Stability of Marriage with Externalities

Isa Emin Hafalir*
Department of Economics, Penn State University
University Park, PA, 16802, USA

December 10, 2004

Abstract

In many matching problems, it is natural to consider that agents may have preferences not only over the set of potential partners but over the whole matching. Once such externalities are considered, the set of stable matchings will depend on what agents believe will happen if they deviate. Sasaki and Toda (1996, *J. of Econ. Theory*, 70, 93) have examined the existence of stable matchings when the beliefs are exogenously specified and shown that stable matchings do not always exist. In this paper, we argue that beliefs should be endogenously generated, that is, they should depend on the preferences. We introduce a particular notion of endogenous beliefs, called sophisticated expectations, and show that with these beliefs, stable matchings always exist.

JEL Classification Numbers: C71, C78, D62.
Keywords: Cooperative Games, Matchings, Externalities.

1 Introduction

In the standard model of matching, introduced by Gale and Shapley [1], agents on one side of the market, say men, are assumed to have preferences over agents on the other side of the market, say women, and vice-versa. The central concern is to identify matchings that are stable in the sense that no unmatched pair of agents prefer each other to their current matches. The reader is referred to Roth and Sotomayor [3] for a detailed discussion of the literature.

Matching models are natural in many contexts—marriages, college admissions and labor markets. In many of these situations, some or all agents may be subject

*I would like to thank Matthew Jackson, Tarik Kara, Semih Koray and especially Kalyan Chatterjee, Federico Echenique and Manabu Toda for their comments and suggestions. I am indebted to Vijay Krishna for his guidance, support and editorial help.
Email address: isaemin@psu.edu

to externalities. A firm may care not only about the quality of its own employees but also the quality of the employees of a competitor. Even in the marriage context, jealousy may play a role.

When such externalities are present, a deviating pair needs to consider how other agents will react to the deviation. Consider a software firm that considers luring away a programmer of a competitor. How the rival firm reacts to this—perhaps by replacing the person with even a better programmer—will affect how profitable the deviation is. In such a situation, the expectations that the original firm has about the reactions of the rival become an important component.

In an important paper, Sasaki and Toda [5] first considered matching problems with such externalities. In their model of one-to-one matching, also called a marriage market, the idea of externalities is captured by specifying preferences over the complete matching rather than just over the set of agents on the other side of the market. The expectations of a deviating pair are specified via what they call *estimation functions*. An estimation function specifies the set of matchings among all other agents that the deviating agents consider to be possible. Given such estimation functions, a matching is blocked if there exists an unmatched pair who are both better off no matter what matching in their estimation arises. A matching that is not blocked is said to be *stable*. Notice that this formulation is non-Bayesian—that is, players do not assign probabilities to the different matches but rather deviate only if the deviation is profitable for all possible matchings that they consider to be possible. Sasaki and Toda [5] show, however, that only the *universal* estimation function—one that considers all matchings to be possible—is compatible with the existence of a stable matching. In their model, however, the estimation functions are specified *exogenously*. In particular, the set of matchings that a deviating pair considers possible does not depend on the preferences of the other agents. But clearly the matchings among other agents that will result from the deviation will depend on their preferences and this dependence should be recognized by the deviating pair.[1]

In this paper, we adapt Sasaki and Toda's [5] model and introduce *endogenous* estimation functions—that is, the set of matchings considered possible by a deviating pair depends on the preferences of other agents. Specifically, we define a plausible notion of *sophisticated expectations* that are not universal, but still guarantee that the set of stable matchings is not empty (Proposition 2 in Section 3).

Sophisticated expectations are determined via an algorithm. The algorithm, defined formally in Section 3, is based on the idea that given some estimation functions, one can define a naturally arising induced matching game in which there are no externalities. The induced matching game has a nonempty set of stable matchings. Sophisticated agents would recognize this and append such stable matchings to their original estimations. The new estimation functions so determined will define a new

[1] Roy Chowdhury [4] studies a model in which the agents assume that a deviation will trigger no response from others (agents are allowed to remain single). Stable matches then exist only under strong assumptions on preferences.

induced matching problem and give a new set of stable matchings, which are then also appended in the same way. The process will continue until there are no additional matchings in the induced game. The resulting estimation function is what we call sophisticated.

The notion of sophisticated expectations is different from that of *rational expectations*. The latter notion leads to estimation functions that satisfy a kind of reduced game consistency common in cooperative game theory—essentially, only stable matchings of the reduced game are considered possible. But as anticipated by Sasaki and Toda [5], rational expectations do not guarantee a nonempty stable set.

In addition to showing that the set of stable matchings is nonempty when agents use sophisticated estimation functions, we also provide a sufficient condition for estimation functions to be compatible with the general existence of the stable matchings (Proposition 3). This result may be used to find other plausible estimation functions.

In Section 4, we consider the question of deviations by larger coalitions—recall that the notion of stability is based on pairwise deviations. As is well-known, in matching models without externalities allowing for deviations by larger coalitions does not affect the set of unblocked matchings. In other words, the core is the same as the set of stable matchings. This equivalence, however, does not extend to situations with externalities. Indeed, it is known that the core may be empty when externalities are present. We thus examine a more permissive notion, that of the *bargaining set*. Although the bargaining set may, in general, also be empty, we provide a sufficient condition on preferences that ensures that this is not so (Proposition 4).

2 The Model and Exogenous Estimations

This paper models two-sided one-to-one matchings. These are called marriage markets in the literature and so let M and W denote the finite sets of men and women and $M \cap W = \varnothing$. We suppose that there are equal numbers of each so that $|M| = |W| = n$.

A bijection $\mu : M \cup W \to M \cup W$ is called a *matching* if (i) $\mu(\mu(a)) = a$ for all $a \in M \cup W$; (ii) $\mu(m) \in W$ for all $m \in M$ and $\mu(w) \in M$ for all $w \in W$. [2]

Thus $(m, w) \in \mu$ means that m and w are paired with each other in the matching μ. $A(M, W)$ denotes the set of all matchings and $A(m, w) = \{\mu \in A(M, W) \mid (m, w) \in \mu\}$ denotes the set of matchings where m and w are matched with each other. Each $a \in M \cup W$ has a strict preference ordering $\succ_a$ over $A(M, W)$. Note that this is the most general way of representing externalities since agents have preferences over the complete set of matchings. Let $\succ$ denote a preference profile, that is, $\succ = \{\succ_a \mid a \in M \cup W\}$. The triplet $(M, W, \succ)$ is called a *matching problem with externalities*.

Let $\varphi_m(w) \subseteq A(m, w)$ be the set of matchings which m considers *possible* when w is matched with him. Similarly, let $\varphi_w(m) \subseteq A(m, w)$ be the set of matchings which

[2] We do not allow agents to be single.

w considers to be possible when m is matched with her. Sasaki and Toda [5] call φ_m and φ_w *estimation functions*.

Given a profile of estimations φ, a matching μ is *φ-admissible* if for any pair $(m, w) \in \mu$,

$$\mu \in \varphi_m(w) \text{ and } \mu \in \varphi_w(m) \tag{1}$$

Given an estimation profile φ, a matching μ is *blocked* by a pair $(m, w) \notin \mu$ if for all $\mu' \in \varphi_m(w)$ and for all $\mu'' \in \varphi_w(m)$,

$$\mu' \succ_m \mu \text{ and } \mu'' \succ_w \mu \tag{2}$$

A matching μ is *φ-stable* if it is φ-admissible and is not blocked. Let $S_\varphi(M, W, \succ)$ denotes the set of all φ-stable matchings.

Sasaki and Toda [5] establish that if the estimation functions for a pair (m, w) such that for at least one of them the estimation function is not the set of all matches, then there exists a preference profile such that the set of φ-stable matchings is empty.

Proposition 1 *For any $n \geq 3$, if there exists an (m, w) pair such that either $\varphi_m(w) \neq A(m, w)$ or $\varphi_w(m) \neq A(m, w)$, then there exists a preference profile $\succ$ such that $S_\varphi(M, W, \succ) = \varnothing$.*

As the statement of the proposition makes apparent, the estimation functions are assumed to be exogenously given—in particular, they do not depend on the preference profile. This seems unnatural, however, since the set of potential matches that m estimates as being likely when he is matched with w, may well depend on the preferences of the other $2n - 2$ agents. As an extreme case, suppose that there is another pair (m', w') such that their preferences are not subject to any external effects but and each considers the other to the best mate. Then it seems natural that every estimation function for both m and w should only allow m' to be matched with w'. Thus estimation functions cannot be independent of the preference profile.

For instance, it seems natural that given a preference profile $\succ$, only estimation functions satisfying the following minimal condition (which is weaker than above extreme case) should be admitted.

Definition 1 *An estimation φ has No Matched-Couple Veto Property (NMCVP) if the following condition is satisfied: Let $(m, w), (m', w') \in \mu$ for some $\mu \in A(M, W)$. If for each $a \in M \cup W - \{m, m', w, w'\}$ and each $\mu^a \in A(m, w) \setminus A(a, \mu(a))$, $\mu \succ_a \mu^a$ then $\mu \in \varphi_m(w) \cap \varphi_w(m)$.*

It can be shown that if we require that estimation functions satisfy NMCVP, then the proof of Proposition 1 no longer goes through. It is because that constructed preference profile in the proof of the above proposition does not satisfy NMCVP. While natural, NMCVP is by itself not the only property that we would like estimation functions to satisfy.

Sasaki and Toda [5] also establish that if for all (m, w), the estimation function is the set of all matches ($\varphi_m(w) = \varphi_w(m) = A(m, w)$), then for any preference profile the set of stable matches in nonempty. But like Proposition 1, this result also relies on the fact that agents consider all matches to be possible even though these may be "irrational" given the preference profile.

We will show below that given any preference profile there exist endogenously generated estimation functions—which do not always equal the set of all matches—such that the resulting stable set is nonempty. Below we develop a procedure for finding such estimation functions.

3 Endogenous Estimations

In this section, we proceed as follows. Fix a particular preference profile $\succ$. We will suppose that if m and w are paired with each other then they have the *same* set of feasible matches. Thus, with a slight abuse of notation we will denote by $\varphi(m, w)$ to be the set of matches considered feasible by both m and w, when m is going to be matched with w. Formally, $\varphi_m(w) = \varphi_w(m) \equiv \varphi(m, w)$. The assumption of equal expectations is not a restriction as we will demonstrate that there are endogenously generated estimation functions satisfying this property which result in nonempty stable sets.

We denote estimation profile of agents by $\varphi = \{\varphi(m, w) \mid (m, w) \in M \times W\}$. Note that $\varphi(m, w)$ depends on preferences of the agents but since $\succ$ is assumed to be fixed, the dependence of φ on $\succ$ is suppressed. Let $S_\varphi(M, W)$ denote the set of stable matchings.

3.1 Rational Expectations

One natural way to formulate the notion of an endogenous estimation is via a "reduced game" consistency condition. Specifically, suppose a pair (m, w) are paired with each other and "exit" the marriage market. Then we have a reduced game with sets $M' = M - \{m\}$ and $W' = W - \{w\}$. Let $S_\varphi(M', W')$ denote the set of stable matches in the reduced game with $n-1$ pairs and preferences $\succ_a'$ which are restrictions of $\succ_a$ to the set of matchings the remaining agents—that is, not including m and w. Agents are said to have *rational expectations* estimations if the estimation of m and w is just $S_\varphi(M', W')$.

Formally, the rational expectations estimation function $\rho(M, W)$ is defined inductively as follows.

When $n = 2$, there are no externalities. This is because if a pair (m, w) match with each other, the only possibility is that other pair is matched as well. So for $n = 2$, the set of stable matchings $S(M, W)$ is nonempty (Gale and Shapley, [1]) and does not depend on any estimation function. This means that when $n = 3$, consistency demand that we set $\rho(m,w) = S(M',W')$ where $M' = M - \{m\}$ and $W' = W - \{w\}$. Now for $n = 3$, denote by $S_\rho(M,W)$ the set of stable matchings.

If this is not empty, we can proceed to the next step and again set for $n = 4$, $\rho(m,w) = S_\rho(M',W')$ where again $M' = M - \{m\}$ and $W' = W - \{w\}$ and so on.

This notion is well defined, however, only if at every stage the set of stable matchings is nonempty. But, as shown by Sasaki and Toda [5] as well, this is false. The following example for $n = 3$ shows that the rational expectations estimation function does not guarantee a nonempty stable set.

Example 1 *Let $n = 3$. Then we have six different matchings, and each agent has preferences over these. Suppose that agents assign utilities 1 to 6 to starting from the least preferred match to the most preferred match (these are only ordinal). Consider the following preferences:*

$$\mu_1 \begin{array}{ccc} 1 & 5 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 3 & 2 & 4 \end{array} \qquad \mu_2 \begin{array}{ccc} 4 & 1 & 6 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 3 & 5 & 2 \end{array} \qquad \mu_3 \begin{array}{ccc} 3 & 6 & 1 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 3 & 4 & 4 \end{array}$$

$$\mu_4 \begin{array}{ccc} 5 & 2 & 5 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 1 & 6 & 5 \end{array} \qquad \mu_5 \begin{array}{ccc} 6 & 4 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 1 & 6 & 5 \end{array} \qquad \mu_6 \begin{array}{ccc} 2 & 3 & 4 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 6 & 2 \end{array}$$

For $n = 2$, there are no externalities and reduced market stable sets, and so the rational expectations estimations can be easily obtained.

$\rho(m_1,w_1) = \mu_4$ *since* μ_1 *is blocked by* (m_3,w_2)
$\rho(m_1,w_2) = \mu_6$ *since* μ_2 *is blocked by* (m_2,w_1)
$\rho(m_1,w_3) = \mu_5$ *since* μ_3 *is blocked by* (m_3,w_1)
$\rho(m_2,w_1) = \mu_3$ *since* μ_6 *is blocked by* (m_1,w_3)
$\rho(m_2,w_2) = \mu_1$ *since* μ_5 *is blocked by* (m_3,w_3)
$\rho(m_2,w_3) = \mu_2$ *since* μ_4 *is blocked by* (m_3,w_1)
$\rho(m_3,w_1) = \mu_5$ *since* μ_2 *is blocked by* (m_2,w_2)
$\rho(m_3,w_2) = \mu_3$ *since* μ_4 *is blocked by* (m_2,w_1)
$\rho(m_3,w_3) = \mu_1$ *since* μ_6 *is blocked by* (m_2,w_2)

For $\rho(m,w) = S(M',W')$, μ_1 is blocked by (m_2,w_1). μ_2 is blocked by (m_2,w_1). μ_3 is blocked by (m_3,w_3). μ_4 is blocked by (m_2,w_1). μ_5 is blocked by (m_3,w_3). μ_6 is blocked by (m_2,w_2). So the stable set is empty.

3.2 Sophisticated Expectations

We now define another notion of an estimation function, called *sophisticated expectations*, and denoted by $\sigma(m,w)$. We will show that unlike the rational expectations, this new notion guarantees that the resulting stable set is nonempty. Recall that the estimation function framework is non-Bayesian. A pair of agents blocks only if the deviation is profitable no matter which matching in their estimation function arises. This is equivalent to saying that a block is profitable even if the worst matching in their estimation function arises. In this sense, deviating agents exhibit "maximin" behavior and it is as if their "point estimate" is the worst outcome. But given these pessimistic estimates, we can define an induced preference profile of agents over the agents on the other side as follows. The ranking that a particular $m \in M$ assigns to a $w \in W$, is the same as the ranking of the worst matching in $\varphi(m, w)$ among all other worst matchings in the collection $\varphi(m, w')$, $w' \in W$. This defines an induced matching game without externalities and we know that this game has a stable set. It is then natural to assume that m considers every matching in this stable set to be also possible and so appends this to his original estimation.

The notion is again defined inductively. For $n = 2$, the estimation function is always a singleton. Because if a pair of man and woman match with each other, the only possibility is that other pair match with each other as well. And we know that the stable set is nonempty for $n = 2$. With the assumption of $S_\sigma(M', W')$, the stable set for $|M'| = |W'| = k$, is nonempty for all (M', W'), we want to show that $S_\sigma(M, W)$, the stable set for $|M| = |W| = k + 1$, is also nonempty.

So suppose that $S_\sigma(M', W')$ is well defined and nonempty for all M' and W' satisfying $|M'| = |W'| = k$. Write $\sigma^1(m, w) = S_\sigma(M', W')$ and call it a 1^{st} degree estimation. Now suppose that if the pair (m, w) deviate then m will believe that as a result of this deviation, the *worst* matching in $\sigma^1(m, w)$ from his perspective will result. This induces a preference ordering for m over W. Similarly, suppose that w also believes that as a result of this deviation, the worst matching in $\sigma^1(m, w)$ from her perspective will result. Again, this induces a preference ordering for w over M. Thus a matching game *without* externalities, called the 1^{st} order induced matching problem is defined. This induced game has a stable set, and we call this the 1^{st} order induced matching stable set and denote it by I^1. For sophisticated agents, it is natural to assume that they will have I^1 also in their estimation functions. I^1, together $\sigma^1(m, w)$ form a new estimation function. We call this 2^{nd} degree estimation and denote it by $\sigma^2(m, w)$.

Then similarly, everybody will take into account the worst matching in $\sigma^2(m, w)$ as a result of their deviation. Again, this will generate an induced preference over the opposite sex and thus define a matching game without externalities. Again, this will have a stable set, called the 2^{nd} order induced matching problem. This game has a stable set, called the 2^{nd} order induced matching stable set and denoted by I^2. Similarly, for sophisticated agents I^2, together $\sigma^2(m, w)$ forms a new estimation function. We call this 3^{rd} order degree estimation and denote it by $\sigma^3(m, w)$ and so on. Agents sophisticated enough will proceed this way until the induced stable matching gives nothing new as compared to last degree estimation and this will be their final estimation function. We denote the final estimation function $\sigma(m, w)$. We call this the *sophisticated estimation function.* We will show that sophisticated expectations estimation function is compatible with the general existence of the stable set.

Formally, let $\sigma^1(m,w) = S_\sigma(M',W')$ and for $k = 1,2,3...$ let $\mu_m^{\sigma^k}(w)$ be the worst matching for m in $\sigma^k(m,w)$ and let $\mu_w^{\sigma^k}(m)$ be the worst matching for w in $\sigma^k(m,w)$. That is:

$$\begin{aligned}\mu_m^{\sigma^k}(w) &= \{\mu \in \sigma^k(m,w): \text{ for all } \mu' \in \sigma^k(m,w),\ \mu' \succeq_m \mu\} \\ \mu_w^{\sigma^k}(m) &= \{\mu \in \sigma^k(m,w): \text{ for all } \mu' \in \sigma^k(m,w),\ \mu' \succeq_w \mu\}\end{aligned}$$

Define the preference without externality $\succ^{\sigma^k}$as:

$$w \succ_m^{\sigma^k} w' \text{ iff } \mu_m^{\sigma^k}(w) \succ_m \mu_m^{\sigma^k}(w') \text{ and } m \succ_m^{\sigma^k} m' \text{ iff } \mu_w^{\sigma^k}(m) \succ_w \mu_w^{\sigma^k}(m')$$

$(M,W,\succ^{\sigma^k})$ is the k^{th} order induced matching problem of agents given the $\sigma^k(m,w)$'s. Let I^k denotes the stable set of $(M,W,\succ^{S_0})$. That is:

$$\begin{aligned}I^k &= \{\mu \in A : \nexists (m,w) \notin \mu \text{ s.t. } w \succ_m^{\sigma^k} \mu(m) \text{ and } m \succ_w^{\sigma^k} \mu(w)\} \\ &= \{\mu \in A : \nexists (m,w) \notin \mu \text{ s.t. } \mu_m^{\sigma^k}(w) \succ_m \mu_m^{\sigma^k}(\mu(m)) \text{ and } \mu_w^{\sigma^k}(m) \succ_w \mu_w^{\sigma^k}(\mu(w))\}\end{aligned}$$

Define $I^k(m,w) = \{\mu \in A(m,w) : \mu \in I^k\}$ to be projection of the stable set onto $M' \times W'$. One should not confuse $I^k(m,w)$ with $\sigma^k(m,w)$, the former is the set of stable matchings k^{th} order induced market in which m and w are matched with each other, and the latter is the set of matchings in the k^{th} order estimation function of m and w.

Now for $k = 1,2,3...$ inductively define,

$$\sigma^{k+1}(m,w) = \sigma^k(m,w) \cup I^k(m,w)$$

Since $S_\sigma(M',W')$ was assumed to be nonempty, σ^k and I^k are well defined and nonempty (Gale and Shapley, 1962).

Note that for all $(m,w) \in M \times W$ and for $k = 1,2,3...$

$$\sigma^k(m,w) \subseteq \sigma^{k+1}(m,w) \text{ and } \sigma^k(m,w) \subseteq A(m,w).$$

Thus the σ^k sequence of sets is monotone and the set of all matchings is finite Thus a limit exists. Now let

$$\sigma(m,w) = \lim_{k\to\infty} \sigma^k(m,w)$$

and denote by $I(m,w)$ the induced market stable set for preferences $\succ^\sigma$. Note that

$$I(m,w) \subset \sigma(m,w), \quad \forall (m,w) \in M \times W. \tag{3}$$

In words, agents consider the stable matchings in the reduced market, $\sigma^1 = S_\sigma(M',W')$, are possible but they also know that everybody is a "maximin" player since he or she considers the worst matching the estimation function to be result of any deviation in which he or she is involved. So they consider the stable matchings in the induced market, I^1, also to be possible. In this way, at every stage, they expand their estimation functions and eventually for a large N, I^N will not give any matching not in σ^N. $\sigma^N \equiv \sigma$ will be their final estimation functions. Note that this estimation function also satisfies NMCVP.

Proposition 2 *For the estimation function σ, the stable set S_σ is nonempty. In fact, $I \subset S_\sigma$.*

Proof. First of all, note that the non-blocking condition (see (2)) can be rewritten as: $\nexists (m,w) \notin \mu$ such that $(\mu_m^\sigma(w) \succ_m \mu$ and $\mu_w^\sigma(m) \succ_w \mu)$. In other words, the worst matching in $\sigma(m,w)$ should be better for m and w compared to μ. Now take any $\mu \in I$. From Gale and Shapley [1], we know that such μ exists. We will show that $\mu \in S_\sigma$.

Since $\mu \in I$ we have that for all $(m,w) \in \mu$, $\mu \in I(m,w)$ and since $I(m,w) \subset \sigma(m,w)$ from (3), we have $\mu \in \sigma(m,w)$ for all $(m,w) \in \mu$. So μ satisfies σ-admissibility (see (1)).

To show that μ is not blocked, we argue by contradiction. So suppose that there exists a pair $(m,w) \notin \mu$ such that $\mu_m^\sigma(w) \succ_m \mu$ and $\mu_w^\sigma(m) \succ_w \mu$. However, since $\mu \in \sigma(m,\mu(m))$ and $\mu \in \sigma(\mu(w),w)$ admissibility implies that $\mu \succeq_m \mu_m^\sigma(\mu(m))$ and $\mu \succeq_w \mu_w^\sigma(\mu(w))$. Hence, by transitivity of preferences $\mu_m^\sigma(w) \succ_m \mu_m^\sigma(\mu(m))$ and $\mu_w^\sigma(m) \succ_w \mu_w^\sigma(\mu(w))$, which contradicts the assumption that $\mu \in I$ because then (m,w) would block μ for preferences $\succ^\sigma$. ■

One may wonder whether $\sigma(m,w) = A(m,w)$ and so whether the result above follows from Sasaki and Toda's existence result. The next example demonstrates however, that we may have $\sigma(m,w) \subsetneq A(m,w)$ and so the stable set is nonempty even though the estimations function does not include all matchings.

Consider Example 1 again. Recall that there exists no stable matching for the rational expectations estimations $\rho(m,w)$.

Example 2 *Suppose $n = 3$ and consider the following utility assignments (1 denotes the least preferred and 6 the most preferred):*

$$\mu_1 \begin{array}{ccc} 1 & 5 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 3 & 2 & 4 \end{array} \qquad \mu_2 \begin{array}{ccc} 4 & 1 & 6 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 3 & 5 & 2 \end{array} \qquad \mu_3 \begin{array}{ccc} 3 & 6 & 1 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 3 & 4 & 4 \end{array}$$

$$\mu_4 \begin{array}{ccc} 5 & 2 & 5 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 1 & 6 & 5 \end{array} \qquad \mu_5 \begin{array}{ccc} 6 & 4 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 1 & 6 & 5 \end{array} \qquad \mu_6 \begin{array}{ccc} 2 & 3 & 4 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 6 & 2 \end{array}$$

For $n = 3$, first find $\sigma^1(m, w) = S_\sigma(M', W')$.

$$\begin{array}{lll} \sigma^1(m_1, w_1) = \mu_4 & \sigma^1(m_1, w_2) = \mu_6 & \sigma^1(m_1, w_3) = \mu_5 \\ \sigma^1(m_2, w_1) = \mu_3 & \sigma^1(m_2, w_2) = \mu_1 & \sigma^1(m_2, w_3) = \mu_2 \\ \sigma^1(m_3, w_1) = \mu_5 & \sigma^1(m_3, w_2) = \mu_3 & \sigma^1(m_3, w_3) = \mu_1 \end{array}$$

So with $\mu^{\sigma^1}_{m_i}(w_j)$ and $\mu^{\sigma^1}_{w_i}(m_j)$ (which is trivial for this case, as $\sigma^1(m_i, w_j)$'s are singleton) values, we have the following induced matching problem with preferences $\succ^{\sigma^1}$:

$$\mu_1 \begin{array}{ccc} 5 & 5 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 1 & 2 & 4 \end{array} \qquad \mu_2 \begin{array}{ccc} 2 & 1 & 2 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 1 & 5 & 5 \end{array} \qquad \mu_3 \begin{array}{ccc} 6 & 6 & 1 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 1 & 4 & 4 \end{array}$$

$$\mu_4 \begin{array}{ccc} 5 & 1 & 1 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 1 & 5 & 4 \end{array} \qquad \mu_5 \begin{array}{ccc} 6 & 5 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 1 & 2 & 5 \end{array} \qquad \mu_6 \begin{array}{ccc} 2 & 6 & 3 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 4 & 4 \end{array}$$

In the induced matching market with the preferences $\succ^{\sigma^1}$ the stable set $I^1 = \{\mu_6\}$ as μ_1 is blocked by (m_2, w_1), μ_2 is blocked by (m_2, w_2), μ_3 is blocked by (m_3, w_3), μ_4 is blocked by (m_3, w_1), and μ_5 is blocked by (m_3, w_3). So $\sigma^2(m_i, w_j) = \sigma^1(m_i, w_j)$ except for $(m_i, w_j) \in \mu_6$. Thus

$$\begin{array}{lll} \sigma^2(m_1, w_1) = \mu_4 & \sigma^2(m_1, w_2) = \mu_6 & \sigma^2(m_1, w_3) = \mu_5 \\ \sigma^2(m_2, w_1) = \{\mu_3, \mu_6\} & \sigma^2(m_2, w_2) = \mu_1 & \sigma^2(m_2, w_3) = \mu_2 \\ \sigma^2(m_3, w_1) = \mu_5 & \sigma^2(m_3, w_2) = \mu_3 & \sigma^2(m_3, w_3) = \{\mu_1, \mu_6\} \end{array}$$

Hence with $\mu^{\sigma^2}_{m_i}(w_j)$ and $\mu^{\sigma^2}_{w_i}(m_j)$ (which is trivial except for (m_2, w_1) and (m_3, w_3)) values, we have the following induced matching problem with preferences $\succ^{\sigma^2}$:

$$\mu_1 \begin{array}{ccc} 5 & 5 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 1 & 2 & 2 \end{array} \qquad \mu_2 \begin{array}{ccc} 2 & 1 & 2 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 1 & 5 & 5 \end{array} \qquad \mu_3 \begin{array}{ccc} 6 & 3 & 1 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 1 & 4 & 4 \end{array}$$

$$\mu_4 \begin{array}{ccc} 5 & 1 & 1 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 1 & 5 & 5 \end{array} \qquad \mu_5 \begin{array}{ccc} 6 & 5 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 1 & 2 & 5 \end{array} \qquad \mu_6 \begin{array}{ccc} 2 & 3 & 3 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 4 & 2 \end{array}$$

In the induced matching market with above preferences of $\succ^{\sigma^2}$ we have $I^2 = \{\mu_1\}$ as μ_2 is blocked by (m_2, w_2), μ_3 is blocked by (m_3, w_3), μ_4 is blocked by (m_1, w_1), μ_5 is *blocked by* (m_3, w_3), *and* μ_6 *is blocked by* (m_2, w_2). *So* $\sigma^3(m_i, w_j) = \sigma^2(m_i, w_j)$ *except for* $(m_i, w_j) \in \mu_1$.

$$\begin{array}{lll}
\sigma^3(m_1, w_1) = \{\mu_4, \mu_1\} & \sigma^3(m_1, w_2) = \mu_6 & \sigma^3(m_1, w_3) = \mu_5 \\
\sigma^3(m_2, w_1) = \{\mu_3, \mu_6\} & \sigma^3(m_2, w_2) = \mu_1 & \sigma^3(m_2, w_3) = \mu_2 \\
\sigma^3(m_3, w_1) = \mu_5 & \sigma^3(m_3, w_2) = \mu_3 & \sigma^3(m_3, w_3) = \{\mu_1, \mu_6\}
\end{array}$$

So with $\mu^{\sigma^3}_{m_i}(w_j)$ *and* $\mu^{\sigma^3}_{w_i}(m_j)$ *(which is left for the reader to verify), we have the following induced matching problem with preferences* $\succ^{\sigma^3}$*:*

$$\mu_1 \begin{array}{ccc} 1 & 5 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 1 & 2 & 2 \end{array} \qquad \mu_2 \begin{array}{ccc} 2 & 1 & 2 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 1 & 5 & 5 \end{array} \qquad \mu_3 \begin{array}{ccc} 6 & 3 & 1 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 1 & 4 & 4 \end{array}$$

$$\mu_4 \begin{array}{ccc} 1 & 1 & 1 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 1 & 5 & 5 \end{array} \qquad \mu_5 \begin{array}{ccc} 6 & 5 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 1 & 2 & 5 \end{array} \qquad \mu_6 \begin{array}{ccc} 2 & 3 & 3 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 4 & 2 \end{array}$$

In the induced matching market with preferences $\succ^{\sigma^3}$ *we have* $I^3 = \{\mu_1\}$ *as* μ_2 *is blocked by* (m_2, w_2), μ_3 *is blocked by* (m_3, w_3), μ_4 *is blocked by* (m_2, w_1), μ_5 *is blocked by* (m_3, w_3), *and* μ_6 *is blocked by* (m_2, w_2). *Since for all* $(m_i, w_j) \in \mu_1$, $\mu_1 \in \sigma^3(m_i, w_j)$ *we have* $S_2(m, w) = S_3(m, w) = S_4(m, w) = \ldots.\ \forall (m, w) \in M \times W$.
So we have $\sigma(m, w) = \sigma^3(m, w)$. *And given* σ, *we have* $\mu_1 \in S_\sigma$.

3.3 Estimations with Nonempty Stable Sets

The following proposition gives a sufficient condition for estimation functions to be compatible with the existence of a stable set. As the previous subsection demonstrates sophisticated estimations σ have this property.

Suppose φ is any estimation function. Define I^φ to be induced market stable set for φ in a manner analogous to the definition of I^k in the previous subsection. That is,

$$I^\varphi = \{\mu \in A : \nexists (m, w) \notin \mu \text{ such that } w \succ^\varphi_m \mu(m) \text{ and } m \succ^\varphi_w \mu(w)\}$$

The next proposition shows that any estimation function φ with the property that for every preference profile, there exists a matching μ which is both φ-admissible and in the stable set of the induced market, results in a nonempty stable set S_φ.

Proposition 3 *For any estimation function* φ, *if* $\mu \in \varphi(m, w) \cap I^\varphi(m, w)$ *for all* $(m, w) \in \mu$, *then* $\mu \in S_\varphi$.

Proof. If $\mu \in \varphi(m,w) \cap I^{\varphi}(m,w)$ for all $(m,w) \in \mu$, then $\mu \in \varphi(m,w)$ for all $(m,w) \in \mu$. So the admissibility is obviously satisfied.

We now show that μ is also unblocked. Assume that $\exists(m,w) \notin \mu$ such that $\mu_m^{\varphi}(w) \succ_m \mu$ and $\mu_w^{\varphi}(m) \succ_w \mu$. Since $\mu \in \varphi(m,\mu(m))$ and $\mu \in \varphi(\mu(w),w)$ because of admissibility, we have $\mu \succeq_m \mu_m^{\varphi}(\mu(m))$ and $\mu \succeq_w \mu_w^{\varphi}(\mu(w))$. Transitivity of the preferences implies that $\mu_m^{\varphi}(w) \succ_m \mu_m^{\varphi}(\mu(m))$ and $\mu_w^{\varphi}(m) \succ_w \mu_w^{\varphi}(\mu(w))$, which contradicts $\mu \in I^{\varphi}$. Hence, $\mu \in S_{\varphi}$. ■

Proposition 3 is a generalization of Proposition 2 (since with sophisticated estimations, $I(m,w) \subset \sigma(m,w)$). It demonstrates that the sophisticated estimation function σ is not the only estimation function that results in a nonempty stable set. In particular, any estimation function φ such that $\sigma \subset \varphi$ will also result in a nonempty stable set. There may be other estimation functions with this property as well. Any matching which is both "seen as possible" by all pairs in the matching and "stable in induced market" is stable.

The proposition also shows that it is not necessary for all agents to be equally sophisticated. It is enough that the estimation functions of all agents are common knowledge and that each agent is sophisticated enough to deduce that the set of stable matchings of the induced problem are possible ($I_{\varphi}(m,w) \subset \varphi(m,w)$). While only an example, the sophisticated estimation functions of the previous subsection are particularly useful nevertheless, because they can be constructed by using an *explicit* algorithm (via the functions σ^k and I^k).

4 The Core and the Bargaining Set

In marriage markets without externalities, we know that the set of stable matchings and the core are equivalent. This is because blocking agents care only about their own matches. Put another way, in marriage markets without externalities the blocking coalition is not affected by the complementary coalition.

Once externalities are considered, however, the equivalence of the two notions is not immediate. The blocking coalition is affected by the complementary coalition. Because of this the core notion also requires the use of estimation functions. One can define the estimation functions of the agents when they form a (not necessarily a pair) coalition.

Ideally, one would like to define endogenously generated estimation functions for coalitions in the same manner as in that in the previous sections. But as an initial investigation, we assume (as do Sasaki and Toda, [5]) that members of a blocking coalition "estimate" all matchings to be possible; that is, the estimation functions are universal. Universal estimations provide the best circumstances for the core to be nonempty—if the core with universal estimation estimations is empty then it is empty with any other estimations. But Sasaki and Toda [5] provide an example in which the core is empty with universal estimations.

The notion of a bargaining set is more permissive than that of the core (see the survey by Maschler, [2]). It requires that blocks which are not credible—in the sense that some members of the block have some other better block options—are ruled out. Below we extend the standard notion of a bargaining set to allow for externalities, again assuming universal estimations. More formally, in our model, the definitions of a block, a counter block, the core and the bargaining set are as follows:

Definition 2 $[(M', W'), \mu']$ *is a block against a matching* μ *if coalition* M' *and* W' *are nonempty subsets of* M *and* W*, respectively with* $|M'| = |W'|$*,* $\mu' \in A(M', W')$ *and if for every* $\mu^* \in A(M - M', W - W')$*,*

$$(\mu' \cup \mu^*) \succeq_a \mu \text{ for all } a \in (M', W')$$

where $(\mu' \cup \mu^*)(a) = \mu'(a)$ *for all* $a \in (M', W')$ *and* $(\mu' \cup \mu^*)(a) = \mu^*(a)$ *for all other* a*.*

A matching is blocked by a coalition of agents, (M', W') if there is a matching, μ' that they can achieve themselves, such that whatever the complementary agents do, the members of the coalition are all better off.

Definition 3 *The Core is the set of all matchings which are not blocked.*

A block is not credible if it is blocked in turn. Formally,

Definition 4 $[(M'', W''), \mu'']$ *is a counter block against a block* $[(M', W'), \mu']$ *at* μ *if both* $[(M'', W''), \mu'']$ *and* $[(M', W'), \mu']$ *are blocks to* μ*,* $(M'' \cup W'') \cap (M' \cup W') \neq \varnothing$*, and for any* $a \in (M'' \cup W'') \cap (M', W')$*, there exists a* $\mu^a \in A(M - M', W - W')$ *such that for all* $\mu^* \in A(M - M'', W - W'')$*,*

$$(\mu'' \cup \mu^*) \succ_a (\mu \cup \mu^a)$$

A block $[(M', W'), \mu']$ is counter blocked by a coalition (M'', W'') if there exists another block, $[(M'', W''), \mu'']$, to the initial matching μ, such that this new coalition (M'', W'') has nonempty intersection with the old coalition (M', W') and all agents in this intersection prefer the new block $[(M'', W''), \mu'']$ to the old block $[(M', W'), \mu']$.

Definition 5 *The bargaining set is the set of all matchings such that every block is counter blocked.*

Even though the bargaining set is larger than the core, it too may be empty.

Example 3 *Suppose $n = 3$ and consider the following utility assignments (1 denotes the least preferred and 6 the most preferred):*

$$
\mu_1 \begin{array}{ccc} 5 & 5 & 5 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 5 & 3 & 5 \end{array}
\qquad
\mu_2 \begin{array}{ccc} 6 & 6 & 1 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 4 & 6 & 6 \end{array}
\qquad
\mu_3 \begin{array}{ccc} 3 & 3 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 3 & 3 & 6 \end{array}
$$

$$
\mu_4 \begin{array}{ccc} 4 & 1 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 4 & 2 & 5 \end{array}
\qquad
\mu_5 \begin{array}{ccc} 1 & 4 & 6 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 1 & 2 & 1 \end{array}
\qquad
\mu_6 \begin{array}{ccc} 2 & 2 & 4 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 2 & 4 \end{array}
$$

The matching μ_1 has only one block which is $[(m_1, w_2), (m_2, w_3)]$. μ_2 also has only one block which is $[(m_3, w_2)]$. μ_3 has three blocks, which are $[(m_1, w_1)], [(m_3, w_3)]$ and $[(m_1, w_1), (m_3, w_3)]$, among these the last one have no counter blocks. μ_4 has two blocks $[(m_3, w_3)]$ and $[(m_1, w_1), (m_3, w_3)]$, and among these the latter has no counter blocks. μ_5 has three blocks $[(m_1, w_1)]$, $[(m_1, w_1), (m_2, w_2)]$ and $[(m_1, w_2), (m_2, w_3)]$, among these the last one has no counter blocks. The matching μ_6 has 8 blocks: $[(m_1, w_1), (m_2, w_2), (m_3, w_3)]$ and all 1 and 2 combinations, and $[(m_1, w_2), (m_2, w_3)]$, among these the last one has no counter blocks. Hence, for this example the bargaining set (and hence also the core) is empty.

In the example, different agents consider different matchings to be the "worst." As the following proposition shows, if all agents agree on the worst matching then the bargaining set is indeed nonempty.

Proposition 4 *If there exists a matching which is worst for all of the agents, then the bargaining set is not empty.*

Proof. Let there be n couples in the marriage market and let $\underline{\mu}$ denote the worst matching.

We will show that if no matching μ other than $\underline{\mu}$ is in the bargaining set then $\underline{\mu}$ is in the bargaining set.

Suppose that all matchings $\mu \neq \underline{\mu}$ are not in the bargaining set. For all nonempty coalitions (M', W') with $|M'| = |W'|$, for all matchings $\mu' \in A(M', W')$ such that μ' is not equal to the projection of $\underline{\mu}$ —that is, for some $a \in M' \cup W'$, $\mu'(a) \neq \underline{\mu}(a)$, $[(M', W'), \mu']$ is a block against $\underline{\mu}$. We will show that there is a counter block for every block. We have two cases:

- Suppose $|M'| < n-1$. In this case, for $m \notin M'$ and $w \notin W'$, let $M'' = M' \cup \{m\}$ and $W'' = W' \cup \{w\}$ and $\mu''(m) = w$, $\mu''(a) = \mu'(a)$ for all $a \in M' \cup W'$. Then $[(M'', W''), \mu'']$ is a counter block against a block $[(M', W'), \mu']$.

- Suppose $|M'| \geq n-1$. In this case, $\mu^* \in A(M - M', W - W')$ is unique (it consists of either just one pair or the empty set). Let us denote the unique matching $(\mu' \cup \mu^*)$ by $\overline{\mu}$. From our assumption we know that the matching $\overline{\mu}$ is not in the bargaining set, so it has a block. Take any block of $\overline{\mu}$, say $[(M'', W''), \mu'']$. Note that μ'' cannot be equal to the projection of $\overline{\mu}$, so we have $(M'' \cup W'') \cap (M' \cup W') \neq \varnothing$. But because $[(M'', W''), \mu'']$ is a block against $\underline{\mu}$ and for all $a \in (M'' \cup W'') \cap (M' \cup W')$ and for all $\mu^* \in A(M - M'', W - W'')$, we have $(\mu'' \cup \mu^*) \succ_a \overline{\mu}$. Thus $[(M'', W''), \mu'']$ is a counter block against $[(M', W'), \mu']$ at $\underline{\mu}$.

■

Interestingly, there can be markets in which *only* the worst matching $\underline{\mu}$ is in the bargaining set. Consider the following example.

Example 4 *Suppose $n = 3$ and consider the following utility assignments (1 denotes the least preferred and 6 the most preferred):*

$$\mu_1 \begin{array}{ccc} 3 & 2 & 3 \\ m_1 & m_2 & m_3 \\ w_1 & w_2 & w_3 \\ 6 & 2 & 2 \end{array} \qquad \mu_2 \begin{array}{ccc} 5 & 3 & 5 \\ m_1 & m_2 & m_3 \\ w_2 & w_3 & w_1 \\ 4 & 4 & 5 \end{array} \qquad \mu_3 \begin{array}{ccc} 2 & 5 & 6 \\ m_1 & m_2 & m_3 \\ w_3 & w_1 & w_2 \\ 6 & 3 & 6 \end{array}$$

$$\mu_4 \begin{array}{ccc} 4 & 6 & 4 \\ m_1 & m_2 & m_3 \\ w_1 & w_3 & w_2 \\ 4 & 3 & 3 \end{array} \qquad \mu_5 \begin{array}{ccc} 6 & 4 & 2 \\ m_1 & m_2 & m_3 \\ w_3 & w_2 & w_1 \\ 5 & 5 & 2 \end{array} \qquad \mu_6 \begin{array}{ccc} 1 & 1 & 1 \\ m_1 & m_2 & m_3 \\ w_2 & w_1 & w_3 \\ 1 & 1 & 1 \end{array}$$

Observe that μ_6 is the worst matching for every agent. Matchings other than μ_6 are not in the bargaining set: μ_1 has five blocks: $[(m_2, w_3)]$, $[(m_3, w_2)]$, $[(m_1, w_2), (m_2, w_3)]$, $[(m_1, w_3), (m_2, w_2)]$, and $[(m_2, w_3), (m_3, w_2)]$; and notice that the last two do not have any counter blocks. The matching μ_2 has one block: $[(m_1, w_3), (m_2, w_2)]$; μ_3 has one block: $[(m_1, w_1)]$; μ_4 has one block $[(m_1, w_2), (m_3, w_1)]$; and μ_5 also has one block $[(m_2, w_1), (m_3, w_2)]$.
Every block of μ_6 has a counter block. Hence μ_6 is the only matching in the bargaining set.

The example illustrates a phenomenon associated with the notion of a bargaining set. A Pareto inefficient outcome—in this case, the worst matching—is the only one that survives the "credible" block test.

5 Concluding Remarks

When externalities are present, the expectations that agents in a coalition hold regarding the complementary coalition are crucial in determining what outcomes are stable. Marriage markets with externalities seem to be relatively simple context in which one may begin to explore what kinds of expectations are natural.

In this paper, we have presented a model in which the expectations that agents in a (pairwise) coalition hold regarding the complementary coalition are endogenously determined—that is, they are consistent with the preferences of agents in the complementary coalition. Sasaki and Toda [5] showed that the rational expectations are, in general, incompatible with the existence of a stable set. At the same time, they also showed that exogenous expectations are also incompatible (except when they are all inclusive). We have defined a notion of sophisticated expectations that have the following features: (a) they are endogenously generated; (b) not all inclusive; and (c) lead to a nonempty stable set. We have also identified a general condition on estimation functions that is sufficient to guarantee a nonempty stable set. It remains for future work to see if there are other natural estimation functions—besides the sophisticated estimations defined here—that also satisfy the sufficient condition and can, perhaps, be extended to other coalitional settings.

6 References

1. D. Gale and L. Shapley, College admission and the stability of marriage, Amer. Math. Monthly 69, 9-15, (1962).

2. M. Maschler, The Bargaining Set, Kernel and Nucleolus: A Survey, in Handbook of Game Theory, (R. Aumannn and S. Hart, Eds) Amsterdam: Elsevier (1992).

3. A. Roth and M.A. Sotomayor, Two-Sided Matching: A Study in Game Theoretic Modelling and Analysis, Cambridge Unive. Press, UK, (1990).

4. P. Roy Chowdhury, Marriage Markets with Externalities, Working Paper, Indian Statistical Institute, Delhi, India (2004).

5. H. Sasaki and M. Toda, Two-sided matching problems with externalities, Journal of Economic Theory 70, 93-108, (1996).