent with our results, many of the markets that suffer from unraveling are ones where the wages are more or less fixed and not up for negotiation. For the example of judicial clerk market, the wage schedule is fixed by the US Congress.

In our model agents are either men or women, are risk neutral, and generate surplus through matching. We compare two regimes of dividing the surplus from matching: in the fixed-transfer regime, there is an exogenously fixed division of surplus, perhaps due to some institutional constraint; in the flexible-transfer regime, the division of surplus is endogenously determined along with the matching. We assume that matching can take place in one of the two periods: in the first period, some men and women are present and have the option of matching early (and exiting), or waiting for the second period; in the second period, which corresponds to the normal timeline of matching, some new men and women arrive, and a stable matching for the agents present is implemented, perhaps by a centralized institution. One can interpret the new agents arriving in the second period in many ways. For example, the new arrivals may be unobserved agents, due to the imperfection of information in the first period. In student-job matching, the new arrivals can be new job openings on one side,

[1] Wetz, Seelig, Khoueiry, and Weiserbs (2010) estimate that in 2007 one in five resident positions in primary care specialties was offered outside of the centralized National Resident Matching Program (NRMP); the data in Wetz, Seelig, Khoueiry, and Weiserbs (2010) suggests that unraveling still exists in the medical resident market, reminiscent of the situation before the introduction of NRMP.

[2] The most recent attempt to coordinate the timing of hiring in the judicial clerk market was initiated in 2004 and abandoned in 2013.

and students unavailable for early interviews (say, due to incomplete preparation) on the other side. In the judicial clerk market, the new arrivals may be judges and students who stick to the regulated timeline of interviewing and hiring for moral reasons.[3]

Regardless of the interpretation, the uncertainty about the new agents in the second period creates incentives for unraveling in the fixed-transfer regime: if all other agents were to wait for the second period, some pairs of agents would have incentives to match early and leave the market. In our setting, this happens when agents on both sides of the market are positioned high relative to their surroundings, and thus for these agents waiting for new arrivals has a larger downside than upside. Our first result states that when the transfers are fixed and as the market gets large, on average at least one quarter of all agents in the first period have strict incentives to match early, independent of the underlying type distributions and surplus function. This is sharply contrasted to the flexible-transfer regime, which eliminates all incentives to match early. The intuition is simple: when the transfers are exogenously fixed, they do not reflect the demand-supply condition in period 2 (i.e., they do not clear the market in period 2); as a result, some agents have incentive to block the outcome in period 2, and some of those agents may engage in such blocking by matching early in period 1. In contrast, when the transfers are endogenously determined in period 2, they accurately reflect the supply-demand condition, so no pair of agents have joint incentive to block the outcome in period 2, and there is also no joint incentive to block in period 1 (no incentive to match early).

Given our result that a lot of agents have incentives to match early in the fixed-transfer regime, it is natural to conjecture that agents in the first period would split into two groups, with one group matching early and the other waiting for the second period; the second group may be empty, representing the possibility of total unraveling. They could achieve such a split as an equilibrium outcome or as a result of an organized dynamic matching mechanism. Avery, Jolls, Posner, and Roth (2007) call this sort of split a "mixed adherence and nonadherence" to the start date (which corresponds to the second period in our model) and write about the possibilities of such split being sustained in equilibrium in the judicial clerks market.

Our second result shows that in large markets with fixed transfers, any mixed adherence and nonadherence to the start date is difficult to sustain in equilibrium; for most realizations of types, in every possible early-matching scheme there are agents with incentives to deviate:

[3] Avery, Jolls, Posner, and Roth (2007) quote judges and students who expressed frustrations over the fact that many interview and offer/accept positions before the regulated timeline.

either from waiting for period-two matching to early matching, or *vice versa*. We use the term *sequential instability* to refer to the presence of these deviations. Sequential instability represents a failure of coordination in the timing of matching and is once again a consequence of fixed transfers: when the transfers are allowed to freely adjust, no one would have an incentive to deviate from matching in period 2, and hence everyone matching in period 2 is sequentially stable.

Finally, we study an intermediate regime between the two extreme points: imagine that the formal matching institution in period 2 cannot implement flexible transfers due to some institutional constraint, but if agents match early (i.e., outside of the formal institution) they may negotiate any desirable level of transfers. We show that flexible transfers in period 1 alone is not sufficient to prevent early matching nor to guarantee sequential stability, and in general expand the possibilities of early matching. These results thus reinforce the importance of flexible transfers in the formal matching institution (i.e., period 2 in our model). Some practical examples of matching institution with flexible transfers include the proposal of Crawford (2008) for incorporating salaries in National Residence Matching Program, and of Sonmez and Switzer (2012) and Sonmez (2012) for matching cadets to military services where the transfers are (additional) years of services. Our paper contributes a new rationale for the desirability of flexible transfers in matching institutions.[4]

Our paper compares two classical models of matching: Gale-Shapley (matching without transfers, or equivalently with fixed transfers) and Becker-Shapley-Shubik (matching with flexible transfers). It is well-known that the predictions of matching with and without transfers have close analogies in the standard (static) setting: for example, the lattice structure of stable matchings, the existence of a median stable matching, the incentive compatibility property of the two extreme stable matchings, etc. (Roth and Sotomayor, 1992). We show that under a natural dynamic extension matching with and without transfers have drastically different implications. On highlighting the advantage of flexible transfers, our results complement Echenique and Galichon (2014), who find that stable matching with fixed transfers can be very inefficient in comparison with stable matching with flexible transfers in the classical, static model. Relatedly, Jaffe and Kominers (2014) embed the fixed transfers and flexible transfers models as two extreme points in a model with taxation and study comparative statics on efficiency as transfers become more flexible (i.e., as the percentage of taxation decreases).

[4]Previously, Bulow and Levin (2006) have shown that stable matching with fixed wages can lead to wage suppression (relative to a stable matching with flexible wages) if firms set wages before entering the stable matching mechanism.

Our paper complements the classic work on unraveling by Li and Rosen (1998), Suen (2000) and Li and Suen (2000), who study early matching as a form of insurance due to risk aversion in settings with flexible transfers; consistent with our results, they show that early matching cannot be sustained in equilibrium when agents are risk neutral and the transfers are flexible. Risk aversion is clearly an important reason for early matching and is applicable whether or not transfers are flexible. We propose a different channel through which unraveling can arise, and this channel only impacts markets with fixed transfers. Our work suggests that markets with flexible transfers could be less prone to unraveling.

The literature has come up with other reasons for unraveling in matching markets: over provision of information (Ostrovsky and Schwarz, 2010), strategic complementarity over the decision to match early (Echenique and Pereyra, 2014), correlation of matching preferences (Halaburda, 2010), search costs (Damiano, Hao, and Suen, 2005), exploding offers (Niederle and Roth, 2009), uncertainty about the imbalance between supply and demand (Niedrele, Roth, and Unver, 2010), and information flow in social network (Fainmesser, 2013). All of these papers work in the setting of fixed transfers, and it remains an open question if flexible transfers could mitigate unraveling through these channels. Sonmez (1999) has shown that in a many-to-one matching setting with fixed transfers and complete information, unraveling is inevitable for some realization of preferences. Finally, Roth (1991) empirically demonstrates that centralized matching mechanisms that are not stable typically leads to unraveling.

The paper proceeds as follows. We present our model in Section 2 and illustrate our results with simple examples in Section 3. Section 4 shows the prevalence of early matching incentives and the likelihood of sequential instability when transfers are fixed. Section 5 shows that flexible transfers eliminate these timing problems and examines the intermediate regime in which flexible transfers are available in period 1, but not in period 2. Section 6 presents some discussion of our results. Most of the proofs are in the Appendix.

2 Model

We study a two-sided, one-to-one matching problem over two periods.

Agents In the first period n men and n women are present in the market. In the second period, k new men and k new women arrive on the market. Both n and k are positive integers, and for simplicity are assumed to be common knowledge among the agents.[5]

[5]As it will become apparent, our results are unchanged if with probability $\epsilon > 0$, k men and k women arrive, and with probability $1 - \epsilon$ no new agents arrive. And likewise if one side of the market has more new

Both men and women have types, which are independently and identically drawn (i.i.d.) from some distributions F and G, respectively. We assume that F is distributed on the interval $[\underline{m}, \overline{m}]$, with a positive and continuous density f on the interior $(\underline{m}, \overline{m})$. Likewise for distribution G on the interval $[\underline{w}, \overline{w}]$ with density g. We assume finite lower bounds $\underline{m}$ and $\underline{w}$, but the upper bounds $\overline{m}$ and $\overline{w}$ can be infinite, i.e., F and G can be distributed over $[\underline{m}, \infty)$ and $[\underline{w}, \infty)$, respectively. A common example of F and G is the uniform distribution on $[0, 1]$; we refer to this distribution as $\mathrm{U}[0, 1]$.

For each $1 \leq i \leq n$ we denote by m_i the i-th lowest type among the n men present in the first period. Likewise, women's types in the first period are denoted by $w_1 < \cdots < w_n$. We let m_i^2 be the i-th lowest type man in the second period, and w_i^2 be the i-th lowest type woman, $1 \leq i \leq l \leq k + n$, where l is the number of men/women present in the second period; we may have $l < n + k$ due to early matching and exiting in the first period.

We assume that in the first period the men and women who are present observe each other's types. (This assumption is stronger than necessary — see Remark 1.) On the other hand, agents in the first period do not know the types of the new arrivals in the second period, although they know the distributions of types.

Utility Agents are risk neural and maximize their expected, quasi-linear utilities. Agents match to generate a surplus that they then split; the surplus does not depend on the time (period 1 or 2) of matching. For simplicity, suppose that the utility of an agent who does not match is zero. We consider two regimes of dividing the surplus between men and women:

1. In the *fixed-transfer regime*[6], when a man of type m is matched to a woman of type w, man m gets value
$$U(w \mid m) \geq 0 \tag{1}$$
and woman w gets value
$$V(m \mid w) \geq 0, \tag{2}$$
where functions U and V represent a fixed scheme of dividing the $U(w \mid m) + V(m \mid w)$ units of surplus produced by man m with woman w.[7] For example, $U(m \mid w)$ can be the fixed schedule of salary received by the worker (man) from the firm (woman) when they are matched together, contingent on their types.

arrivals than the other with probability 1/2 each.

[6]We have in mind applications of our model to firm-worker matching, so we emphasize the term "fixed transfer" instead of the term "no transfer" which is commonly used in the literature.

[7]Mailath, Postlewaite, and Samuelson (2013) call $U(w \mid m)$ and $V(m \mid w)$ premuneration values.

We assume that the functions U and V are twice continuously differentiable, increasing in matched type:

$$\frac{\partial U(w \mid m)}{\partial w} > 0, \quad \frac{\partial V(m \mid w)}{\partial m} > 0, \tag{3}$$

and have bounded ratio of derivatives: there exists a positive constant $\bar{b}$ such that

$$\frac{\partial U(w \mid m)}{\partial w} \Big/ \frac{\partial U(w' \mid m)}{\partial w} \leq \bar{b}, \quad \frac{\partial V(m \mid w)}{\partial m} \Big/ \frac{\partial U(m' \mid w)}{\partial m} \leq \bar{b}, \tag{4}$$

for all $m, m' \in [\underline{m}, \overline{m}]$ and $w, w' \in [\underline{w}, \overline{w}]$.

2. In the *flexible-transfer regime*, men and women can freely negotiate the division of surplus: when man of type m is matched to woman of type w, man m gets

$$U(w \mid m) + P(m, w) \geq 0, \tag{5}$$

and woman w gets

$$V(m \mid w) - P(m, w) \geq 0, \tag{6}$$

where the transfer $P(m, w)$, which can be positive or negative, is endogenously determined (i.e., as a part of the stable matching). We assume that the total surplus $U(w \mid m) + V(m \mid w)$ features complementarities between the type of man and the type of woman:

$$\frac{\partial^2 \left(U(w \mid m) + V(m \mid w)\right)}{\partial m\, \partial w} > 0. \tag{7}$$

2.1 Fixed-transfer Regime

Second Period We assume that in the second period there is a centralized institution that implements a stable matching for all agents who are present in the second period — which consist of first-period agents who have decided to wait and the new arrivals. Let $m_l^2 > \cdots > m_1^2$ and $w_l^2 > \cdots > w_1^2$ be the men and women present in the second period. A matching in the second period is a subset $\mu^2 \subset \{m_1^2, \ldots, m_l^2\} \times \{w_1^2, \ldots, w_l^2\}$ such that every man is matched to a distinct woman.[8] The matching μ^2 depends on the types in the second period: $\mu^2 = \mu^2(m_1^2, \ldots, m_l^2, w_1^2, \ldots, w_l^2)$. To keep the notation simple, we omit this dependence.

[8] Formally, this means that (1) for every m_i^2, there exists a j such that $(m_i^2, w_j^2) \in \mu^2$, and (2) if $(m_i^2, w_j^2), (m_i^2, w_{j'}^2) \in \mu^2$, then $j = j'$.

For the fixed-transfer regime, we use the stable matching of Gale and Shapley (1962):

Definition 1. The matching μ^2 of the second period is *stable* if for every man m_i^2 and every woman w_j^2 in the second period, we have

$$U(w_{i'}^2 \mid m_i^2) \geq U(w_j^2 \mid m_i^2), \text{ or } V(m_{j'}^2 \mid w_j^2) \geq V(m_i^2 \mid w_j^2), \tag{8}$$

where $(m_i^2, w_{i'}^2) \in \mu^2$ and $(m_{j'}^2, w_j^2) \in \mu^2$. [9]

Condition (8) says that no pair of man and woman has a joint incentive to deviate from μ^2 (no blocking). By the assumption in (3), a man always prefers a woman of higher type over a woman of lower type, and likewise for the preference of woman. Therefore, a stable matching here is simply an assortative matching, i.e., the highest type man is matched to the highest type woman, the second highest to the second highest, etc. So we assume a stable/assortative matching is always implemented in period 2.

First Period and Early Matching In the first period men and women may match early and permanently leave the market, gaining utility from their match partner; or they may wait for the stable matching in the second period. Given men $m_n > \cdots > m_1$ and women $w_n > \cdots > w_1$ in the first period, an *early matching* $\mu \subset \{w_1, \ldots, w_n\} \times \{w_1, \ldots, w_n\}$ is a set of man-woman pairs such that each agent belongs to at most one pair.[10] We emphasize that unlike the second-period matching μ^2 which is a complete matching, the early matching μ could be a partial matching; in particular, we could have $\mu = \emptyset$, which represents the situation where all agents wait to the second period.

We are interested in early matchings from which no pair of agents has an incentive to deviate. Let μ^2 be the stable (i.e., assortative) matching in the second period. Given a list $L = \{m_i : i \in I\} \cup \{w_i : i \in J\}$ of men and women who wait to the second period (where $I, J \subseteq \{1, \ldots, n\}$ and $|I| = |J|$), and anticipating the new arrivals and the assortative matching μ^2 in the second period, man $m_i \in L$, has expected utility:

$$\mathbf{U}(L, m_i) \equiv \sum_{j=0}^{k} \binom{k}{j} F(m_i)^j (1 - F(m_i))^{k-j} \cdot \mathbb{E}_k[\, U(w_{r+j}^2 \mid m_i)\,], \tag{9}$$

where r is the rank (from the bottom) of m_i in $\{m_{i'} : i' \in I\}$, w_{r+j}^2 is the $(r+j)$-th

[9]Given conditions (1) and (2) on utility functions, the individual rationality requirement in stability is automatically satisfied.

[10]Formally, we have $(m_i, w_j), (m_i, w_{j'}) \in \mu \Rightarrow j = j'$ and $(m_{i'}, w_j), (m_i, w_j) \in \mu \Rightarrow i = i'$.

lowest woman in the second period, among the k new women and $\{w_{i'} : i' \in J\}$, and the expectation $\mathbb{E}_k$ is taken over the type realizations of the k new women. In (9), with probability $\binom{k}{j}F(m_i)^j(1-F(m_i))^{k-j}$ there are j number of new men with types below m_i; in this event man m_i is assortatively matched to the woman of rank $r+j$ in the second period.

Likewise, by waiting to the second period woman $w_i \in L$, has expected utility:

$$\mathbf{V}(L, w_i) \equiv \sum_{j=0}^{k} \binom{k}{j} G(w_i)^j (1 - G(w_i))^{k-j} \cdot \mathbb{E}_k[\, V(m^2_{r+j} \mid w_i)\,], \tag{10}$$

where r is the rank (from the bottom) of w_i is in $\{w_{i'} : i' \in J\}$, and m^2_{r+j} is the $(r+j)$-th lowest man in the second period, among the k new men and $\{m_{i'} : i' \in I\}$.

We are interested in early matchings that are stable in conjunction with the second-period assortative matching:

Definition 2. Fix a realization of types in the first period. The matching scheme (μ, μ^2) is *sequentially stable* for this realization if:

1. the second-period matching μ^2 is stable in the sense of Definition 1; so μ^2 is an assortative matching;

2. for any man m_i and woman w_j who both wait for the second period ($(m_i, w_{j'}) \notin \mu$ for every j', and $(m_{i'}, w_j) \notin \mu$ for every i'), we have

$$\mathbf{U}(L(\mu), m_i) \geq U(w_j \mid m_i) \text{ or } \mathbf{V}(L(\mu), w_j) \geq V(m_i \mid w_j), \tag{11}$$

   where $L(\mu)$ is the list of men and women who wait for the second period according to μ;

3. for any couple (m_i, w_j) who matches early ($(m_i, w_j) \in \mu$), we have both

$$U(w_j \mid m_i) \geq \mathbf{U}(L(\mu) \cup \{m_i, w_j\}, m_i) \text{ and } V(m_i \mid w_j) \geq \mathbf{V}(L(\mu) \cup \{m_i, w_j\}, w_j). \tag{12}$$

The sequential aspect of Definition 2 is similar to that of the subgame perfect equilibrium concept: agents anticipate a stable matching μ^2 in the second period (Point 1), and the early matching μ is pairwise stable given this anticipation (Points 2 and 3). Point 2 says that every pair of agents designated to wait on the market by μ does not want to deviate and match early (at least one member of the pair does not have incentive to match early). Point 3 says that both members of a pair designated by μ to match early prefer the designated match over waiting for the second period.

In Definition 2 we have intentionally proposed a minimal notion of sequential stability: the definition does not require the early matching μ to have no blocking (i.e., does not require the early matching μ to be assortative). Such minimal definition strengthens our negative result (Theorem 2) in the fixed-transfer regime.

We sometimes abuse the terminology by saying that an early matching μ is sequentially stable, which means that the early matching μ, in conjunction with an assortative matching in the second period, forms a sequentially stable matching scheme in the sense of Definition 2.

Remark 1. We have assumed full observability of the types in period 1 because we want, as a benchmark, an unambiguous distinction between the known types in period 1 and the unknown types from the later period. We note, however, that the full observability of the first-period types is not strictly necessary for agents' early matching decision. When thinking about his expected utility from period 2, an agent in period 1 needs not know the types of those who have matched early. And among those from period 1 who wait, an agent needs not know the types of those who are clearly out of his reach, i.e., those whose rank differs by more than k from him. When k is small compared to n, which is assumed for our fixed-transfer results, this amounts to assuming an agent knowing the types in his local surrounding in period 1.

2.2 Flexible-transfer Regime

Second Period In the second period of the flexible-transfer regime, when man m_i^2 is matched to woman w_j^2 (i.e., $(m_i^2, w_j^2) \in \mu^2$), they negotiate a transfer $P^2(m_i^2, w_j^2)$ (which could be positive or negative), from which man m_i^2 gets

$$U(w_j^2 \mid m_i^2) + P^2(m_i^2, w_j^2) \geq 0, \tag{13}$$

and woman w_j^2 gets

$$V(m_i^2 \mid w_j^2) - P^2(m_i^2, w_j^2) \geq 0. \tag{14}$$

The transfer $P^2(m_i^2, w_j^2)$ specifies a division of the surplus $U(w_j^2 \mid m_i^2) + V(m_i^2 \mid w_j^2)$ created between man m_i^2 and woman w_j^2. We require the payoffs of man and woman to be greater than or equal to zero, which is a normalization of their outside options. In general, the transfer $P^2(m_i^2, w_j^2)$ depend on all types present in the second period, i.e., we abuse the notation by writing

$$P^2(m_i^2, w_j^2) = P^2(m_i^2, w_j^2)(m_1^2, \dots, m_l^2, w_1^2, \dots, w_l^2)$$

where $m_1^2 < m_2^2 < \cdots < m_l^2$ and $w_1^2 < w_2^2 < \cdots < w_l^2$ are the types of all agents present in the second period.

As in the fixed-transfer regime, we assume that a stable matching is implemented for the agents present in the second period. In the flexible-transfer regime the stable matching is defined by Shapley and Shubik (1971) and Becker (1973) and specifies transfers as a part of the matching:

Definition 3. The matching and transfers (μ^2, P^2) of the second period is *stable* if the non-negativity conditions (13) and (14) hold for every matched pair, and for every man m_i^2 and every woman w_j^2 in the second period, we have

$$\big(U(w_{i'}^2 \mid m_i^2) + P^2(m_i^2, w_{i'}^2)\big) + \big(V(m_{j'}^2 \mid w_j^2) - P^2(m_{j'}^2, w_j^2)\big) \geq U(w_j^2 \mid m_i^2) + V(m_i^2 \mid w_j^2), \tag{15}$$

where $(m_i^2, w_{i'}^2) \in \mu^2$ and $(m_{j'}^2, w_j^2) \in \mu^2$.

Condition (15) is the analogue of condition (8) in the flexible-transfer regime and says that no pair of man and woman can find a division of their surplus that both prefer over what they get in (μ^2, P^2) (no blocking). We collectively refer to the matching and transfers (μ^2, P^2) in Definition 3 as a stable matching when there is no possibility of confusion.

Shapley and Shubik (1971) and Becker (1973) have proved that in the flexible-transfer regime a stable matching (μ^2, P^2) always exists, and moreover given complementarities in the surplus function (Assumption (7)), in any stable matching μ^2 must be the assortative matching. So as in the fixed-transfer regime we assume that an assortative matching is always implemented in the second period. For every realization of types in the second period, we fix a schedule of transfers P^2 that make the assortative matching stable. For example, P^2 can be the transfers that make the stable matching woman-optimal, which can be implemented by an ascending-price auction (Demange, Gale, and Sotomayor, 1986). Another natural choice is the transfers associated with the median stable matching (Schwarz and Yenmez, 2011).

First Period and Early Matching As in the fixed-transfer regime, in the first period agents can match early and skip the stable matching in the second period, which motivates the definition of sequential stability. Let (μ^2, P^2) be the stable matching and transfers in the second period, which means that μ^2 is an assortative matching. Given a list $L = \{m_i : i \in I\} \cup \{w_i : i \in J\}$ of men and women who wait to the second period (where $I, J \subseteq \{1, \ldots, n\}$ and $|I| = |J|$), and anticipating the new arrivals and (μ^2, P^2) in the second period, man $m_i \in L$ has expected utility:

$$\mathbf{U}(L, m_i; P^2) \equiv \sum_{j=0}^{k} \binom{k}{j} F(m_i)^j (1 - F(m_i))^{k-j} \cdot \mathbb{E}_k[\, U(w^2_{r+j} \mid m_i) + P^2(m_i, w^2_{r+j})\,], \tag{16}$$

where r is the rank (from the bottom) of m_i in $\{m_{i'} : i' \in I\}$, w^2_{r+j} is the $(r+j)$-th lowest woman in the second period, among the k new women and $\{w_{i'} : i' \in J\}$, and the expectation $\mathbb{E}_k$ is taken over the type realizations of the k new women.

Likewise, by waiting to the second period woman $w_i \in L$ has expected utility:

$$\mathbf{V}(L, w_i; P^2) \equiv \sum_{j=0}^{k} \binom{k}{j} G(w_i)^j (1 - G(w_i))^{k-j} \cdot \mathbb{E}_k[\, V(m^2_{r+j} \mid w_i) - P(m^2_{r+j}, w_i)\,], \tag{17}$$

where r is the rank (from the bottom) of w_i is in $\{w_{i'} : i' \in J\}$, and m^2_{r+j} is the $(r+j)$-th lowest man in the second period, among the k new men and $\{m_{i'} : i' \in I\}$.

Given these expected utilities in period 1, we are led to the following definition of sequential stability:

Definition 4. Fix a realization of types in the first period. The matching scheme (μ, μ^2, P^2) is *sequentially stable* for this realization if:

1. the second-period matching (μ^2, P^2) is stable in the sense of Definition 3; so μ^2 is an assortative matching;

2. for any man m_i and woman w_j who both wait for the second period ($(m_i, w_{j'}) \notin \mu$ for every j', and $(m_{i'}, w_j) \notin \mu$ for every i'), we have

$$\mathbf{U}(L(\mu), m_i; P^2) + \mathbf{V}(L(\mu), w_j; P^2) \geq U(w_j \mid m_i) + V(m_i \mid w_j), \tag{18}$$

   where $L(\mu)$ is the list of men and women who wait for the second period according to μ;

3. for any couple (m_i, w_j) who matches early $((m_i, w_j) \in \mu)$, we have

$$U(w_j \mid m_i) + V(m_i \mid w_j) \geq \mathbf{U}(L(\mu) \cup \{m_i, w_j\}, m_i; P^2) + \mathbf{V}(L(\mu) \cup \{m_i, w_j\}, w_j; P^2). \tag{19}$$

Definition 4 is the analogue of Definition 2 in the flexible-transfer regime: agents anticipates a stable matching (μ^2, P^2) in the second period (Point 1), and the early matching μ is pairwise stable given this anticipation (Points 2 and 3). Point 2 says that for any pair of agents designated to wait on the market by μ and for any division of their surplus $(U(w_j \mid m_i) + V(m_i \mid w_j))$ from matching early, at least one member of the pair prefers his/her expected utility from not matching early over his/her division from matching early. Point 3 says that members of a pair designated by μ to match early can find a division of their surplus of which both prefer over their expected utilities from not matching early.

As in the fixed transfer regime, Definition 4 is a weak notion of sequential stability since it does not require the early matching μ to feature no blocking. This weak definition is irrelevant for the first part of our main result Theorem 3 (since no one matches early), and it strengthens the claim in the second part that no one matching early is the unique sequentially stable outcome when the second period matching is woman-optimal (or man-optimal).

3 Examples

In this section we illustrate with examples the incentive to match early (Inequality (12)) as well as the possible failure of sequential stability (Definition 2) in the fixed-transfer regime. We then show in the examples how flexible transfers can address these problems.

For the examples in this section, we assume either $n = 1$ or 2 pairs of agents in period 1, and $k = 1$ pair of new arrivals in period 2. Moreover, we assume a uniform distribution of types: $F = G = \mathrm{U}[0, 1]$, and multiplicative value functions $U(w \mid m) = V(m \mid w) = mw/2$. Clearly, in the fixed-transfer regime these value functions are equivalent to the man getting a value of w and the woman getting a value of m when man m is matched to woman w.

3.1 Example 1: early matching

We first examine the case of $n = 1$ and $k = 1$ in the fixed-transfer regime. Let $m \in [0, 1]$ be the type of the man in period 1, and let $w \in [0, 1]$ be the type of the woman in period 1.

The incumbent man m prefers to leave the market in the first period with the incumbent woman w if his expected match in the second period (from the anticipated assortative matching) is lower than w. The second-period assortative matching can fall into three cases:

1. if both entrants are of higher type than the incumbents, or if both entrants are of lower type than the incumbents, then the incumbents are matched in the second period;

2. if the entrant man is of a higher type than m while the entrant woman is of lower type than w, then the incumbent man is matched to the entrant woman, and gets a match of quality lower than w. These are the realizations in which the man loses from waiting, and the woman gains;

3. if the entrant man is of a lower type than m while the entrant woman is of higher type than w, then the incumbent man is matched to the entrant woman, and gets a match of quality higher than w. These are the realizations in which the man gains by waiting, and the woman loses.

Formally, man m's expected utility from waiting to the second period is (after normalizing $U(w \mid m) = \frac{1}{2}mw$ to w):

$$\big(mw + (1-m)(1-w)\big)w + (1-m)\int_0^w x\,dx + m\int_w^1 x\,dx$$

where the three terms correspond to the three cases above. Comparing this to w and simplifying, man m strictly prefers to match early with w if and only if:[11]

$$(1-m)(w)^2 > m(1-w)^2. \tag{20}$$

Symmetrically, woman w strictly prefers to match early with m if and only if:

$$(1-w)(m)^2 > w(1-m)^2. \tag{21}$$

Therefore, pairs (m, w) which satisfy both of the above conditions would rather match early and leave the market over waiting for the assortative matching in the second period. The set of such pairs is non-empty, and is plotted in Figure 1.

The figure reveals that with a non-negligible probability (approximately 9.7%) the first-period agents have strict incentive to match early. Agents that prefer an early match (in the

[11] We write $(w)^2$ for the square of w because we use w^2 for the type of woman in period 2.

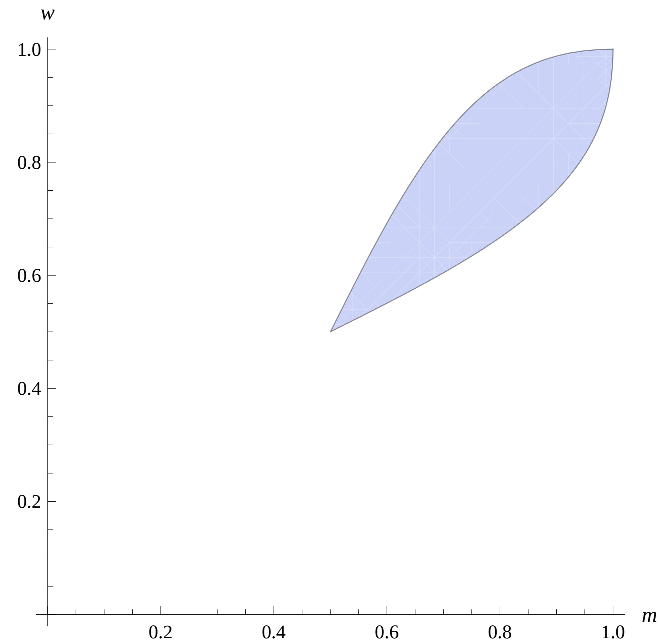


Figure 1: Types in the shaded area prefer to match early (both (20) and (21) are satisfied).

shaded region) have two notable characteristics: they are of *similar type* and have relatively *high types*. Intuitively, pairs with similar types have a similar probability of gaining or losing from waiting to period 2. However, since they are of high type and hence the upside from waiting is smaller than the possible downside. Thus, these agents experience an "endogenous discounting," preferring to leave the market. Other agents prefer to wait — if the two types are not similar, the higher of the two knows that he/she can expect with high probability a better draw in the second period, and thus prefers to wait. Alternatively, if the two types are similar but both low, then both agents have a higher upside than downside from waiting, since a good draw is expected to improve their match by more than a bad one would.

Summarizing, for pairs in shaded region of Figure 1, early matching $\mu = \{(m, w)\}$, in conjunction with the second period assortative matching, forms a sequentially stable matching scheme (Definition 2). For pairs outside of the shaded region in Figure 1, not matching early $\mu = \emptyset$ is sequentially stable.

3.2 Example 2: failure of sequential stability

We now assume $n = 2$ (and $k = 1$ as before) in the fixed-transfer regime. Let $m_2 > m_1$ be the types of the two men in the first period, and let $w_2 > w_1$ be the types of the two women.

The expected utility of man m_i in the second period, $1 \leq i \leq 2$, is (after normalizing $U(w \mid m) = \frac{1}{2}mw$ to w):

$$\left(m_i w_i + (1-m_i)(1-w_i)\right) w_i + (1-m_i)\left((w_-)^2 + \int_{w_-}^{w_i} x\,dx\right) + m_i\left((1-w_+)w_+ + \int_{w_i}^{w_+} x\,dx\right), \tag{22}$$

assuming that first-period women w_- and w_+ ($w_- < w_i < w_+$) are also present in the second period (if w_i is the highest among women waiting for the second period, then let $w_+ = 1$; likewise for $w_- = 0$). For example, if $i = 2$, then we always have $w_+ = 1$, with $w_- = w_1$ when w_1 waits for period 2, and with $w_- = 0$ when w_1 matches early (which reduces to Example 1). The first term in (22) represents the events in which m_i is assortatively matched to w_i in the second period, the second term represents events in which m_i is assortatively matched to a worse type than w_i, and the last term represents the events in which m_i is assortatively matched to a better type than w_i.

It is simple to show (this is a special case of the integration-by-part derivation in Page 47) that m_i strictly prefers to match early with w_i, i.e., w_i strictly dominates (22), if and only if:

$$(1-m_i)((w_i)^2 - (w_-)^2) > m_i((1-w_i)^2 - (1-w_+)^2), \tag{23}$$

where the left-hand side represents m_i's downside (being matched downward) in the second period, and the right-hand side represents his upside (being matched upward). Consistent with our intuition, the downside increases with the difference between w_i and w_-, and the upside increases with the difference between w_+ and w_i.

Symmetrically, w_i prefers to match early with m_i if and only if:

$$(1-w_i)((m_i)^2 - (m_-)^2) > w_i((1-m_i)^2 - (1-m_+)^2), \tag{24}$$

where m_- and m_+ are the types of men waiting to the second period who are just below and above m_i.

Let us specialize to the following realization of types in the first period and consider four cases:

$$m_1 = w_1 = \frac{2}{5}, \qquad m_2 = w_2 = \frac{3}{5}.$$

1. Consider the option of the top pair (m_2, w_2) waiting for the second period. If this is the case, for the incentives of the bottom pair we have: $m_1 = w_1 = 2/5$, $w_- = 0$ and $w_+ = w_2 = 3/5$, so inequality (23) holds, i.e., the bottom man's downside in the second period dominates his upside, and likewise for the woman. Therefore, the bottom pair

has a strict incentive to match early. Intuitively, the presence of the top pair in the second period imposes strong competition for the bottom pair and caps their upside: they can get at most 3/5 in the assortative matching of the second period.

2. If the bottom pair indeed matches early, for the incentives of the top pair we have $m_2 = w_2 = 3/5$, $w_- = 0$ and $w_+ = 1$, so again inequality (23) holds, i.e., the top man's downside in the second period dominates his upside, and likewise for the woman. Therefore, the top pair also wants to match early. An alternative way to see this is to return to Example 1, with $n = k = 1$, and notice that the pair $(3/5, 3/5)$ is in the shaded, early-matching region of Figure 1.

3. If indeed the top pair matches early, for the incentives of the bottom pair we have $m_1 = w_1 = 2/5$, $w_- = 0$ and $w_+ = 1$, so now inequality (23) does not hold, i.e, the bottom man's upside in the second period now dominates his downside, and likewise for the woman. Therefore, the bottom pair now wants to wait to the second period. And indeed in Figure 1 of Example 1, we see that the pair $(2/5, 2/5)$ is not in the shaded, early-matching region.

4. Finally, if the bottom pair waits for the second period, for the incentives of the top pair we have $m_2 = w_2 = 3/5$, $w_- = w_1 = 2/5$ and $w_+ = 1$, so inequality (23) does not hold, i.e., the top man's upside in the second period dominates his downside, and likewise for the woman. Therefore, the top pair also prefers to wait to the second period. Intuitively, the bottom pair's presence in the second period provides insurance (fallback options) for the top pair: they are guaranteed of at least 2/5 in the second period.

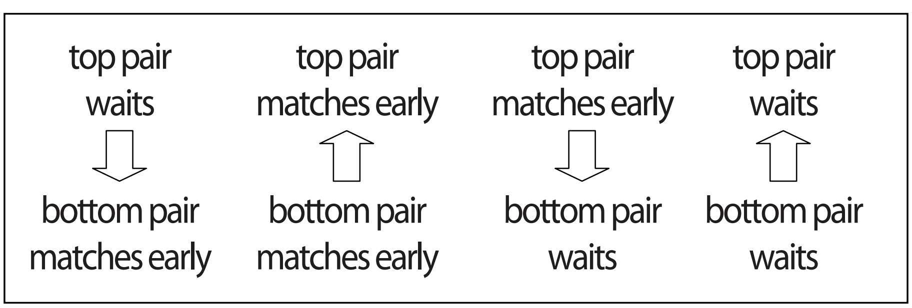


Figure 2: Failure of sequential stability.

As displayed in Figure 2, where the first arrow corresponds to case 1, etc., each choice of one pair implies another by the other pair, but none of these choices is consistent with each other. In other words, no assortative early matching is sequentially stable in the sense of Definition 2. To complete the argument, we must also consider early matching of cross matches between higher-ranked and lower-ranked agents. It is easy to see that an early matching consisting of a single pair of cross match (say, $\mu = \{(m_1, w_2)\}$) is never profitable to the higher-ranked agent in the cross match (i.e., woman w_2, because she can do no worse than m_1 in the second period). Furthermore, the early matching of two cross matches $\mu = \{(m_1, w_2), (m_2, w_1)\}$ is not sequentially stable: one can check that in this example if the early matching $\mu = \{(m_1, w_2), (m_2, w_1)\}$ were sequentially stable, then the early matching $\mu = \{(m_1, w_1), (m_2, w_2)\}$ would be sequentially stable as well[12] (it is not, as shown by Case 3 in page 17).

Therefore, realization of $m_1 = w_1 = 2/5, m_2 = w_2 = 3/5$ does not admit a sequentially stable matching scheme.

3.3 Example 3: flexible transfers

Example 1 and 2 are set in the fixed-transfer regime. We now demonstrate in the context of these examples how flexible transfers can eliminate the incentives to match early and restore sequential stability.

In the flexible-transfer regime, an agent's transfer is consistent with his outside options, i.e., what he can get when he deviates from the assortative matching. For simplicity, we assume that the woman-optimal stable matching is implemented in period 2; see Example 4 on the role of the woman-optimal stable matching.

Let us first work in the setting of Example 1: $n = k = 1$. Consider first the case where the two first-period agents wait to the second period, and let $m_2^2 > m_1^2$ and $w_2^2 > w_1^2$ be the types in the second period. Recall that in this example the total surplus created by man m and woman w is $U(w \mid m) + V(m \mid w) = mw$. Then the woman optimal stable matching in period 2 is the assortative matching of (m_i^2, w_i^2), $i \in \{1, 2\}$, together with man m_1^2 getting

$$\pi_1 = 0 \tag{25}$$

[12]If m_2 has an incentive to match early with w_1, then m_1 must also have an incentive to match early with w_1, since m_2 is in a strictly better position than m_1 (formally, let $w_i = w_1$, $w_- = 0$ and $w_+ = 1$, if (23) holds for $m_i = m_2$, then it also holds for $m_i = m_1$). Moreover, if m_2 has incentive to match early with w_1, then he must also have incentive to match early with a better woman w_2. And likewise for the women.

(after netting his transfer), man m_2^2 getting

$$\pi_2 = (m_2^2 - m_1^2) \cdot w_1^2, \tag{26}$$

and woman w_i^2 getting

$$\Pi_i = m_i^2 \cdot w_i^2 - \pi_i, \tag{27}$$

for $i \in \{1, 2\}$. Intuitively, in the woman optimal stable matching each man m_i^2 gets minimally his outside option which is matching with woman w_{i-1}^2 and displacing man m_{i-1}^2. Thus, in (26) man m_2^2's outside option is to match with woman w_1^2, producing $m_2^2 \cdot w_1^2$ and getting $m_2^2 \cdot w_1^2 - \Pi_1$ (leaving Π_1 to w_1^2 to ensure that she would agree to displace man m_1^2).

By construction, we have the following "no blocking" condition:

$$\pi_i + \Pi_j \geq m_i^2 \cdot w_j^2, \tag{28}$$

with a strict inequality when $i = 1$ and $j = 2$. [13]

Inequality (28) eliminates the early matching incentives identified in Example 1 and 2. To see this, let m and w denote the types of man and woman present in period 1, and suppose that they wait to period 2, so $m = m_i^2$ and $w = m_j^2$, where the indices $i, j \in \{1, 2\}$ depend on the realization of the new arrivals' types. Therefore, man m and woman w have a joint incentive to match early if and only if

$$m \cdot w \geq \mathbb{E}[\pi_i + \Pi_j], \tag{29}$$

where the expectation is taken over the new arrivals' types (which determine the values of π_i and Π_j). By refusing to match early man m gets an expected payoff of $\mathbb{E}[\pi_i]$ in period 2, and woman w an expected payoff of $\mathbb{E}[\Pi_j]$. Therefore, if (29) holds, there exists an division of $m \cdot w$ of which both m and w prefer over their expected payoffs; clearly, the converse also holds. Inequality (29) is the analogue of inequalities (20) and (21) in the flexible-transfer regime.

However, inequality (28) implies that inequality (29), the incentive to match early, cannot hold if $m < 1$ and $w > 0$: for every realization of types in period 2 inequality (28) implies that $\pi_i + \Pi_j \geq m_i^2 \cdot w_j^2 = m \cdot w$, with a strict inequality when $i = 1$ and $j = 2$ (which happens with a positive probability when $m < 1$ and $w > 0$). Therefore, man m and woman w

[13] By construction, we have $\pi_2 + \Pi_1 = m_2^2 \cdot w_1^2$. This implies that $\pi_1 + \Pi_2 = 0 + (m_2^2 \cdot w_2^2 - (m_2^2 - m_1^2) \cdot w_1^2) > m_1^2 \cdot w_2^2$.

cannot have a joint incentive to match early. Intuitively, the flexibility of transfers mitigate the undesirable (to the high type) event of a high type agent being matched to a low type in period 2 by permitting the high type to extract from the low type a large share of their joint surplus in such event.

Clearly, the previous analysis extends to Example 2 with $n = 2$ and $k = 1$, which shows that flexible transfers restore sequential stability: since nobody has an incentive to match early, the outcome that everyone participates in period 2 is sequentially stable.

4 Fixed Transfers

4.1 Prevalence of early matching incentives

Example 1 reveals that in the fixed-transfer regime there is a strictly positive probability of types with incentive to match early and skip the assortative matching of the second period. In this subsection we analyze the prevalence of these early matching incentives as the number of agents gets large. As a measure of prevalence we fix the benchmark of all first-period agents waiting for the second period and count how many of them have an incentive to deviate from this benchmark by matching early.

Formally, fix the empty set of early matching $\mu = \emptyset$ in Definition 2 of sequential stability, and let ξ_n be the number of first-period pairs (m_i, w_i) for whom the incentive to *not* match early (Inequality (11) in Definition 2) is violated:

$$\mathbf{U}(L, m_i) < U(w_i \mid m_i) \text{ and } \mathbf{V}(L, w_i) < V(m_i \mid w_i),$$

where L is the set of all agents in the first period, and $\mathbf{U}(L, m_i)$ and $\mathbf{V}(L, w_i)$ are, respectively, m_i and w_i's expected utilities from the second period (see Equation (9) and (10)). That is, $\xi_n = \sum_{i=1}^{n} \mathbf{1}_{\{\mathbf{U}(L,m_i)<U(w_i|m_i) \text{ and } \mathbf{V}(L,w_i)<V(m_i|w_i)\}}$.

Note that when counting ξ_n we restrict to man and woman of the same rank in the first period. Therefore, ξ_n is a lower bound on all pairs of man and woman with an incentive to match early.

Theorem 1. *Fix $k \geq 1$ pairs of new arrivals. In the fixed-transfer regime, $\mathbb{E}[\xi_n]/n$ tends to $1/4$ as the number n of first-period agents tends to infinity.*

Proof. See Section A.2. □

Theorem 1 says that with fixed transfers a lot of first-period agents have incentives to deviate from the stable matching in period 2, so the empty early matching $\mu = \emptyset$ is far from being sequentially stable. The intuition for Theorem 1 is simple. Consider a pair (m_i, w_i) in the first period — because of the new arrivals, m_i could either do better or do worse than w_i in the assortative matching of the second period. Because m_i and w_i are of the same percentile in the distribution, the probability that m_i would do better in the second period is approximately the probability that he would do worse. Therefore, m_i has an incentive to match early with w_i if and only if w_i is closer to the higher end of her surrounding (women that m_i could be matched to when he does better) than to the lower end (women that m_i could be matched to when he does worse); in large markets, this event happens with probability 1/2. A symmetric argument works for w_i: with probability 1/2 she has an incentive to match early with m_i. Therefore, the ex-ante probability is $1/4 = 1/2 \times 1/2$ that both m_i and w_i have incentives to match early. We then reinterpret the ex-ante probability of 1/4 as the expected fraction of man-woman pairs with an incentive to match early.

4.2 Failure of sequential stability

Theorem 1 implies that in the fixed-transfer regime all first-period agents going to the assortative matching in the second period cannot be sequentially stable. It leaves open the possibility that some other arrangement of early matching would be sequentially stable. We now show that this is unlikely in large markets:

Theorem 2. *Assume a single pair $k = 1$ of new arrivals. Then in the fixed-transfer regime, the probability that the first-period types do not admit a sequentially stable matching scheme tends to 1 as the number n of first-period agents tends to infinity.*

Proof. See Section 4.2.2 and Section A.3. □

Theorem 2 says that with fixed transfers and as the market gets large, with probability tending to 1 *any* early-matching arrangement (including no early matching) is not sequentially stable: anticipating an assortative matching in the second period, either an individual man or woman would have an incentive to deviate from his/her early matching by waiting for the second period, or a pair would have incentives to jointly deviate from waiting for the second period by matching early. The probability of sequential stability vanishes with the market size n, despite the fact that the number of possible early matchings grows exponentially with n.

The instability in Theorem 2 is driven by the interdependence in the first-period agents' decisions to match early or to wait for the second period. When the higher-ranked agents wait they act as strong competitors to the lower-ranked agents, which caps the upside from waiting (being matched upward in period 2) for the lower-ranked agents, thus making the lower-ranked agents less willing to wait. On the other hand, when the lower-ranked agents wait they provide insurance (fallback options) against negative outcome (being matched downward in period 2) for the higher-ranked agents, thus making the higher-ranked agents more willing to wait. Thus, the higher-ranked and the lower-ranked agents essentially engage in a "game" of matching pennies, in which the higher-ranked agents want to synchronize their time of matching with the lower-ranked agents, while the lower-ranked agents do not want such synchronization.[14] As the number of agents increases these interdependencies become difficult to reconcile. Therefore, a seemingly minor informational perturbation in the form of one pair of arriving agents in the second period has a large rippling effect on the intertemporal stability of the matching market, *especially* when the market is thick. Obviously, with more pairs of arrivals the first-period agents care even more about being matched up or down in period 2, so we expect the same effect to be present; this is indeed confirmed by Monte Carlo simulations of the next section.

4.2.1 Monte Carlo simulation

We use Monte Carlo simulations to understand how quickly the probability of sequential stability converge to 0 with n (the number of existing men/women) when $k = 1$ (the number of arrivals). Moreover, simulations can tell us what happens when agents have large uncertainty about the second period, i.e., when k is on the same magnitude as n, an interesting case for which we currently do not have any analytical result. For simplicity, we consider uniform distribution of types (on the interval $[0, 1]$) $F = G = \mathrm{U}[0, 1]$, and multiplicative value function $U(w \mid m) = V(m \mid w) = mw/2$; this gives closed-form expressions for agents' expected utility from period 2 (see Appendix C).[15] We plot the results of the Monte Carlo simulation in Figure 3. For $k = 1$, the probability of sequential stability rapidly decreases to 32% when $n = 20$. When $k = n$, the probabilities of sequential stability are slightly larger and also decrease with n and k; when $n = k = 11$ the probability of sequential stability is around 84%. Figure 3 shows that Theorem 2 is not merely a result in "asymptopia" and that sequential stability fails with a non-trivial probability even when n is small. Moreover,

[14] See Example 2 in Section 3.2 for a concrete illustration of this "game" of matching pennies.

[15] The Monte Carlo simulation code is available at `http://www.sfu.ca/~songzid/probstablemc.nb`.

Figure 3 suggests that sequential instability persists when $k = n$.

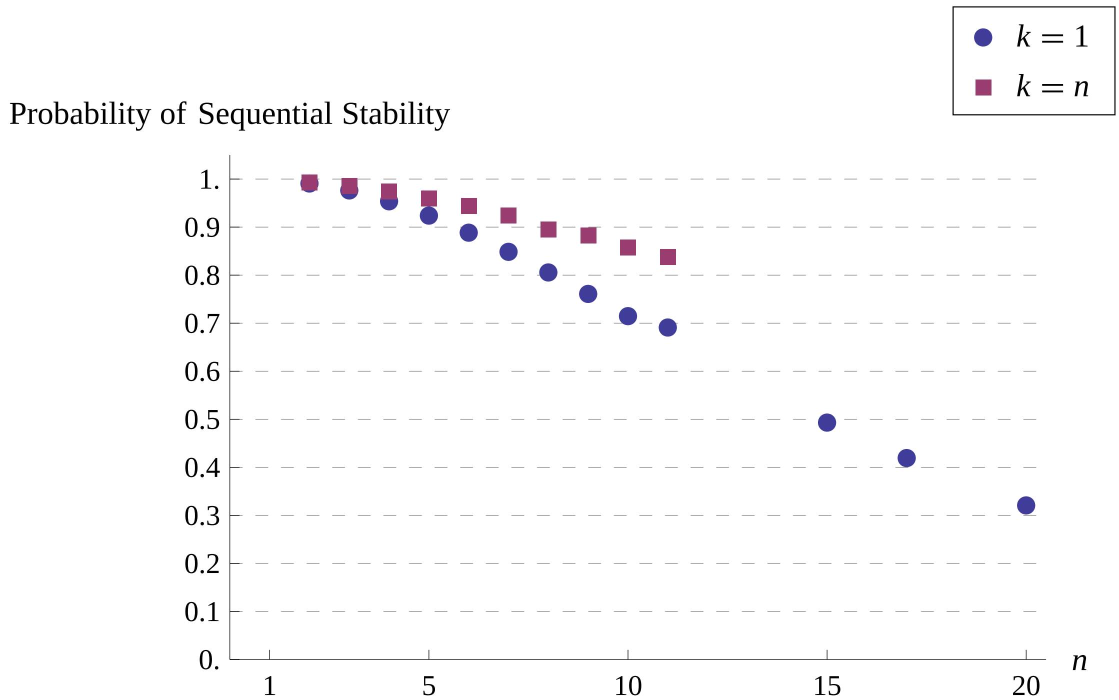


Figure 3: Monte Carlo computation of the probability that a sequentially stable matching scheme exists in the fixed-transfer regime, given $F = \mathrm{U}[0,1] = G$ and $U(w \mid m) = mw/2 = V(m \mid w)$.

4.2.2 Main Step in the Proof of Theorem 2

To illustrate why Theorem 2 is true, we now sketch the main step of the proof in which we study sequential stability in a related and simpler model; let us call this model the exponential model.

In the exponential model with parameter r, there are $r+1$ men and $r+1$ women in period 1, with types $m_r > m_{r-1} > \cdots > m_0$ and $w_r > w_{r-1} > \cdots > w_0$. We assume that the differences in consecutive types, $m_{i+1} - m_i$ and $w_{i+1} - w_i$, $0 \leq i \leq r-1$, are i.i.d. exponential random variables (with mean 1). Moreover, we assume that agents m_r, w_r, m_0 and w_0 always wait for period 2, and we study the early matching decisions of agents m_i and w_i, $1 \leq i \leq r-1$. We restrict attention to early matching that is assortative: any assortative early matching can be represented by a subset $I \subseteq \{1, 2, \ldots, r-1\}$ that lists the indices of agents who choose *not* to match early. Finally, we assume that the early matching incentives are given by the following definition of sequential stability:

Definition 5 (Sequential Stability in the Exponential Model)**.** Fix a realization of m_i and w_i, $0 \le i \le r$. The early matching in which a subset $I = \{i_l : 1 \le l \le L\}$ wait for period 2, where $i_0 \equiv 0 < i_1 < \cdots < i_L < r \equiv i_{L+1}$, is *sequentially stable* for this realization of types if:

1. for any $i \in I$ (there exists a unique l such that $i = i_l$), either woman w_i has no incentive to match early with man m_i in the sense of

$$\underbrace{m_{i_{l+1}} - m_i}_{\text{upside of waiting}} \ge \underbrace{m_i - m_{i_{l-1}}}_{\text{downside of waiting}} \quad , \tag{30}$$

or man m_i has no incentive to match early with woman w_i in the sense of

$$\underbrace{w_{i_{l+1}} - w_i}_{\text{upside of waiting}} \ge \underbrace{w_i - w_{i_{l-1}}}_{\text{downside of waiting}} \quad ; \tag{31}$$

2. for any $i \notin I$ (there exists a unique l such that $i_l < i < i_{l+1}$), the pair (m_i, w_i) both have incentives to match early with each other:

$$\underbrace{m_{i_{l+1}} - m_i}_{\text{upside of waiting}} \le \underbrace{m_i - m_{i_l}}_{\text{downside of waiting}} \quad , \text{ and } \quad \underbrace{w_{i_{l+1}} - w_i}_{\text{upside of waiting}} \le \underbrace{w_i - w_{i_l}}_{\text{downside of waiting}} \quad . \tag{32}$$

To understand Definition 5, consider the first point, which says that man m_{i_l} has an incentive to match early with woman w_{i_l} if and only if the downside $w_{i_l} - w_{i_{l-1}}$ from waiting for period 2 exceeds the upside $w_{i_{l+1}} - w_{i_l}$ from waiting. The expression of upside and downside as the difference in types comes from the assumptions of $k = 1$ and of a large market of which the agents $m_0, \ldots, m_r$, $w_0, \ldots, w_r$ are a part, so with large probability m_{i_l} is either matched to $w_{i_{l+1}}$ (when the new man is worse than m_{i_l} and the new woman is better than $w_{i_{l+1}}$) or to $w_{i_{l-1}}$ (when the new man is better than m_{i_l} and the new woman is better than $w_{i_{l-1}}$). (We formalize this point in Section A.3.) Since when deciding about early matching the agents simply compare the differences in types in their surrounding, in the exponential model it is without loss of generality to restrict to assortative, sequentially stable early matching: any non-assortative early matching that is sequentially stable can be "uncrossed" and converted into an assortative early matching that is also sequentially stable (types in their surrounding are not affected by the uncrossing). Finally in a large market, the differences of order statistics $m_{i+1} - m_i$ and $w_{i+1} - w_i$, $0 \le i \le r-1$, can be normalized to be i.i.d. exponential (Lemma 4); see also Pyke (1965).

Definition 5 critically assumes that agents m_0, m_r, w_0 and w_r exogenously wait for period 2. In Section A.3 we justify this assumption by proving that in large markets it is unlikely to have a large number of consecutively ranked pairs all choosing to match early, so in any sequentially stable matching scheme agents who wait (candidates for m_0, m_r, w_0 and w_r) can be found everywhere in the support of type.

Let $\hat{\mathbb{P}}_r$ be the probability measure for the random variables m_i and w_i with exponentially distributed differences. For any $I \subseteq \{1, 2, \ldots, r-1\}$, let G_I denote the event (i.e., a set of realizations of m_i and w_i) that the early matching implied by I is sequentially stable in the sense of Definition 5; see Figure 4 for an example of G_I. We now show the probability of a sequentially stable I tends to zero with r in the exponential model:

$$\lim_{r\to\infty} \hat{\mathbb{P}}_r \left(\bigcup_{I\subseteq\{1,2,\ldots,r-1\}} G_I \right) = 0. \tag{33}$$

We first exploit the memoryless property of the exponential distribution to simplify the expression for each individual $\hat{\mathbb{P}}_r(G_I)$:

Lemma 1. *Suppose that $I = \{i_l : 1 \le l \le L\}$, where $i_0 \equiv 0 < i_1 < \cdots < i_L < r \equiv i_{L+1}$. Then we have:*

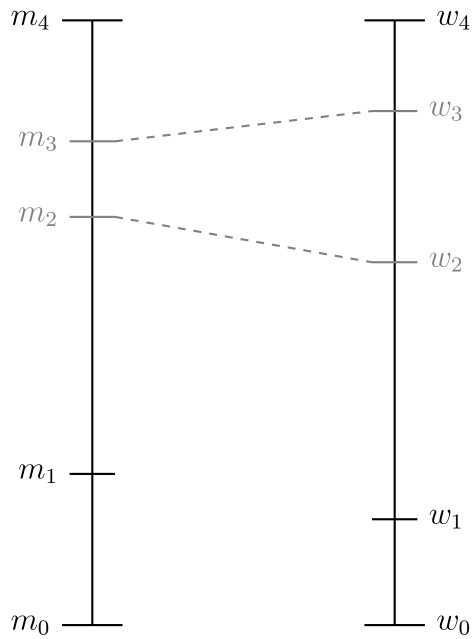
$$\hat{\mathbb{P}}_r(G_I) = \frac{1}{4^{r-L-1}} \hat{\mathbb{P}}_r \left(\bigcap_{l=1}^{L} \left(\{m_{i_{l+1}} - m_{i_l} \ge m_{i_l} - m_{i_{l-1}}\} \cup \{w_{i_{l+1}} - w_{i_l} \ge w_{i_l} - w_{i_{l-1}}\} \right) \right). \tag{34}$$

Proof. In Point 2 of Definition 5, if $i_{l+1} > 2+i_i$, then we only need to check the early matching incentives for the pair (m_{1+i_l}, w_{1+i_l}), since if (m_{1+i_l}, w_{1+i_l}) have incentives to match early, then so do (m_{2+i_l}, w_{2+i_l}). Thus, we have

$$\begin{aligned}\hat{\mathbb{P}}_r(G_I) = \hat{\mathbb{P}}_r \Bigg(& \left(\bigcap_{l=0}^{L} \{m_{i_{l+1}} - m_{1+i_l} \le m_{1+i_l} - m_{i_l}\} \cap \{w_{i_{l+1}} - w_{1+i_l} \le w_{1+i_l} - w_{i_l}\} \right) \\ & \bigcap \left(\bigcap_{l=1}^{L} \left(\{m_{i_{l+1}} - m_{i_l} \ge m_{i_l} - m_{i_{l-1}}\} \cup \{w_{i_{l+1}} - w_{i_l} \ge w_{i_l} - w_{i_{l-1}}\} \right) \right) \Bigg).\end{aligned}$$

The exponential distribution has no "memory," that is, conditional on the event $m_{i_{l+1}} - m_{1+i_l} \le m_{1+i_l} - m_{i_l}$, the random variable $m_{i_{l+1}} - m_{i_l}$ has the same distribution as its

Figure 4: An example of early matching (indicated by the grey dashed lines) in the exponential model with $r = 4$. Those who wait for period 2 have indices in $I = \{1\}$. $G_I = \{m_4 - m_2 \leq m_2 - m_1\} \cap \{w_4 - w_2 \leq w_2 - w_1\} \cap (\{m_4 - m_1 \geq m_1 - m_0\} \cup \{w_4 - w_1 \geq w_1 - w_0\})$. Lemma 1 says that $\hat{\mathbb{P}}_r(G_I) = \hat{\mathbb{P}}_r(\{m_4 - m_1 \geq m_1 - m_0\} \cup \{w_4 - w_1 \geq w_1 - w_0\})/16$.



unconditional distribution,[16] and likewise for w_i's. Therefore, we have

$$
\begin{aligned}
\hat{\mathbb{P}}_r(G_I) =& \prod_{l=0}^{L} \hat{\mathbb{P}}_r \left(\{m_{i_{l+1}} - m_{1+i_l} \leq m_{1+i_l} - m_{i_l}\} \right) \hat{\mathbb{P}}_r \left(\{w_{i_{l+1}} - w_{1+i_l} \leq w_{1+i_l} - w_{i_l}\} \right) \\
& \cdot \hat{\mathbb{P}}_r \left(\bigcap_{l=1}^{L} \left(\{m_{i_{l+1}} - m_{i_l} \geq m_{i_l} - m_{i_{l-1}}\} \cup \{w_{i_{l+1}} - w_{i_l} \geq w_{i_l} - w_{i_{l-1}}\} \right) \right) \\
=& \frac{1}{4^{r-L-1}} \cdot \hat{\mathbb{P}}_r \left(\bigcap_{l=1}^{L} \left(\{m_{i_{l+1}} - m_{i_l} \geq m_{i_l} - m_{i_{l-1}}\} \cup \{w_{i_{l+1}} - w_{i_l} \geq w_{i_l} - w_{i_{l-1}}\} \right) \right),
\end{aligned}
$$

[16] Suppose that $a_1, a_2, \ldots, a_i$ are i.i.d. exponential random variables. Conditioning on the event that $a_1 \geq a_2 + \ldots + a_n$, the random variable $\sum_{j=1}^{n} a_j$ has conditional density function $h(z)$, where (note that $\mathbb{P}(a_1 \geq a_2 + \ldots + a_n) = 1/2^{n-1}$):

$$
h(z) = \frac{1}{1/2^{n-1}} \int_{a_1=z/2}^{z} \exp(-a_1) \frac{(z-a_1)^{n-2} \exp(-(z-a_1))}{(n-2)!} \, da_1 = \frac{z^{n-1} \exp(-z)}{(n-1)!}
$$

which is the unconditional density function of $\sum_{j=1}^{n} a_j$.

where the second line follows from Lemma 5 which implies that

$$\hat{\mathbb{P}}_r\left(\{m_{i_{l+1}} - m_{1+i_l} \leq m_{1+i_l} - m_{i_l}\}\right) = \hat{\mathbb{P}}_r\left(\{w_{i_{l+1}} - w_{1+i_l} \leq w_{1+i_l} - w_{i_l}\}\right) = \frac{1}{2^{i_{l+1}-i_l-1}}.$$

$\square$

In light of Lemma 1, for any $r \geq 2$ and any $I = \{i_l : 1 \leq l \leq L\}$, where $i_0 \equiv 0 < i_1 < \cdots < i_L < r \equiv i_{L+1}$, we define

$$\pi_r(I) \equiv \frac{1}{4^{r-L-1}}\hat{\mathbb{P}}_r\left(\bigcap_{l=1}^{L}\left(\{m_{i_{l+1}} - m_{i_l} \geq m_{i_l} - m_{i_{l-1}}\} \cup \{w_{i_{l+1}} - w_{i_l} \geq w_{i_l} - w_{i_{l-1}}\}\right)\right) \quad (35)$$

and

$$\pi_r \equiv \sum_{I \subseteq \{1,2,\ldots,r-1\}} \pi_r(I). \quad (36)$$

Lemma 2. *For any $I \neq I' \subseteq \{1, \ldots, r-1\}$, we have $\hat{\mathbb{P}}_r(G_I \cap G_{I'}) = 0$. Hence we have*

$$\hat{\mathbb{P}}_r\left(\bigcup_{I \subseteq \{1,2,\ldots,r-1\}} G_I\right) = \pi_r. \quad (37)$$

The proof of Lemma 2 is a bit involved so we defer it to the appendix (page 40).

Proposition 1. *π_r tends to 0 as $r \to \infty$.*

Proof. For $r \geq 2$, we have (with the convention that $\pi_0 = \pi_1 \equiv 1$):

$$\pi_r \leq \pi_{\lfloor r/2\rfloor}\pi_{\lceil r/2\rceil} + \sum_{i=1}^{\lfloor r/2\rfloor}\sum_{j=1}^{\lceil r/2\rceil}\frac{1}{4^{i+j-1}}\pi_{\lfloor r/2\rfloor-i}\pi_{\lceil r/2\rceil-j}, \quad (38)$$

where we first sum $\pi_r(I)$ over $I \subseteq \{1, \ldots, r-1\}$ such that $\lfloor r/2\rfloor \in I$, and then sum $\pi_r(I)$ over $I \subseteq \{1, \ldots, r-1\}$ such that $\lfloor r/2\rfloor \notin I$ and the closest elements in $I \cup \{0, r\}$ to $\lfloor r/2\rfloor$ are $\lfloor r/2\rfloor - i$ and $\lfloor r/2\rfloor + j$; we use the fact that for $I' = \{i_1, i_2, \ldots, i_l\}$ and $I'' = \{i_{l+1}, i_{l+2}, \ldots, i_L\}$ with $0 < i_1 < i_2 < \cdots < i_l \leq i_{l+1} < i_{l+2} < \cdots < i_L < r$, we have

$$\pi_r(I' \cup I'') \leq \begin{cases} \pi_{i_l}(I' - \{i_l\}) \cdot \pi_{r-i_{l+1}}(I'' - \{i_{l+1}\}) & i_l = i_{l+1} \\ \pi_{i_l}(I' - \{i_l\}) \cdot \pi_{r-i_{l+1}}(I'' - \{i_{l+1}\})/4^{i_{l+1}-i_l-1} & i_l < i_{l+1} \end{cases}. \quad (39)$$

Therefore, for any $l \leq \lfloor r/2 \rfloor$ we have

$$\begin{aligned}\pi_r &\leq \pi_{\lfloor r/2 \rfloor}\pi_{\lceil r/2 \rceil} + \sum_{i=1}^{l}\sum_{j=1}^{l}\frac{1}{4^{i+j-1}}\pi_{\lfloor r/2 \rfloor - i}\pi_{\lceil r/2 \rceil - j} + \sum_{i=1}^{l}\sum_{j=l+1}^{\infty}\frac{1}{4^{i+j-1}} + \sum_{i=l+1}^{\infty}\sum_{j=1}^{\infty}\frac{1}{4^{i+j-1}} \\ &\leq \frac{13}{9}\max_{-1\leq j\leq l}\pi^2_{\lfloor r/2 \rfloor - j} + \frac{8}{9}\left(\frac{1}{4}\right)^l, \end{aligned} \tag{40}$$

where we use Lemma 2 which implies that $\pi_{\lfloor r/2 \rfloor - i}$ and $\pi_{\lceil r/2 \rceil - j}$ are always less than or equal to 1.

We can directly calculate π_r when r is small. We use the following facts from our calculations: $\pi_7 \approx 0.595$ and $\pi_r < 0.595$ for $8 \leq r \leq 17$.

Define the function

$$H_l(y) = \frac{13}{9}y^2 + \frac{8}{9}\left(\frac{1}{4}\right)^l, \tag{41}$$

and let $(H_l)^i$ be H_l iterated i times.

First let $l = 2$. When y is in between 0.0609154 and 0.631392 which are the two solutions to $H_l(y) = y$, we have $H_l(y) < y$ and $\lim_{i\to\infty}(H_l)^i(y) = 0.0609154$. By Equation (40) and our calculations of π_r for $7 \leq r \leq 17$, we have $\pi_r < H_l(0.595)$ for $r \geq (7+2)\times 2 = 18$; $\pi_r < (H_l)^2(0.595)$ for $r \geq (18+2)\times 2 = 40$; $\pi_r < (H_l)^3(0.595)$ for $r \geq (40+2)\times 2 = 84$; and so on. Therefore, we have $\limsup_{r\to\infty}\pi_r \leq \lim_{i\to\infty}(H_l)^i(0.595) = 0.0609154$.

We then assume progressively larger values of l and use similar reasoning in the above paragraph to conclude that $\limsup_{r\to\infty}\pi_r \leq 0$. □

Proposition 1 implies that the probability of sequential stability vanishes with r in the exponential model. Section A.3 leverages this fact to prove that the probability of sequential stability vanishes with n in our actual model.

5 Flexible Transfers

In this section we supplement the fixed-transfer model with transfers that are endogenously determined as a part of the stable matching. We show that flexible transfers eliminate the incentives to match early and restore sequential stability.

Theorem 3. *Consider any $n \geq 1$ and $k \geq 1$ in the flexible-transfer regime. For any selection of stable matching (μ^2, P^2) in the second period, the matching scheme consisting of no early matching $(\mu = \emptyset, \mu^2, P^2)$ is sequentially stable for every realization of first-period* *types. Moreover, when (μ^2, P^2) is always the woman-optimal stable matching, for almost every realization of first-period types (μ, μ^2, P^2) is sequentially stable if and only if $\mu = \emptyset$.*[17]

Proof. Fix a realization of types in the first period, consider no early matching: $\mu = \emptyset$. For any man m_i and woman w_j in the first period, and for any type realization of the new arrivals, suppose that m_i is ranked r-th and w_j is ranked s-th in the second period (among the new arrivals and the first-period agents), i.e., $m_i = m_r^2$ and $w_j = w_s^2$. By the assumption that a stable matching (μ^2, P^2), where μ^2 is assortative, is implemented in the second period, we have the no-blocking condition:

$$\begin{aligned}\big(U(w_r^2 \mid m_r^2) + P^2(m_r^2, w_r^2)\big) + \big(V(m_s^2 \mid w_s^2) - P^2(m_s^2, w_s^2)\big) &\geq U(w_s^2 \mid m_r^2) + V(m_r^2 \mid w_s^2) \\ &= U(w_j \mid m_i) + V(m_i \mid w_j) \quad (42)\end{aligned}$$

Take expectation over the type realization of the new arrivals in (42) gives (where L here is the list of all first-period agents since $\mu = \emptyset$):

$$\mathbf{U}(L, m_i; P^2) + \mathbf{V}(L, w_j; P^2) \geq U(w_j \mid m_i) + V(m_i \mid w_j), \tag{43}$$

i.e., man m_i and woman w_j do not have a joint incentive to match early. This proves the first part of the theorem.

Now suppose (μ^2, P^2) is always the woman-optimal stable matching. Let $L \supseteq \{m_i, w_j\}$ now be a list of first-period agents who wait for period 2. If the first-period types are all distinct (happens with probability 1), then inequality (42) is strict unless $r = s + 1$ or $r = s$ (where m_i is ranked r-th and w_j is ranked s-th in the second period), which fails to occur with a positive probability over the new arrivals. Thus, inequality (43) is also strict. Thus, no pair of first-period agents can have a joint incentive to match early (inequality (19) in Definition 4 cannot hold). □

Theorem 3 presents a sharp contrast to the results from the fixed-transfer regime: with flexible transfers in period 2, *none* of the agents have an incentive to deviate from the second-period stable matching (in contrast to Theorem 1), which implies that everyone participating in the second-period stable matching is sequentially stable (in contrast to Theorem 2). Moreover, under the woman-optimal implementation of the second-period transfers everyone participating in period 2 is the unique sequentially stable matching scheme; we conjecture

[17] In the second part of Theorem 3 we select (μ^2, P^2) to be the woman-optimal stable matching for every realization of types in period 2 (those who wait from period 1 plus the new arrivals). Clearly, the same result holds if (μ^2, P^2) is always the man-optimal stable matching.

that the same result also holds if the second-period matching is always the median stable matching. Intuitively, the flexible transfers mitigate the risk in period 2: when a high type woman is assortatively matched to a low type man in period 2 (because of strong women and weak men in the new arrivals), the stable matching gives the high type woman a sufficiently large fraction of her joint surplus with the low type man, such that her fraction of surplus dominates all incentives to match early.

Theorem 3 implies that the timing problems identified in Section 4 (the prevalence of early matching incentives and the failure of sequential stability) are caused by the lack of flexible transfers. Theorem 3 thus gives a rationale for the desirability of flexible transfers in matching.

Example 4. We now show that early matching can occur in a sequentially stable matching scheme when the stable matching is not "consistently" implemented in period 2. This example clarifies the role of man/woman-optimal stable matching in Theorem 3. Consider the setting of Example 3 (Section 3.3) with $n = 1$ and $k = 1$. Let (m, w) denote the types of agents in the first period and (m^2, w^2) the new arrivals' types. Instead of always having a woman-optimal stable matching in period 2 as in Example 3, in the situation when man m and woman w wait for period 2, suppose that the man-optimal stable matching is implemented in period 2 when woman w is doing better than man m ($w > w^2$ and $m^2 > m$), and that the woman-optimal stable matching is implemented in period 2 when man m is doing better than woman w ($m > m^2$ and $w^2 > w$). In all other cases, assume an arbitrary stable matching is implemented in period 2 (It does not matter because in these cases man m is assortatively matched to woman w, if they wait for period 2.)

Then, for every realization of the first-period type (m, w), the early matching of $\mu = \{(m, w)\}$ is sequentially stable given the second-period stable matchings described in the previous paragraph: by those specific choices of stable matching in period 2, the total surplus of $m \cdot w$ from matching early is exactly equal to the sum of utilities of man m and woman w in period 2 for every realization of (m^2, w^2); for example, in the woman-optimal stable matching which is implemented when $w < w^2$ and $m^2 < m$, man m is exactly indifferent between matching with w^2 and deviating to match with w (see the specification of transfers in Example 3). Of course, by the first part of Theorem 3 no early matching is also sequentially stable in this example.

5.1 An intermediate regime

In this subsection we analyze an intermediate regime in which flexible transfers are possible in period 1, but not in period 2. The motivation is that while some institutional constraint may impose a rigid schedule of transfers in the formal matching institution (period 2), if agents match outside of the formal institution they may be free to negotiate any division of the surplus (period 1). This intermediate regime is plausible if we interpret early matching as attempts to cheat or to circumvent the formal matching institution of period 2: if agents are cheating then it is likely that they also ignore the institutionally fixed schedule of transfers. We show that flexible transfers in period 1 *alone* is not sufficient to prevent early matching nor to guarantee sequential stability. These results reinforce the importance of the flexible transfers in the formal matching institution (i.e., period 2 in our model). Moreover, flexible transfers in period 1 alone may enable a kind of cross match of higher-ranked woman and lower-ranked man (or vice versa) in early matching that would not be possible with fixed transfers.

In the second period of this intermediate regime, an assortative matching with fixed transfers is implemented — when man m_i^2 is assortatively matched to woman w_i^2 in period 2, man m_i^2 gets $U(w_i^2 \mid m_i^2)$ and woman w_i^2 gets $V(m_i^2 \mid w_i^2)$. Then, given a list L of first-period agents waiting to period 2, man $m_i \in L$ has an expected utility of $\mathbf{U}(L, m_i)$ (defined in (9)) from the assortative matching in period 2, and woman $w_i \in L$ has an expected utility of $\mathbf{V}(L, w_i)$ (defined in (10)). In the first period, a pair of agents can negotiate any division of surplus if they choose to match early. Then, man m_i and woman w_i have a strict incentive to match early if their joint surplus (the amount available to divide) is greater than the sum of their expected utilities from period 2:

$$U(w_i \mid m_i) + V(m_i \mid w_i) > \mathbf{U}(L, m_i) + \mathbf{V}(L, w_i). \tag{44}$$

As before, we first consider the benchmark where all first-period agents wait for the assortative matching in the second period, i.e., let L be the list of all first-period agents in (44). Let ξ_n be the number of first-period pairs (m_i, w_i) with a strict incentive to match early given this benchmark (i.e., inequality (44) holds). We have the following analogue of Theorem 1:

Theorem 4. *Fix $k \geq 1$ pairs of new arrivals. Then in the intermediate regime, the expected fraction of first-period pairs with an incentive to match early, $\mathbb{E}[\xi_n]/n$, tends to $1/2$ as the number n of first-period agents tends to infinity.*

Proof. We omit the proof because it is virtually identical to that of Theorem 1. □

Theorem 4 reveals that flexible transfers in period 1 actually exacerbates the prevalence of incentives to match early: the early matching fraction increases from 1/4 with fixed transfers (Theorem 1) to 1/2 in the this intermediate regime. Intuitively, with flexible transfers in period 1 we no longer need both agents to prefer the early matching (as we do with fixed transfers): one side who is eager to match early now has the option of contributing some transfers to the other side (who might be unwilling otherwise) to "sweeten" the early matching deal. Since we now have one early-matching incentive condition instead of two, we have the early-matching probability of 1/2 instead of 1/4.

Next, we adapt the definition of sequential stability, Definition 2, for this intermediate regime. Since agents anticipate an assortative matching with fixed transfers in period 2, condition (1) in Definition 2 is unchanged. Condition (2) in Definition 2, the incentive for not matching early, is now:

$$\mathbf{U}(L(\mu), m_i) + \mathbf{V}(L(\mu), w_j) \geq U(w_j \mid m_i) + V(m_i \mid w_j), \tag{45}$$

since agents can negotiate transfers in period 1. Similarly, condition (3) in Definition 2, the incentive for matching early, becomes:

$$U(w_j \mid m_i) + V(m_i \mid w_j) \geq \mathbf{U}(L(\mu) \cup \{m_i, w_j\}, m_i) + \mathbf{V}(L(\mu) \cup \{m_i, w_j\}, w_j). \tag{46}$$

Given this definition sequential stability in the intermediate regime, one can show that the realization of types in Example 2 (Section 3.2) still does not admit a sequentially stable matching scheme. Indeed, because of the symmetry of man and woman in Example 2, we have $U(w_i \mid m_i) = V(m_i \mid w_i)$ as well as $\mathbf{U}(L(\mu), m_i) = \mathbf{V}(L(\mu), w_i)$ for any assortative early matching μ, and hence $\mathbf{U}(L(\mu), m_i) + \mathbf{V}(L(\mu), w_i) \geq U(w_i \mid m_i) + V(m_i \mid w_i)$ if and only if $\mathbf{U}(L(\mu), m_i) \geq U(w_i \mid m_i)$. Therefore, the sequential instability of any assortative early matching follows directly from the argument when the transfers are fixed (see Section 3.2). It is also easy to check that any non-assortative early matching is not sequentially stable (for example, the cross match in Example 5 cannot work here). Thus, flexible transfers in period 1 alone cannot guarantee sequential stability. We conjecture that Theorem 2 continues to hold with flexible transfers in period 1.

Finally, we illustrate that flexible transfers in period 1 may enable a kind of cross match between a high-ranked woman and a low-ranked man in early matching that would not be possible with fixed transfers. This example is in the same spirit as Theorem 4 and shows that flexible transfers in period 1 expand the possibilities of early matching.

Example 5. Suppose that $n = 2$, $k = 1$, $F = G = \mathrm{U}[0,1]$ and $U(w \mid m) = V(m \mid w) = mw/2$. Fix the realization of

$$m_2 = \frac{9}{10}, m_1 = \frac{4}{10}, w_2 = \frac{4.2}{10}, w_1 = \frac{1}{10}, \tag{47}$$

in the first period. We claim that the early matching $\mu = \{(m_1, w_2)\}$ is sequentially stable with flexible transfers in period 1. Notice that the early matching $\mu = \{(m_1, w_2)\}$ cannot be sequentially stable with fixed transfers, because of the presence of m_2 and w_1 in period 2 and that $k = 1$, w_2 will always be matched to a type higher than m_1 in period 2 (and potentially even higher than $m_2 = 9/10$), so with fixed transfers woman w_2 will never have an incentive to match early with man m_1. On the other hand, man m_1 is always matched to someone worse than w_2 in period 2 (and potentially even worse than $w_1 = 1/10$), so woman w_2 could leverage her advantage and m_1's disadvantage in period 2 to match early with m_1 and demand a large share of the surplus from this early matching. Indeed with transfers in period 1 this is sequentially stable for the realization of types in (47), and both man m_1 and woman w_2 strictly benefits from this early matching. (Detail can be found in Appendix B.)

One can show that for the realization in (47), $\mu = \{(m_1, w_2)\}$ is the unique sequentially stable early matching with flexible transfers in period 1, and $\mu = \emptyset$ is the unique sequentially stable early matching with fixed transfers. Therefore, fixed transfers may lead to greater efficiency in equilibrium than flexible transfers in period 1 alone.

6 Discussion

6.1 Tradeoffs in Large Markets

When the types of agents are distributed on a bounded interval and given a fixed k, the arrangement in which nobody matches early ($\mu = \emptyset$) becomes approximately stable in the fixed-transfer regime as n tends to infinity, because k pairs of arrivals could displace the existing agents by at most k ranks, and the difference between the types of two consecutively ranked agents tends to 0 at a rate of $O_p(1/n)$ given a bounded type distribution. Theorem 1 and Theorem 2 are of interest for bounded type distribution because they identify novel tradeoffs, as n gets large, between the decrease in the impact of the new arrivals on the existing agents on the one hand, and the increase in the number of agents wanting to matching early and the increase in the probability of sequential instability caused by the new arrivals on the other hand. The main message of our paper is that such tradeoff does not exist when the transfers are flexible.

To understand which of these effects dominates we must study the rates at which these effects appear or vanish as n becomes large. Theorem 1 implies that the expected number of pairs with an incentive to match early increases at least linearly with n. More research is needed to quantify the rate at which the probability of sequential instability tends to 1 with n. The Monte Carlo results presented in Figure 3 suggest that the probability of sequential instability increases to 1 linearly with n when the type distribution is uniform and the values are multiplicative.

Theorem 1 and Theorem 2 are also true for unbounded type distributions. When the agents' types have unbounded distribution and when the surplus function features large complementarities between the high-type agents (e.g., $U(w \mid m) = V(m \mid w) = mw/2$), nobody matching early is not approximately stable under $k = 1$ as n tends to infinity, since even a small increase in the type of the matched partner can generate a large increase in surplus for the high-type agent due to their complementarities. Unbounded types are plausible when the number of agents tends to infinity, since it allows the possibility of exceptional talent/productivity in large population.

Finally, we note that under $k = 1$ and a bounded distribution of types, the decision by a first-period agent to wait for period 2 is not approximately a dominant strategy even when n is large, because waiting for period 2 could be very costly for an agent when others below him have all committed to early matchings, thus exposing this agent to a large risk of being matched to a low type in period 2.

6.2 Interpreting Theorem 2

As we have noted the failure of sequential stability in the fixed transfers regime has a "matching pennies" logic. To sustain an equilibrium in a matching pennies game convexity in the form of mixed strategy is needed. In our setting a natural source of convexity is the transfers between the matched partners, and indeed we show that flexible transfers can sustain sequential stability. So we view Theorem 2 as an argument for introducing flexible transfers to stabilize the market (and to prevent early matching). In practice, it is quite plausible that an employer can promise an employee who proposes early matching to pay the employee more if they are matched in period 2 and if the employee is in high demand then. This kind of promise would introduce flexible transfers in period 2 and would help stabilize the market.

Comparing with transfers, explicit randomization in early matching decisions is less natural in our setting. On the other hand, randomization in early matching decisions could be viewed as due to uncertainty about an agent's type in period 1, in the spirit of Harsanyi's purification theorem. Since in this paper we focus on the role of transfers in early matching, we leave exploring the purification argument to future research.

6.3 Assumption on n and k

For our fixed-transfer results we have assumed a large n in period 1 and a small k in period 2. This assumption corresponds to the natural benchmark in which most of the agents (i.e., those in period 1) *endogenously* choose their time of matching; the small k arrivals in period 2 can be interpreted as a small shock to the market. The case of large n and large k is also interesting and raises intriguing theoretical questions (see Equation (70) in Appendix C for the expected utility with arbitrary n and k). The case of large n and large k could be approximated with continuums of agents in both periods, which is another interesting direction for the future.

References

Alon, N. and J. H. Spencer (2008): *The Probabilistic Method*, Wiley-Interscience, 3 ed.

Avery, C., C. Jolls, R. A. Posner, and A. E. Roth (2007): "The New Market for Federal Judicial Law Clerks," *University of Chicago Law Review*, 74, 447–486.

Becker, G. S. (1973): "A Theory of Marriage: Part 1," *Journal of Political Economy*, 81, 813–846.

Bulow, J. and J. Levin (2006): "Matching and Price Competition," *American Economic Review*, 96, 652–668.

Crawford, V. P. (2008): "The Flexible-Salary Match: A Proposal to Increase the Salary Flexibility of the National Resident Matching Program," *Journal of Economic Behavior and Organization*, 66, 149–160.

Damiano, E., L. Hao, and W. Suen (2005): "Unravelling of Dynamic Sorting," *The Review of Economic Studies*, 72, 1057–1076.

Demange, G., D. Gale, and M. Sotomayor (1986): "Multi-Item Auctions," *Journal of Political Economy*, 94, 863–872.

Echenique, F. and A. Galichon (2014): "Ordinal and cardinal solution concepts for two-sided matching," Working Paper.

Echenique, F. and J. S. Pereyra (2014): "Strategic Complementarities and Unraveling in Matching Markets," *Theoretical Economics*.

Fainmesser, I. P. (2013): "Social Networks and Unraveling in Labor Markets," *Journal of Economic Theory*, 148, 64–103.

Gale, D. and L. S. Shapley (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

Halaburda, H. (2010): "Unravelling in two-sided matching markets and similarity of preferences," *Games and Economic Behavior*, 69, 365–393.

Jaffe, S. and S. D. Kominers (2014): "Taxation in Matching Markets," Working Paper.

Li, H. and S. Rosen (1998): "Unraveling in Matching Markets," *The American Economic Review*, 88, 371–387.

Li, H. and W. Suen (2000): "Risk Sharing, Sorting, and Early Contracting," *Journal of Political Economy*, 108, 1058–1091.

Mailath, G. J., A. Postlewaite, and L. Samuelson (2013): "Pricing and investments in matching markets," *Theoretical Economics*, 8, 535–590.

Niederle, M. and A. E. Roth (2009): "Market Culture: How Rules Governing Exploding Offers Affect Market Performance," *American Economic Journal: Microeconomics*, 1, 199?219.

Niedrele, M., A. E. Roth, and M. U. Unver (2010): "Unraveling Results from Comparable Demand and Supply: An Experimental Investigation," Working Paper.

Ostrovsky, M. and M. Schwarz (2010): "Information Disclosure and Unraveling in Matching Markets," *American Economic Journal: Microeconomics*, 2, 34–63.

Pyke, R. (1965): "Spacing," *Journal of the Royal Statistical Society. Series B (Methodological)*, 27, 395–449.

Roth, A. E. (1991): "A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Marketsfor New Physicians and Surgeons in the United Kingdom," *American Economic Review*, 81, 415–440.

Roth, A. E. and M. A. O. Sotomayor (1992): *Two-Sided Matching: A Study in Game-Theoretic Modeling and Analysis*, Cambridge University Press.

Roth, A. E. and X. Xing (1994): "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Transactions," *The American Economic Review*, 84, 992–1044.

Schwarz, M. and M. B. Yenmez (2011): "Median stable matching for markets with wages," *Journal of Economic Theory*, 146, 619–637.

Shapley, L. S. and M. Shubik (1971): "The assignment game I: The core," *International Journal of Game Theory*, 1, 111–130.

Sonmez, T. (1999): "Can Pre-arranged Matches Be Avoided in Two-Sided Matching Markets?" *Journal of Economic Theory*, 86, 148–156.

——— (2012): "Bidding for Army Career Specialties: Improving the ROTC Branching Mechanism," *Journal of Political Economy*.

Sonmez, T. and T. Switzer (2012): "Matching with (Branch-of-Choice) Contracts at the United States Military Academy," *Econometrica*.

Suen, W. (2000): "A Competitive Theory of Equilibrium and Disequilibrium Unravelling in Two- Sided Matching," *The RAND Journal of Economics*, 31, 101–120.

Wetz, R., C. Seelig, G. Khoueiry, and K. Weiserbs (2010): "Out-of-Match Residency Offers: The Possible Extent and Implications of Prematching in Graduate Medical Education," *Journal of Graduate Medical Education*, 2, 327–333.

Appendix

A Proofs

A.1 Preliminary Lemmas

We use the following lemmas in the proofs of Theorem 1 and Theorem 2.

Lemma 3 (Chernoff)**.**

$$\mathbb{P}\left(\left|F(m_i) - \frac{i}{n}\right| \geq \epsilon\right) \leq 2\exp(-2\epsilon^2 n)$$

for any $\epsilon > 0$.

Proof of Lemma 3. Let $u_1 \leq \ldots \leq u_n$ be the order statistics of n i.i.d. U$[0,1]$ random variables, and set $m_i \equiv F^{-1}(u_i)$ where $F^{-1}(u) \equiv \inf\{x \in \mathbb{R} : F(x) \geq u\}$. Since the density f is everywhere positive, we have $F(m_i) = u_i$.

Clearly,

$$\mathbb{P}(|u_i - i/n| \geq \epsilon) \leq \mathbb{P}(u_i \geq i/n + \epsilon) + \mathbb{P}(u_i \leq i/n - \epsilon).$$

By definition, we have

$$\mathbb{P}(u_i \leq i/n - \epsilon) = \mathbb{P}\left(\sum_{j=1}^{n} \mathbf{1}(z_j \leq i/n - \epsilon) \geq i\right)$$

where $z_1, \ldots, z_n$ are n i.i.d. U$[0,1]$ random variables.

We now apply a standard Chernoff bound to i.i.d. random variables $\mathbf{1}(z_j \leq i/n - \epsilon)$'s (e.g., Alon and Spencer (2008), Theorem A.1.4):

$$\begin{aligned}\mathbb{P}\left(\sum_{j=1}^{n} \mathbf{1}(z_j \leq i/n - \epsilon) \geq i\right) &= \mathbb{P}\left(\sum_{j=1}^{n} (\mathbf{1}(z_j \leq i/n - \epsilon) - (i/n - \epsilon)) \geq n\epsilon\right) \\ &\leq \exp(-2\epsilon^2 n).\end{aligned}$$

Similarly,

$$\begin{aligned}\mathbb{P}(u_i \geq i/n + \epsilon) &\leq \mathbb{P}\left(\sum_{j=1}^{n} \mathbf{1}(z_j \geq i/n + \epsilon) \geq n - i\right) \\ &= \mathbb{P}\left(\sum_{j=1}^{n} (\mathbf{1}(z_j \leq i/n + \epsilon) - (1 - i/n - \epsilon)) \geq n\epsilon\right) \\ &\leq \exp(-2\epsilon^2 n).\end{aligned}$$

□

Lemma 4 connects the differences between consecutive U$[0,1]$ order statistics to exponential random variables. It is well-known, and the proof can be found in Pyke (1965).

Lemma 4. *Let $u_1 \leq \ldots \leq u_n$ be the order statistics of n i.i.d.* U$[0,1]$ *random variables. Then, $(1 - u_n, u_n - u_{n-1}, u_{n-1} - u_{n-2}, \ldots, u_2 - u_1, u_1)$ has the same distribution as $(x_1/x, x_2/x, \ldots, x_{n+1}/x)$, where $x_1, \ldots, x_{n+1}$ are i.i.d. exponential random variables and $x \equiv \sum_{i=1}^{n+1} x_i$.*

The following lemma is an easy exercise in integration:

Lemma 5. *Suppose that $x_1, \ldots x_l, y$ are i.i.d. unit exponential random variables. Then for any $c > 0$,*

$$\mathbb{P}\left(c \sum_{i=1}^{l} x_i \leq y\right) = (1 + c)^{-l}.$$

Proof of Lemma 2 (page 27). Fix a realization $(m_i, w_i)_{0 \leq i \leq r}$ such that

$$m_j - m_i \neq m_l - m_j, \quad w_j - w_i \neq w_l - w_j$$

for all $0 \leq i < j < l \leq r$. We will prove that $(m_i, w_i)_{0 \leq i \leq r}$ admits at most one assortative early matching that is sequentially stable in the sense of Definition 5.

Let $I = \{i_1, \ldots, i_L\}$ and $I' = \{i'_1, \ldots, i'_K\}$, where $i_0 \equiv 0 < i_1 < \cdots < i_L < r \equiv i_{L+1}$ and $i'_0 \equiv 0 < i'_1 < \cdots < i'_K < r \equiv i'_{K+1}$. Suppose that $I \neq I'$. For the sake of contradiction, suppose that both I and I' are sequentially stable for $(m_i, w_i)_{0 \leq i \leq r}$ in the sense of Definition 5.

First, find the smallest $\bar{l} \leq \min(L+1, K+1)$ in which $i_{\bar{l}} \neq i'_{\bar{l}}$. Without loss assume that $i_{\bar{l}} < i'_{\bar{l}}$. This means that $i_{\bar{l}} \in I$ but $i_{\bar{l}} \notin I'$. Since both I and I' are sequentially stable, for the couple of index $i_{\bar{l}}$, we have simultaneously:

$$m_{i_{\bar{l}}} - m_{i_{\bar{l}-1}} < m_{i_{\bar{l}+1}} - m_{i_{\bar{l}}} \text{ or } w_{i_{\bar{l}}} - w_{i_{\bar{l}-1}} < w_{i_{\bar{l}+1}} - w_{i_{\bar{l}}}$$

and (since $i_{\bar{l}-1} = i'_{\bar{l}-1} < i_{\bar{l}} < i'_{\bar{l}}$)

$$m_{i_{\bar{l}}} - m_{i'_{\bar{l}-1}} > m_{i'_{\bar{l}}} - m_{i_{\bar{l}}} \text{ and } w_{i_{\bar{l}}} - w_{i'_{\bar{l}-1}} > w_{i'_{\bar{l}}} - w_{i_{\bar{l}}}$$

which imply that $i_{\bar{l}+1} > i'_{\bar{l}}$.

Now consider index $i'_{\bar{l}}$: analogues of the above two inequalities imply that $i'_{\bar{l}+1} > i_{\bar{l}+1}$. That $i'_{\bar{l}+1} > i_{\bar{l}+1}$ in turn implies that $i_{\bar{l}+2} > i'_{\bar{l}+1}$ (considering index $i_{\bar{l}+1}$), and so on. Eventually we would conclude either $i'_{\bar{l}+j} > i_{L+1} = r$ or $i_{\bar{l}+j} > i'_{K+1} = r$, which are both impossible. □

A.2 Proof of Theorem 1

Let $q_{i,n}$ denote the probability that the first-period pair (m_i, w_i) has an incentive to match early, given that everyone is waiting for the assortative matching in the second period. Let $\mathbb{P}_n$ denote the probability measure over the types in the first period (and $\mathbb{E}_n$ for the corresponding expectation). Finally, let $\lfloor x \rfloor$ be the largest integer less than or equal to x.

Fix an arbitrary $\delta > 0$. We prove that there exists N such that for any $n \geq N$ we have:

$$\left| q_{i,n} - \frac{1}{4} \right| \leq \delta, \text{ for every } \delta n < i < (1-\delta)n. \tag{48}$$

As a consequence, for $n \geq N$ we have:

$$\mathbb{E}_n[\xi_n] = \sum_{i=1}^{n} q_{i,n} \geq \sum_{i=\lfloor n\delta \rfloor}^{\lfloor n(1-\delta) \rfloor} q_{i,n} \geq \left(\frac{1}{4} - \delta \right) (1-2\delta)n,$$

which implies a lower bound of 1/4 for pairs with an incentive to match early. The derivation of the upper bound is analogous and proves the theorem.

We now prove (48). By waiting to the second period, assuming everyone else does so as well, man $m_{\lfloor np \rfloor}$'s expected utility is:

$$\sum_{0 \le i < j \le k} \binom{k}{i} F(m_{\lfloor np \rfloor})^i (1 - F(m_{\lfloor np \rfloor}))^{k-i} \binom{k}{j} G(w_{\lfloor np \rfloor})^j (1 - G(w_{\lfloor np \rfloor}))^{k-j} \times \mathbb{E}_{(G_-(w_{\lfloor np \rfloor}))^{\otimes j}}[U(w^2_{\lfloor np \rfloor - (j-i)} \mid m_{\lfloor np \rfloor})]$$

$$+ \sum_{0 \le j < i \le k} \binom{k}{i} F(m_{\lfloor np \rfloor})^i (1 - F(m_{\lfloor np \rfloor}))^{k-i} \binom{k}{j} G(w_{\lfloor np \rfloor})^j (1 - G(w_{\lfloor np \rfloor}))^{k-j} \times \mathbb{E}_{(G_+(w_{\lfloor np \rfloor}))^{\otimes (k-j)}}[U(w^2_{\lfloor np \rfloor + (i-j)} \mid m_{\lfloor np \rfloor})]$$

$$+ \sum_{0 \le i \le k} \binom{k}{i} F(m_{\lfloor np \rfloor})^i (1 - F(m_{\lfloor np \rfloor}))^{k-i} \binom{k}{i} G(w_{\lfloor np \rfloor})^i (1 - G(w_{\lfloor np \rfloor}))^{k-i} \times U(w_{\lfloor np \rfloor} \mid m_{\lfloor np \rfloor}),$$

where $(G_-(w_{\lfloor np \rfloor}))^{\otimes j}$ is the probability measure of j i.i.d. random variables from the conditional distribution $G(x)/G(w_{\lfloor np \rfloor})$, $x \le w_{\lfloor np \rfloor}$; $w^2_{\lfloor np \rfloor - (j-i)}$ is the $(\lfloor np \rfloor - (j-i))$-th lowest woman among the j new women (from the distribution $G(x)/G(w_{\lfloor np \rfloor})$) and the first period women $w_1 \le \ldots \le w_n$; and likewise for other terms.

Comparing this to $U(w_{\lfloor np \rfloor} \mid m_{\lfloor np \rfloor})$, we see that man $m_{\lfloor np \rfloor}$ strictly prefers to match early with woman $w_{\lfloor np \rfloor}$ if and only if:

$$\sum_{0 \le i < j \le k} \binom{k}{i} F(m_{\lfloor np \rfloor})^i (1 - F(m_{\lfloor np \rfloor}))^{k-i} \binom{k}{j} G(w_{\lfloor np \rfloor})^j (1 - G(w_{\lfloor np \rfloor}))^{k-j} \times \Big(U(w_{\lfloor np \rfloor} \mid m_{\lfloor np \rfloor}) - \mathbb{E}_{(G_-(w_{\lfloor np \rfloor}))^{\otimes j}}[U(w^2_{\lfloor np \rfloor - (j-i)} \mid m_{\lfloor np \rfloor})]\Big)$$

$$> \sum_{0 \le j < i \le k} \binom{k}{i} F(m_{\lfloor np \rfloor})^i (1 - F(m_{\lfloor np \rfloor}))^{k-i} \binom{k}{j} G(w_{\lfloor np \rfloor})^j (1 - G(w_{\lfloor np \rfloor}))^{k-j} \times \Big(\mathbb{E}_{(G_+(w_{\lfloor np \rfloor}))^{\otimes (k-j)}}[U(w^2_{\lfloor np \rfloor + (i-j)} \mid m_{\lfloor np \rfloor})] - U(w_{\lfloor np \rfloor} \mid m_{\lfloor np \rfloor})\Big), \tag{49}$$

which has the interpretation of comparing man $m_{\lfloor np \rfloor}$'s downside in period 2 with his upside.

We have the following uniform-convergence result, and analogously for the terms on the RHS of (49) and in the inequality for woman $w_{\lfloor np \rfloor}$'s early-matching incentive.

Lemma 6. *Fix $j > i$. Without loss of generality, let $v_1 \leq \ldots \leq v_n$ be the order statistics of n i.i.d.* $\mathrm{U}[0,1]$ *random variables, and set $w_{i'} \equiv G^{-1}(v_{i'})$, $1 \leq i' \leq n$.*

For every $\epsilon > 0$, we have

$$\sup_{p\in[\delta,1-\delta]} \mathbb{P}_n \left(\left| \frac{\begin{array}{c} F(m_{\lfloor np \rfloor})^i(1-F(m_{\lfloor np \rfloor}))^{k-i} G(w_{\lfloor np \rfloor})^j(1-G(w_{\lfloor np \rfloor}))^{k-j} \\ \times \Big(U(w_{\lfloor np \rfloor} \mid m_{\lfloor np \rfloor}) - \mathbb{E}_{(G_-(w_{\lfloor np \rfloor}))^{\otimes j}}[U(w^2_{\lfloor np \rfloor-(j-i)} \mid m_{\lfloor np \rfloor})] \Big) \end{array}}{p^i(1-p)^{k-i}p^j(1-p)^{k-j}U'(G^{-1}(p) \mid F^{-1}(p))(v_{\lfloor np \rfloor} - v_{\lfloor np \rfloor-(j-i)})/g(G^{-1}(p))} - 1 \right| > \epsilon \right)$$

converges to zero with n.

Proof. Lemma 6 is a consequence of Lemma 3 and of the uniform continuities of $f(x)$, $g(y)$, $U(y \mid x)$ and $V(x \mid y)$ for $(x,y) \in [F^{-1}(\delta/2), F^{-1}(1-\delta/2)] \times [G^{-1}(\delta/2), G^{-1}(1-\delta/2)]$. □

Let

$$(a_{-k}, \ldots, a_{-1}, a_1, \ldots, a_k, \alpha_{-k}, \ldots, \alpha_{-1}, \alpha_1, \ldots, \alpha_k)$$

be $4k$ i.i.d. unit-exponential random variables. Define

$$\nu(p, C) \equiv \mathbb{P} \left(\begin{array}{l} C \sum_{0 \leq i < j \leq k} \binom{k}{i} p^i(1-p)^{k-i} \binom{k}{j} p^j (1-p)^{k-j} \cdot \sum_{l=i-j}^{-1} a_l \\ > \sum_{0 \leq j < i \leq k} \binom{k}{i} p^i(1-p)^{k-i} \binom{k}{j} p^j (1-p)^{k-j} \cdot \sum_{l=1}^{i-j} a_l, \\ C \sum_{0 \leq i < j \leq k} \binom{k}{i} p^i(1-p)^{k-i} \binom{k}{j} p^j (1-p)^{k-j} \cdot \sum_{l=i-j}^{-1} \alpha_l \\ > \sum_{0 \leq j < i \leq k} \binom{k}{i} p^i(1-p)^{k-i} \binom{k}{j} p^j (1-p)^{k-j} \cdot \sum_{l=1}^{i-j} \alpha_l \end{array} \right).$$

By Lemma 6 and Lemma 4, for any $\epsilon > 0$, there exists N such that

$$\nu\left(p, \frac{1-\epsilon}{1+\epsilon}\right) - \epsilon \leq q_{\lfloor np \rfloor, n} \leq \nu\left(p, \frac{1+\epsilon}{1-\epsilon}\right) + \epsilon$$

for all $p \in [\delta, 1-\delta]$ and $n \geq N$.

Furthermore, $\nu(p, C)$ converges to $\nu(p, 1) = 1/4$ as $C \to 1$, uniformly in $p \in [\delta, 1-\delta]$, which implies our conclusion.

A.3 Proof of Theorem 2

Throughout the proof we abuse our terminology and say that an early matching is sequentially stable if it, together with the assortative matching in the second period, forms a sequentially stable matching scheme in the sense of Definition 2; moreover, we focus on indices instead of type realizations and let $\mu \subseteq \{1, \ldots, n\}^2$. As in the proof of Theorem 1, we use $\mathbb{P}_n$ to denote the probability measure over the types in the first period (and $\mathbb{E}_n$ for the corresponding expectation), and use $\lfloor x \rfloor$ to denote the largest integer less than or equal to x. And let

$$\mathcal{M} \equiv \left\{ \mu \subset \{1, \ldots, n\}^2 : \begin{array}{l} (i,j), (i,j') \in \mu \Rightarrow j = j', \\ (i',j), (i,j) \in \mu \Rightarrow i = i' \end{array} \right\}$$

be the set of all possible early matchings.

For any $\mu \in \mathcal{M}$, let

$$C_\mu \subseteq \{(m_1, \ldots, m_n, w_1, \ldots, w_n) \in \mathbb{R}^{2n} \mid m_n \geq \ldots \geq m_1, w_n \geq \ldots \geq w_1\}$$

be the set of first-period types that admit μ as a sequentially stable early matching. Theorem 2 states that under the assumption of $k = 1$, we have

$$\lim_{n \to \infty} \mathbb{P}_n \left(\bigcup_{\mu \in \mathcal{M}} C_\mu \right) = 0.$$

We first note that under the assumption $k = 1$, any sequentially stable early matching μ satisfies the following "intermediate" property:

$$\begin{aligned} &(i,j) \in \mu, \ i \leq l \leq j \text{ or } j \leq l \leq i \\ \Longrightarrow &\text{ there exists } i' \text{ and } j' \text{ such that } (i',l) \in \mu \text{ and } (l,j') \in \mu \end{aligned} \tag{Int}$$

This property follows from the fact that if μ is sequentially stable and $(i,j) \in \mu$, then if one of them breaks the early matching, man i and woman j must be of the same rank among those who wait (according to μ) plus themselves. Suppose otherwise, say man i is of a lower rank than woman j, then woman j cannot have incentive to match early with man i because the new arrivals can change woman j's ranking by at most one place in the second period (recall that $k = 1$), so in any case she gets a better (or equal) match than m_i.

Define

$$\mathcal{M}_{\text{int}} \equiv \{\mu \in \mathcal{M} : \mu \text{ satisfies property (Int)}\}.$$

We have that $C_\mu = \emptyset$ for any $\mu \notin \mathcal{M}_{\text{int}}$.

Notice that if $\mu \in \mathcal{M}_{\text{int}}$, then the set of man-ranks who do not match early under μ

equals the set of woman-ranks who do not match early under μ:

$$\mathbf{I}(\mu) \equiv \{i \in \{1, \ldots, n\} : \forall j, (i,j) \notin \mu\} = \{j \in \{1, \ldots, n\} : \forall i, (i,j) \notin \mu\}. \tag{50}$$

We fix an arbitrary $p \in (0,1)$ throughout the proof of Theorem 2.

For fixed integers $s > 0$ and $t > 0$, define:

$$\mathcal{I} \equiv \{I : I \subseteq \{1, \ldots, n\}\}.$$

$$\mathcal{I}_{j,l} \equiv \{I \in \mathcal{I} : \lfloor np \rfloor + t + j \in I \text{ and } \lfloor np \rfloor - t - l \in I\}. \tag{51}$$

$$\mathcal{I}' \equiv \left\{ I \in \mathcal{I} : \begin{array}{l} \min(\{i \in I : i \geq \lfloor np \rfloor + t\}) > \lfloor np \rfloor + t + s, \text{ or} \\ \max(\{i \in I : i \leq \lfloor np \rfloor - t\}) < \lfloor np \rfloor - t - s \end{array} \right\}.$$

In words, $\mathcal{I}'$ is the set of agents' ranks in which there is a "gap" of size at least s, either starting at $\lfloor np \rfloor + t$ or ending at $\lfloor np \rfloor - t$.

Clearly,

$$\mathcal{I} = \mathcal{I}' \cup \bigcup_{\substack{0 \leq j \leq s \\ 0 \leq l \leq s}} \mathcal{I}_{j,l}.$$

We can further divide $\mathcal{I}'$ into $\mathcal{I}'_1$ and $\mathcal{I}'_2$: $\mathcal{I} = \mathcal{I}'_1 \cup \mathcal{I}'_2$, where

$$\mathcal{I}'_1 \equiv \{I \in \mathcal{I} : \min(\{i \in I : i \geq \lfloor np \rfloor + t\}) > \lfloor np \rfloor + t + s\}, \tag{52}$$

$$\mathcal{I}'_2 \equiv \{I \in \mathcal{I} : \max(\{i \in I : i \leq \lfloor np \rfloor - t\}) < \lfloor np \rfloor - t - s\}.$$

Let

$$C_I \equiv \bigcup_{\substack{\mu \in \mathcal{M}_{\text{int}}, \\ \mathbf{I}(\mu) = I}} C_\mu$$

for $I \in \mathcal{I}$, and

$$C(\mathcal{I}'') \equiv \bigcup_{I \in \mathcal{I}''} C_I$$

for $\mathcal{I}'' \subseteq \mathcal{I}$.

Clearly, we have

$$C(\mathcal{I}) = \bigcup_{\mu \in \mathcal{M}} C_\mu.$$

On the other hand,

$$\mathbb{P}_n(C(\mathcal{I})) \le \mathbb{P}_n(C(\mathcal{I}'_1)) + \mathbb{P}_n(C(\mathcal{I}'_2)) + \sum_{\substack{0\le j\le s\\ 0\le l\le s}} C(\mathcal{I}_{j,l}).$$

Therefore,

$$\begin{aligned}\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I})) \le& \limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_1)) + \limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_2))\\ &+ \sum_{\substack{0\le j\le s\\ 0\le l\le s}} \limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_{j,l})).\end{aligned} \tag{53}$$

In the next two subsections we show that for a fixed s, $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_{j,l}))$ goes to 0 as t goes to infinity for any j and l, at a rate independent of s (Section A.3.1, Equation (58) and Proposition 1); and that for a fixed t, $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_1))$ and $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_2))$ go to 0 as s goes to infinity, at a rate independent of t (Section A.3.2, Equations (65) and (66)). This implies that by choosing s and t sufficiently large, we can make the left hand side of (53), which is independent of s and t, as close to zero as we want. Thus, the left hand side of (53) must be exactly zero, which proves Theorem 2.

A.3.1 Reduction to the Exponential Model of Section 4.2.2

We first bound the term $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_{j,l}))$ in (53). An early matching μ with $\mathbf{I}(\mu) \in \mathcal{I}_{j,l}$ has the property that men and women of ranks $\lfloor np\rfloor + t + j$ and $\lfloor np\rfloor - t - l$ do not match early under μ. Thus, types in $C(\mathcal{I}_{j,l})$ satisfy the property that assuming that men and women of ranks $\lfloor np\rfloor + t + j$ and $\lfloor np\rfloor - t - l$ wait to the second period, men and women between ranks $\lfloor np\rfloor + t + j$ and $\lfloor np\rfloor - t - l$ admit a sequentially stable early matching among themselves.

Therefore, in this subsection we solve the following local problem. Fix an integer $r > 1$. Assuming that *exogenously* men and women of ranks $\lfloor np\rfloor$ and $\lfloor np\rfloor + r$ wait to the second period, what is the probability that agents between ranks $\lfloor np\rfloor$ and $\lfloor np\rfloor + r$ admit a sequentially stable early matching *among themselves*? [18]

Denote

$$U'(w \mid m) \equiv \frac{\partial U(w \mid m)}{\partial w}, V'(m \mid w) \equiv \frac{\partial V(m \mid w)}{\partial m}. \tag{54}$$

We first note that given $k = 1$, the expected utility of m_i in the second period is (where

[18] It will be obvious that we get the same result by assuming that men and women of ranks $\lfloor np\rfloor + i$ and $\lfloor np\rfloor + i + r$ wait to the second period, for any fixed i.

$w_+ > w_i > w_-$ are the types of women who are going to the second period (women in the list L) and who are just around the rank of m_i):

$$\begin{aligned}\mathbf{U}(L, m_i) =& \Big(F(m_i)G(w_i) + (1 - F(m_i))(1 - G(w_i)) \Big) U(w_i \mid m_i) \\ &+ \ (1 - F(m_i)) \left(U(w_- \mid m_i)G(w_-) + \int_{w_-}^{w_i} U(x \mid m_i)g(x)\, dx \right) \\ &+ \ F(m_i) \left(U(w_+ \mid m_i)(1 - G(w_+)) + \int_{w_i}^{w_+} U(x \mid m_i)g(x)\, dx \right),\end{aligned} \tag{55}$$

where the first term represents the events in which m_i is matched to w_i in the second period, the second term represents events in which m_i is matched to a worse type than w_i, and the last term represents the events in which m_i is matched to a better type than w_i. Clearly, a symmetric formula holds for woman w_i.

It is easy to use integration by parts to verify that m_i has strict incentive to match early with w_i (i.e., $U(w_i \mid m_i)$ dominates (55)) if and only if:

$$(1 - F(m_i)) \int_{w_-}^{w_i} U'(x \mid m_i)G(x)\, dx > F(m_i) \int_{w_i}^{w_+} U'(x \mid m_i)(1 - G(x))\, dx, \tag{56}$$

which has the interpretation of comparing m_i's downside in period 2 with his upside.

Before putting (56) to use, let us adapt some of the previous notations to the present, local setting. Let

$$\mathcal{M}(r) \equiv \left\{ \mu \subset \{1, \dots, r-1\}^2 : \begin{array}{l} (i,j), (i,j') \in \mu \Rightarrow j = j', \\ (i',j), (i,j) \in \mu \Rightarrow i = i' \end{array} \right\};$$

an early matching among agents between ranks $\lfloor np \rfloor$ and $\lfloor np \rfloor + r$ can be represented a $\mu \in \mathcal{M}(r)$: man of rank $\lfloor np \rfloor + i$ matches early with woman of rank $\lfloor np \rfloor + j$ according to μ if and only if $(i,j) \in \mu$.

And as before, define

$$\mathcal{M}(r)_{\text{int}} \equiv \{\mu \in \mathcal{M}(r) : \mu \text{ satisfies property (Int)}\},$$

where the property (Int) is defined in page 44.

Finally, for each $\mu \in \mathcal{M}(r)_{\text{int}}$, define

$$\mathbf{I}_r(\mu) \equiv \{i \in \{1, \dots, r-1\} : \forall j, (i,j) \notin \mu\} = \{j \in \{1, \dots, r-1\} : \forall i, (i,j) \notin \mu\}.$$

to be the set of ranks of men/women who wait to second period according to μ.

Given (56), we can simplify Definition 2:

Definition 6. Fix an ordered list of types $P = \big((m_i)_{\lfloor np\rfloor \le i \le \lfloor np\rfloor + r}, (w_i)_{\lfloor np\rfloor \le i \le \lfloor np\rfloor + r}\big)$ and an early matching $\mu \in \mathcal{M}(r)_{\text{int}}$. Suppose that $\mathbf{I}_r(\mu) = \{i_l : 1 \le l \le L\}$, where $0 \equiv i_0 < i_1 < \ldots < i_L < i_{L+1} \equiv r$. Then, early matching μ is *sequentially stable* for P if:

1. for every $1 \le l \le L$, we have either

$$\begin{aligned}&(1 - F(m_{\lfloor np\rfloor+i_l})) \int_{w_{\lfloor np\rfloor+i_{l-1}}}^{w_{\lfloor np\rfloor+i_l}} U'(x \mid m_{\lfloor np\rfloor+i_l}) G(x)\, dx \\ \le & F(m_{\lfloor np\rfloor+i_l}) \int_{w_{\lfloor np\rfloor+i_l}}^{w_{\lfloor np\rfloor+i_{l+1}}} U'(x \mid m_{\lfloor np\rfloor+i_l})(1 - G(x))\, dx,\end{aligned}$$

   or

$$\begin{aligned}&(1 - G(w_{\lfloor np\rfloor+i_l})) \int_{m_{\lfloor np\rfloor+i_{l-1}}}^{m_{\lfloor np\rfloor+i_l}} V'(x \mid w_{\lfloor np\rfloor+i_l}) F(x)\, dx \\ \le & G(w_{\lfloor np\rfloor+i_l}) \int_{m_{\lfloor np\rfloor+i_l}}^{m_{\lfloor np\rfloor+i_{l+1}}} V'(x \mid w_{\lfloor np\rfloor+i_l})(1 - F(x))\, dx,\end{aligned}$$

2. for the couple $(i, j) \in \mu$ who matches early (there exists a unique l such that $i_l < i < i_{l+1}$ and $i_l < j < i_{l+1}$), we have

$$\begin{aligned}&(1 - F(m_{\lfloor np\rfloor+i})) \int_{w_{\lfloor np\rfloor+i_l}}^{w_{\lfloor np\rfloor+j}} U'(x \mid m_{\lfloor np\rfloor+i}) G(x)\, dx \\ \ge & F(m_{\lfloor np\rfloor+i}) \int_{w_{\lfloor np\rfloor+j}}^{w_{\lfloor np\rfloor+i_{l+1}}} U'(x \mid m_{\lfloor np\rfloor+i})(1 - G(x))\, dx,\end{aligned}$$

   and

$$\begin{aligned}&(1 - G(w_{\lfloor np\rfloor+j})) \int_{m_{\lfloor np\rfloor+i_l}}^{m_{\lfloor np\rfloor+i}} V'(x \mid w_{\lfloor np\rfloor+j}) F(x)\, dx \\ \ge & G(w_{\lfloor np\rfloor+j}) \int_{m_{\lfloor np\rfloor+i}}^{m_{\lfloor np\rfloor+i_{l+1}}} V'(x \mid w_{\lfloor np\rfloor+j})(1 - F(x))\, dx.\end{aligned}$$

For $\mu \in \mathcal{M}(r)_{\text{int}}$, let

$$D_\mu \subseteq \left\{ \big((m_i)_{\lfloor np\rfloor \le i \le \lfloor np\rfloor + r}, (w_i)_{\lfloor np\rfloor \le i \le \lfloor np\rfloor + r}\big) \in \mathbb{R}^{2(r+1)} : \begin{array}{l} m_{\lfloor np\rfloor+r} \ge \ldots \ge m_{\lfloor np\rfloor}, \\ w_{\lfloor np\rfloor+r} \ge \ldots \ge w_{\lfloor np\rfloor} \end{array} \right\}$$

be the set of types in between ranks $\lfloor np \rfloor$ and $\lfloor np \rfloor + r$ that admit μ as a sequentially stable early matching according to Definition 6.

For each $I \subseteq \{1, \ldots, r-1\}$, let $G_I \subseteq \mathbb{R}^{2(r+1)}$ be the types that make I sequentially stable according to Definition 5 in Section 4.2.2. Let $\hat{\mathbb{P}}_r$ be the probability measure of the exponential model described in Section 4.2.2 and recall the definition of π_r in Equation (36).

Proposition 2. *Fix $p \in (0,1)$ and integer $r \geq 2$. We have*

$$\lim_{n\to\infty} \mathbb{P}_n \left(\bigcup_{\mu \in \mathcal{M}(r)_{int}} D_\mu \right) = \hat{\mathbb{P}}_r \left(\bigcup_{I \subseteq \{1,\ldots,r-1\}} G_I \right) = \pi_r. \tag{57}$$

Proof. The proof is omitted because it is identical to that of Theorem 1. Essentially, because $m_{\lfloor np \rfloor + r} - m_{\lfloor np \rfloor}$ and $w_{\lfloor np \rfloor + r} - w_{\lfloor np \rfloor}$ converge to zero in probability with n, the condition in Point 1 of Definition 6 becomes approximately

$$w_{\lfloor np \rfloor + i_l} - w_{\lfloor np \rfloor + i_{l-1}} \leq w_{\lfloor np \rfloor + i_{l+1}} - w_{\lfloor np \rfloor + i_l} \text{ or } m_{\lfloor np \rfloor + i_l} - m_{\lfloor np \rfloor + i_{l-1}} \leq m_{\lfloor np \rfloor + i_{l+1}} - m_{\lfloor np \rfloor + i_l},$$

and the condition in point 2 of Definition 6 becomes approximately

$$w_{\lfloor np \rfloor + j} - w_{\lfloor np \rfloor + i_l} \geq w_{\lfloor np \rfloor + i_{l+1}} - w_{\lfloor np \rfloor + j} \text{ and } m_{\lfloor np \rfloor + i} - m_{\lfloor np \rfloor + i_l} \geq m_{\lfloor np \rfloor + i_{l+1}} - m_{\lfloor np \rfloor + i},$$

when n is large. (Compare with Definition 5 for the definition of G_I in the exponential model.) $\square$

Now going back to bounding (53): by Proposition 2 we have

$$\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_{j,l})) = \pi_{2t+j+l} \tag{58}$$

because positions $\lfloor np \rfloor + t + j$ and $\lfloor np \rfloor - t - l$ are of distance $2t + j + l$ apart (cf. Equation (51)).

Therefore, Proposition 1 proves that $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_{j,l}))$ goes to 0 as t goes to infinity, for any fixed $s > 0$, $0 \leq j \leq s$ and $0 \leq l \leq s$.

A.3.2 Consecutive pairs matching early

In this subsection we give the required bounds for $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_1))$ and $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_2))$ in (53). Recall that p is an arbitrary percentile in $(0,1)$. Intuitively, we are removing the "local" assumption of the previous subsection that men and women of ranks $\lfloor np \rfloor$ and $\lfloor np \rfloor + r$

exogenously wait to the second period.

Let us focus $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_1))$; the bound for $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}'_2))$ is analogous.

Let $\bar{t} \equiv t + s + 1$, $m_0 \equiv \underline{m} > -\infty$ and $w_0 \equiv \underline{w} > -\infty$.

For every $1 \leq i \leq \lfloor np \rfloor + t - 1$, define

$$\mathcal{I}'_1(i) \equiv \left\{ I \in \mathcal{I} : \begin{array}{l} \max(\{i' \in I : i' < \lfloor np \rfloor + t\}) = i, \text{ and} \\ \min(\{i' \in I : i' \geq \lfloor np \rfloor + t\}) \geq \lfloor np \rfloor + \bar{t} \end{array} \right\},$$

and define

$$\mathcal{I}'_1(0) \equiv \left\{ I \in \mathcal{I} : \begin{array}{l} \{i' \in I : i' < \lfloor np \rfloor + t\}) = \emptyset, \text{ and} \\ \min(\{i' \in I : i' \geq \lfloor np \rfloor + t\}) \geq \lfloor np \rfloor + \bar{t} \end{array} \right\}.$$

By construction, we have (cf. (52) for the definition of $\mathcal{I}'_1$):

$$C(\mathcal{I}'_1) = \bigcup_{0 \leq i \leq \lfloor np \rfloor + t - 1} C(\mathcal{I}'_1(i)) \tag{59}$$

For any $0 \leq i \leq \lfloor np \rfloor + t - 1$ we have

$$\mathbb{P}_n(C(\mathcal{I}'_1(i))) \leq \mathbb{P}_n \left(\begin{array}{l} \bar{b} \int_{w_i}^{w_{i+1}} G(x)\, dx \geq F(m_{i+1}) \int_{w_{i+1}}^{w_{\lfloor np \rfloor + \bar{t}}} (1 - G(x))\, dx, \\ \bar{b} \int_{m_i}^{m_{i+1}} F(x)\, dx \geq G(w_{i+1}) \int_{m_{i+1}}^{m_{\lfloor np \rfloor + \bar{t}}} (1 - F(x))\, dx \end{array} \right) \tag{60}$$

because types in $C(\mathcal{I}'_1(i))$ are such that given agents of rank i wait to the second period while the next rank above i who waits is at least $\lfloor np \rfloor + \bar{t}$, some woman of rank $j \geq i + 1$ must have incentive to match early with the man of rank $i + 1$, and some man of rank $j' \geq i+1$ must have incentive to match early with the woman of rank $i+1$. Let's look at the incentive of the woman of rank j: her "downside" from waiting must dominate her "upside" (see equation (56) on page 47); she has an "downside" of at most

$$(1 - G(w_j)) \int_{m_i}^{m_{i+1}} V'(x \mid w_j) F(x)\, dx \leq \int_{m_i}^{m_{i+1}} V'(x \mid w_j) F(x)\, dx\,,$$

and an "upside" of at least

$$G(w_j) \int_{m_{i+1}}^{m_{\lfloor np \rfloor + \bar{t}}} V'(x \mid w_j)(1 - F(x))\, dx \geq G(w_{i+1}) \int_{m_{i+1}}^{m_{\lfloor np \rfloor + \bar{t}}} V'(x \mid w_j)(1 - F(x))\, dx\,;$$

and likewise for man of rank j'. Finally, for any x and y we have $V'(x \mid w_j)/V'(y \mid w_j) \leq \bar{b}$ by assumption (4). This explains the inequalities inside of the probability in (60).

Fix a $\epsilon > 0$ such that $p + \epsilon < 1$, $p - 3\epsilon > 0$ and $\bar{t}/n < \epsilon/2$.

Equations (59) and (60) imply that:

$$
\begin{aligned}
&\mathbb{P}_n(C(\mathcal{I}_1')) \\
&\leq \sum_{i=1}^{\lfloor np \rfloor + t} \mathbb{P}_n \begin{pmatrix} \bar{b} \int_{w_{i-1}}^{w_i} G(x)\,dx \geq F(m_i) \int_{w_i}^{w_{\lfloor np \rfloor + \bar{t}}} (1 - G(x))\,dx, \\ \bar{b} \int_{m_{i-1}}^{m_i} F(x)\,dx \geq G(w_i) \int_{m_i}^{m_{\lfloor np \rfloor + \bar{t}}} (1 - F(x))\,dx \end{pmatrix} \\
&\leq \sum_{i=\lfloor np_0 \rfloor + 1}^{\lfloor np \rfloor + t} \mathbb{P}_n \begin{pmatrix} \bar{b} \int_{w_{i-1}}^{w_i} G(x)\,dx \geq F(m_i) \int_{w_i}^{w_{\lfloor np \rfloor + \bar{t}}} (1 - G(x))\,dx, \\ \bar{b} \int_{m_{i-1}}^{m_i} F(x)\,dx \geq G(w_i) \int_{m_i}^{m_{\lfloor np \rfloor + \bar{t}}} (1 - F(x))\,dx \end{pmatrix} \\
&\quad + \sum_{i=1}^{\lfloor np_0 \rfloor} \mathbb{P}_n \begin{pmatrix} \bar{b} \int_{w_{i-1}}^{w_i} G(x)\,dx \geq F(m_i) \int_{w_i}^{w_{\lfloor np \rfloor + \bar{t}}} (1 - G(x))\,dx, \\ \bar{b} \int_{m_{i-1}}^{m_i} F(x)\,dx \geq G(w_i) \int_{m_i}^{m_{\lfloor np \rfloor + \bar{t}}} (1 - F(x))\,dx \end{pmatrix} \\
&\leq \sum_{i=\lfloor np_0 \rfloor + 1}^{\lfloor np \rfloor + t} \mathbb{P}_n \begin{pmatrix} \bar{b}(w_i - w_{i-1}) \geq F(m_{\lfloor np_0 \rfloor})(1 - G(w_{\lfloor np \rfloor + \bar{t}}))(w_{\lfloor np \rfloor + \bar{t}} - w_i), \\ \bar{b}(m_i - m_{i-1}) \geq G(w_{\lfloor np_0 \rfloor})(1 - F(m_{\lfloor np \rfloor + \bar{t}}))(m_{\lfloor np \rfloor + \bar{t}} - m_i) \end{pmatrix} \qquad (61) \\
&\quad + \sum_{i=1}^{\lfloor np_0 \rfloor} \mathbb{P}_n \begin{pmatrix} \bar{b} G(w_i)(w_i - w_{i-1}) \geq F(m_i)(1 - G(w_{\lfloor np \rfloor + \bar{t}}))(w_{\lfloor np \rfloor + \bar{t}} - w_{\lfloor np_0 \rfloor}), \\ \bar{b} F(m_i)(m_i - m_{i-1}) \geq G(w_i)(1 - F(m_{\lfloor np \rfloor + \bar{t}}))(m_{\lfloor np \rfloor + \bar{t}} - m_{\lfloor np_0 \rfloor}) \end{pmatrix}, \qquad (62)
\end{aligned}
$$

where $p_0 \in (\epsilon, p - 2\epsilon)$ is arbitrary.

Our goal is then to bound (61) and (62).

Choose $\overline{a} > \underline{a} > 0$ so that

$$\underline{a} \leq f(x), g(y) \leq \overline{a},$$

for all $(x, y) \in [F^{-1}(p_0 - \epsilon), F^{-1}(p + \epsilon)] \times [G^{-1}(p_0 - \epsilon), G^{-1}(p + \epsilon)]$. For (61) we have:

$$
\begin{aligned}
&\mathbb{P}_n\begin{pmatrix} \bar{b}(w_i - w_{i-1}) \geq F(m_{\lfloor np_0 \rfloor})(1 - G(w_{\lfloor np \rfloor + \bar{t}}))(w_{\lfloor np \rfloor + \bar{t}} - w_i), \\ \bar{b}(m_i - m_{i-1}) \geq G(w_{\lfloor np_0 \rfloor})(1 - F(m_{\lfloor np \rfloor + \bar{t}}))(m_{\lfloor np \rfloor + \bar{t}} - m_i) \end{pmatrix} \\
\leq &\mathbb{P}_n\begin{pmatrix} \bar{b}(w_i - w_{i-1}) \geq F(m_{\lfloor np_0 \rfloor})(1 - G(w_{\lfloor np \rfloor + \bar{t}}))(w_{\lfloor np \rfloor + \bar{t}} - w_i), \\ G(w_{\lfloor np \rfloor + \bar{t}}) \leq p + \epsilon, F(m_{\lfloor np_0 \rfloor}) \geq p_0 - \epsilon, \\ \bar{b}(m_i - m_{i-1}) \geq G(w_{\lfloor np_0 \rfloor})(1 - F(m_{\lfloor np_0 \rfloor + \bar{t}}))(m_{\lfloor np \rfloor + \bar{t}} - m_i), \\ F(m_{\lfloor np \rfloor + \bar{t}}) \leq p + \epsilon, G(w_{\lfloor np_0 \rfloor}) \geq p_0 - \epsilon \end{pmatrix} \\
&+ \mathbb{P}_n(G(w_{\lfloor np \rfloor + \bar{t}}) > p + \epsilon) + \mathbb{P}_n(F(m_{\lfloor np \rfloor + \bar{t}}) > p + \epsilon) \\
&+ \mathbb{P}_n(F(m_{\lfloor np_0 \rfloor}) < p_0 - \epsilon) + \mathbb{P}_n(G(w_{\lfloor np_0 \rfloor}) < p_0 - \epsilon) \\
\leq &\mathbb{P}_n\begin{pmatrix} (\overline{a}/\underline{a})\bar{b}(v_i - v_{i-1}) \geq (p_0 - \epsilon)(1 - p - \epsilon)(v_{\lfloor np \rfloor + \bar{t}} - v_i), \\ (\overline{a}/\underline{a})\bar{b}(u_i - u_{i-1}) \geq (p_0 - \epsilon)(1 - p - \epsilon)(u_{\lfloor np \rfloor + \bar{t}} - u_i) \end{pmatrix} \\
&+ \mathbb{P}_n(G(w_{\lfloor np \rfloor + \bar{t}}) > p + \epsilon) + \mathbb{P}_n(F(m_{\lfloor np \rfloor + \bar{t}}) > p + \epsilon) \\
&+ \mathbb{P}_n(F(m_{\lfloor np_0 \rfloor}) < p_0 - \epsilon) + \mathbb{P}_n(G(w_{\lfloor np_0 \rfloor}) < p_0 - \epsilon),
\end{aligned}
$$

where it is w.l.o.g. to set $m_i \equiv F^{-1}(u_i)$ and $w_i \equiv G^{-1}(v_i)$: $u_1 \leq \ldots \leq u_n$ and $v_1 \leq \ldots \leq v_n$ are (independent copies of) order statistics of n i.i.d. U$[0,1]$ random variables (let $u_0 \equiv 0$ and $v_0 \equiv 0$).

By construction, we have $0 < \underline{a} \leq f(m), g(w) \leq \overline{a} < \infty$ for all $m \in [F^{-1}(p_0 - \epsilon), F^{-1}(p + \epsilon)]$ and $w \in [G^{-1}(p_0 - \epsilon), G^{-1}(p + \epsilon)]$. Thus, by the mean value theorem, we have for $i > j$:

$$(u_i - u_j)/\overline{a} \leq m_i - m_j \leq (u_i - u_j)/\underline{a}$$

and

$$(v_i - v_j)/\overline{a} \leq w_i - w_j \leq (v_i - v_j)/\underline{a},$$

which explains the presence of $\overline{a}/\underline{a}$ in the above probabilities.

By Lemma 3, we have

$$
\begin{aligned}
\lim_{n\to\infty} n\Big(&\mathbb{P}_n(G(w_{\lfloor np \rfloor + \bar{t}}) > p + \epsilon) + \mathbb{P}_n(F(m_{\lfloor np \rfloor + \bar{t}}) > p + \epsilon) \\
&+ \mathbb{P}_n(F(m_{\lfloor np_0 \rfloor}) < p_0 - \epsilon) + \mathbb{P}_n(G(w_{\lfloor np_0 \rfloor}) < p_0 - \epsilon)\Big) = 0.
\end{aligned}
$$

Therefore, to bound $\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_1'))$ we can replace (61) by

$$\sum_{i=\lfloor np_0\rfloor+1}^{\lfloor np\rfloor+t} \mathbb{P}_n \left(\begin{array}{l} (\overline{a}/\underline{a})\overline{b}(v_i - v_{i-1}) \geq (p_0 - \epsilon)(1 - p - \epsilon)(v_{\lfloor np\rfloor+\bar{t}} - v_i), \\ (\overline{a}/\underline{a})\overline{b}(u_i - u_{i-1}) \geq (p_0 - \epsilon)(1 - p - \epsilon)(u_{\lfloor np\rfloor+\bar{t}} - u_i) \end{array} \right). \tag{63}$$

Similarly, we can replace (62) by

$$\sum_{i=1}^{\lfloor np_0\rfloor} \mathbb{P}_n \left(\begin{array}{l} \overline{a}\overline{b}G(w_i)(w_i - w_{i-1}) \geq F(m_i)(1 - p - \epsilon)(p - p_0 - 2\epsilon), \\ \overline{a}\overline{b}F(m_i)(m_i - m_{i-1}) \geq G(w_i)(1 - p - \epsilon)(p - p_0 - 2\epsilon) \end{array} \right). \tag{64}$$

The following lemma takes care of (64):

Lemma 7.

$$\lim_{n\to\infty} \sum_{i=1}^{\lfloor np_0\rfloor} \mathbb{P}_n \left(\begin{array}{l} \overline{a}\overline{b}G(w_i)(w_i - w_{i-1}) \geq F(m_i)(1 - p - \epsilon)(p - p_0 - 2\epsilon), \\ \overline{a}\overline{b}F(m_i)(m_i - m_{i-1}) \geq G(w_i)(1 - p - \epsilon)(p - p_0 - 2\epsilon) \end{array} \right) = 0.$$

Proof. For any $1 \leq i \leq \lfloor np_0\rfloor$, we have:

$$\begin{aligned}
&\mathbb{P}_n \left(\begin{array}{l} \overline{a}\overline{b}G(w_i)(w_i - w_{i-1}) \geq F(m_i)(1 - p - \epsilon)(p - p_0 - 2\epsilon), \\ \overline{a}\overline{b}F(m_i)(m_i - m_{i-1}) \geq G(w_i)(1 - p - \epsilon)(p - p_0 - 2\epsilon) \end{array} \right) \\
\leq &\mathbb{P}_n \left((w_i - w_{i-1})(m_i - m_{i-1}) \geq \left(\frac{(1 - p - \epsilon)(p - p_0 - 2\epsilon)}{\overline{a}\overline{b}} \right)^2 \right) \\
\leq &\mathbb{P}_n \left((m_i - m_{i-1}) \geq \frac{(1 - p - \epsilon)(p - p_0 - 2\epsilon)}{\overline{a}\overline{b}} \right) \\
&+ \mathbb{P}_n \left((w_i - w_{i-1}) \geq \frac{(1 - p - \epsilon)(p - p_0 - 2\epsilon)}{\overline{a}\overline{b}} \right).
\end{aligned}$$

Let

$$C \equiv \frac{(1 - p - \epsilon)(p - p_0 - 2\epsilon)}{\overline{a}\overline{b}},$$

and

$$\delta \equiv \sup_{x\in[\underline{m}, F^{-1}(p_0+\epsilon)]} \frac{1 - F(x + C)}{1 - F(x)} < 1.$$

We have

$$\begin{aligned}
\mathbb{P}_n((m_i - m_{i-1}) \geq C) &\leq \mathbb{P}_n((m_i - m_{i-1}) \geq C, F(m_{i-1}) \leq p_0 + \epsilon) + 2\exp(-2\epsilon^2 n) \\
&\leq \delta^{n-\lfloor np_0\rfloor} + 2\exp(-2\epsilon^2 n),
\end{aligned}$$

for every $1 \leq i \leq \lfloor np_0 \rfloor$.

And likewise for $\mathbb{P}_n((w_i - w_{i-1}) \geq C)$. $\square$

By Lemma 4 and 5, we have:

$$\begin{aligned}
&\sum_{i=\lfloor np_0 \rfloor+1}^{\lfloor np \rfloor + t} \mathbb{P}_n \begin{pmatrix} (\overline{a}/\underline{a})\overline{b}(v_i - v_{i-1}) \geq (p_0 - \epsilon)(1-p-\epsilon)(v_{\lfloor np \rfloor + \bar{t}} - v_i), \\ (\overline{a}/\underline{a})\overline{b}(u_i - u_{i-1}) \geq (p_0 - \epsilon)(1-p-\epsilon)(u_{\lfloor np \rfloor + \bar{t}} - u_i) \end{pmatrix} \\
\leq &\sum_{i=\lfloor np_0 \rfloor+1}^{\lfloor np \rfloor + t} \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-(\lfloor np \rfloor + \bar{t} - i)} \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-(\lfloor np \rfloor + \bar{t} - i)} \\
\leq &\left(1 - \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-2}\right) \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-2s},
\end{aligned}$$

since $\bar{t} - t = s + 1$.

Therefore, we have

$$\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_1')) \leq \left(1 - \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-2}\right) \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-2s}. \tag{65}$$

By exact same argument, we have

$$\limsup_{n\to\infty} \mathbb{P}_n(C(\mathcal{I}_2')) \leq \left(1 - \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-2}\right) \left(1 + \frac{\underline{a}(p_0-\epsilon)(1-p-\epsilon)}{\overline{a}\overline{b}}\right)^{-2s}. \tag{66}$$

These are the required bounds for Equation (53) and complete the proof of Theorem 2.

B Detail for Example 5

Consider the setting of Example 5, and let $L = \{m_1, m_2\} \cup \{w_1, w_2\}$, i.e., all first-period agents go to period 2.

As in Example 2 (Inequality (23)), one can show that with fixed transfers, man m_1 has an incentive to match early with woman w_2, i.e., $m_1 w_2/2 > \mathbf{U}(L, m_1)$ if and only if

$$\frac{m_1}{2} \cdot (w_2 - w_1) + \frac{m_1}{2} \cdot (1 - m_1)(w_1)^2 > \frac{m_1}{2} \cdot m_1((1-w_1)^2 - (1-w_2)^2), \tag{67}$$

where the differences with Inequality (23) are that we do not divide out $\frac{m_1}{2}$, and that we have the correction term $\frac{m_1}{2}(w_2 - w_1)$ on the left-hand side because m_1 is thinking about matching early with w_2 instead of w_1. Inequality (67) always holds (because $w_2 - w_1 > (1-w_1)^2 - (1-w_2)^2$), which reflects the fact that m_1 is always matched to a type lower than w_2 in period 2 (since $k = 1$), so he has an incentive to match early with w_2.

Likewise, with fixed transfers woman w_2 has an incentive to match early with man m_1, i.e., $m_1 w_2/2 > \mathbf{V}(L, w_2)$ if and only if

$$\frac{w_2}{2} \cdot (m_1 - m_2) + \frac{w_2}{2} \cdot (1 - w_2)((m_2)^2 - (m_1)^2) > \frac{w_2}{2} \cdot w_2 (1 - m_2)^2. \tag{68}$$

Inequality (68) can never hold (because $m_2 - m_1 > (m_2)^2 - (m_1)^2$), which reflects the fact that w_2 is always matched to a type higher than m_1 in period 2, so without transfers she does not have an incentive to match early with m_1.

Adding (67) and (68) together, we see that m_1 and w_2 have an incentive to match early given some transfers, i.e., $m_1 w_2 > \mathbf{U}(L, m_1) + \mathbf{V}(L, w_2)$, if and only if

$$\begin{aligned} &\frac{m_1}{2} \cdot (w_2 - w_1) + \frac{m_1}{2} \cdot (1 - m_1)(w_1)^2 + \frac{w_2}{2} \cdot (m_1 - m_2) + \frac{w_2}{2} \cdot (1 - w_2)((m_2)^2 - (m_1)^2) \\ &> \frac{m_1}{2} \cdot m_1((1 - w_1)^2 - (1 - w_2)^2) + \frac{w_2}{2} \cdot w_2(1 - m_2)^2 \end{aligned} \tag{69}$$

Finally, it is trivial to check that (69) holds for the realization of Example 5 in (47).

C Monte Carlo Simulation

We assume $F = G = \mathrm{U}[0,1]$ and $U(w \mid m) = V(m \mid w) = mw/2$. To compute the probability of sequential stability for fixed values of n and k, we make repeated draws of the $2n$ agents' types in the first period (from the $\mathrm{U}[0,1]$ distribution), and for each draw we enumerate all possible early matching scheme to check if one is sequentially stable in the sense of Definition 2. To check sequential stability, we need an explicit formula of an agent' expected utility in the second period, given a list of first-period agents who are waiting for the second period. We now normalize $U(w \mid m) = mw/2$ to w and derive this formula from the man's perspective. This formula should be useful for future research when k is large.

Without loss of generality, let $L = \{m_i : 1 \le i \le r\} \cup \{w_i : 1 \le i \le r\}$ be a ranked[19] list of first-period men and women who wait for the second period. Man m_i's expected utility

[19] That is, we have $m_1 \le \ldots \le m_r$ and $w_1 \le \ldots \le w_r$.

in the second period is:

$$\mathbf{U}(L, m_i) \equiv \sum_{j=0}^{k} \binom{k}{j} (m_i)^j (1-m_i)^{k-j} \cdot \mathbb{E}_k[\, w^2_{i+j} \,],$$

where w^2_{i+j} is the $(i+j)$-th lowest woman in the second period, among the k new arrivals and $w_1 \leq \ldots \leq w_r$.

An explicit formula for $\mathbb{E}[\, w^2_s \,]$, where $s \leq k+r$, can be obtained as follows:

$$\begin{aligned}
\mathbb{E}_k[\, w^2_s \,] &= \int_0^1 \mathbb{P}_k(w^2_s \geq x)\, dx \\
&= \int_0^{w_1} \sum_{l=0}^{s-1} \binom{k}{l} x^l (1-x)^{k-l}\, dx \\
&\quad + \int_{w_1}^{w_2} \sum_{l=0}^{s-2} \binom{k}{l} x^l (1-x)^{k-l}\, dx \\
&\quad + \ldots \\
&\quad + \int_{w_{s-1}}^{w_s} (1-x)^k\, dx \\
&= \sum_{l=1}^{s} \int_0^{w_l} \binom{k}{s-l} x^{s-l} (1-x)^{k-(s-l)}\, dx \\
&= \frac{1}{k+1} \sum_{l=1}^{s} (1 - B(s-l \mid w_l, k+1)),
\end{aligned} \tag{70}$$

where if $s > r$, we set $w_{r+1} = \ldots = w_s = 1$ (which does not effect the value of w^2_s because $s \leq k+r$), and $B(\,\cdot \mid p, n)$ is the CDF of a binomial distribution with n independent trials and probability p of success in each trial. Note that if $w_1 = w_2 = \ldots = w_s = 1$, then Equation (70) gives $\mathbb{E}_k[\, w^2_s \,] = s/(k+1)$, clearly the right answer.