STABILITY CONCEPTS IN MATCHING UNDER DISTRIBUTIONAL CONSTRAINTS

YUICHIRO KAMADA AND FUHITO KOJIMA

Abstract. Many real matching markets are subject to distributional constraints. To guide market designers faced with constraints, we propose new stability concepts. A matching is strongly stable if satisfying blocking pairs inevitably violates a constraint. We show that a strongly stable matching may not exist, and that existence is guaranteed if and only if all distributional constraints are trivial. To overcome this difficulty, we propose a more permissive concept, weak stability. We demonstrate a weakly stable matching always exists, implies efficiency, and is characterized by standard normative axioms. These results are obtained in a more general environment than those in existing studies, accommodating a wide variety of applications. *JEL Classification Numbers*: C70, D47, D61, D63.

Keywords: matching, distributional constraints, efficiency, stability, weak stability, strong stability, strategy-proofness

---

Date: September 13, 2016.

Kamada: Haas School of Business, University of California, Berkeley, Berkeley, CA 94720, y.cam.24@gmail.com. Kojima: Department of Economics, Stanford University, Stanford, CA 94305, fkojima@stanford.edu. We are grateful to the Editor, the Associate Editor, and two referees of the journal, Mustafa Oguz Afacan, Péter Biró, Erich Budish, Sylvain Chassang, Vincent Crawford, Hisao Endo, Clayton Featherstone, Tamás Fleiner, Drew Fudenberg, Tadashi Hashimoto, John William Hatfield, Toshiaki Iizuka, Rob Irving, Ryo Jinnai, Onur Kesten, Scott Duke Kominers, Hideo Konishi, Mihai Manea, David Manlove, Taisuke Matsubae, Aki Matsui, Yusuke Narita, Muriel Niederle, Parag Pathak, Al Roth, Dan Sasaki, Tayfun Sönmez, Satoru Takahashi, William Thomson, Alexis Akira Toda, Kentaro Tomoeda, Utku Ünver, Jun Wako, Alex Westkamp, Yosuke Yasuda, and participants at numerous seminars and conferences for helpful comments. Doctors Keisuke Izumi, Yoshiaki Kanno, Masataka Kawana, and Masaaki Nagano answered our questions about medical residency in Japan and introduced us to the relevant medical literature. We are grateful to officials at the Ministry of Health, Labor and Welfare and the Japan Residency Matching Program for discussion. Jin Chen, Irene Hsu, Seung Hoon Lee, Bobak Pakzad-Hurson, Neil Prasad, Fanqi Shi, Pete Troyan, Akhil Vohra and Rui Yu provided excellent research assistance. Kojima acknowledges financial support from the National Research Foundation (through its Global Research Network Grant, NRF-2013S1A2A2035408) as well as the Sloan Foundation.

1. Introduction

Many real matching markets are subject to distributional constraints. For example, medical residency matching in Japan is subject to the "regional cap" constraint, which is an upper-bound constraint on the total number of residents that can be assigned in each region. Policies that are mathematically equivalent to the regional-cap policy can be found in many different applications, such as graduate school admission in China, college admission in several European countries, residency match in the U.K., and teacher assignment in Scotland.

For cases without any distributional constraint, the theory of two-sided matching has been extensively studied ever since the seminal contribution by Gale and Shapley (1962), and it has been applied to the design of clearinghouses in various markets in practice. Stability has emerged as the key feature to the success of matching market design; a matching is stable if there is no blocking pair, that is, there is no pair of agents (say a doctor and a hospital) who prefer matching with each other to accepting the current matching. Unfortunately, all stable matchings may violate the given distributional constraint. This fact poses a challenge to market designers faced with such constraints.

The present paper addresses this challenge by analyzing stability concepts that respect the existence of distributional constraints. In order to guide our pursuit of the "right" stability concepts, we build upon the idea that there are different types of blocking pairs. Based on this idea, we consider stability concepts that require certain blocking pairs be eliminated while tolerating others.

Depending on the requirement placed on tolerated blocking pairs, we consider two related concepts, strong and weak stability. We begin by defining strong stability. We say that a matching is strongly stable if satisfying a blocking pair inevitably results in a violation of the distributional constraint.

While strong stability is perhaps the most natural stability concept under distributional constraints, we find a number of senses in which this concept is too demanding and is unlikely to be useful for market designers faced with distributional constraints. First, we find that a strongly stable matching does not necessarily exist, unlike a stable matching without constraints. In addition, we show that no mechanism is strategy-proof for doctors and produces a strongly stable matching whenever one exists.

Given these findings, we seek necessary and sufficient conditions under which these negative conclusions about strong stability can be avoided. Our characterization result demonstrates that the cases under which the above negative conclusions can be avoided are exactly the cases in which the distributional constraint reduces to each individual

hospital's capacity constraint. Thus, the difficulty with strong stability is an inevitable conclusion as long as there exists a nontrivial constraint.

Motivated by these negative conclusions, we introduce a more permissive concept, which we call weak stability. We say that a matching is weakly stable if it eliminates two most intuitive, and therefore likely most worrisome, types of blocking pairs. More specifically, it requires that no blocking pair exists such that either (i) adding a doctor at the blocking hospital does not violate the distributional constraint or (ii) the blocking hospital prefers the blocking doctor to one of its current employees.

In a sharp contrast to strong stability, weak stability turns out to produce a number of positive results. First and foremost, a weakly stable matching always exists. Second, it can be characterized by feasibility, individual rationality, no justified envy, and non-wastefulness.[1] And yet, we also show that weak stability is strong enough to imply (constrained) efficiency. In particular, the concept is strong enough to exclude unappealing matchings such as those produced in the current Japanese and British residency matching clearinghouses, college admission in some European countries, Chinese graduate school admission, and Scottish teacher matching.[2] These results suggest that weak stability is a useful concept for the market designer who seeks a normatively appealing outcome.

All of these analyses are conducted in the environment with a general constraint structure. Specifically, For each input of the numbers of doctors at different hospitals, a function called the *feasibility constraint* judges if the given input is feasible or not. This setting entails the "partitional regional cap" models as in Kamada and Kojima (2015), in which the set of hospitals are partitioned into regions and each region is assigned an upper bound of the number of matched doctors. In addition, it also captures many other situations. For example, imagine there are two hospitals with government-subsidized seats, and one seat in the first hospital costs half of that of the second. In such a case the difference in their costs may lead to different "weights" as a natural constraint (e.g., because of the government's budget constraint), so that there is a cap on the sum of the number of doctors in the first hospital and half the number of doctors in the second hospital. Another possibility is that constraints do not form partitions, as in the case in which there are restrictions not only in terms of regions but also in terms of medical specialties.[3]

[1]Here non-wastefulness is adapted to account for distributional constraints.

[2]See Kamada and Kojima (2015) for details of these markets and senses in which their mechanisms may result in matchings that violate weak stability.

[3]Section 5.1 lists various possibilities that the model with a feasibility constraint can capture.

Perhaps the most important lesson of this paper is that a market designer should consider normative appeal of stability concepts when it is not straightforward to set up an objective of the design. In our context, the distributional constraints imply a possibility of non-existence of strongly stable matchings. This does not mean we should completely abandon the idea of stability as it entails normative implications. This is exactly why, when looking for a weaker stability concept, we not only aimed to guarantee existence but also sought characterizations in terms of normative criteria such as efficiency and the no-justified-envy property. The approach we take here may be a good reference point for the prevalent market design problems in which traditional concepts do not apply.

Related Literature. The present paper is closely related to Kamada and Kojima (2015, 2016). They define a concept that they call stability, which is stronger than weak stability but weaker than strong stability. In their setting, hard distributional constraints are supplemented by governmental goals that induce soft preferences over doctor distributions within each region. They show the existence of a stable matching with respect to the given governmental goal. Our contributions over Kamada and Kojima (2015, 2016) are the following. First, our stability concepts, both strong and weak, can be defined only based on hard distributional constraints, without any reference to soft governmental goals. Although stability using governmental goals can potentially achieve an outcome deemed more desirable by the policy maker, the information about such soft goals may not be available to the analyst and, in such a case, analyzing solution concepts independent of such information may be more useful. Second, even weak stability turns out to imply efficiency, and is strong enough to preclude unappealing matchings such as those produced in the current Japanese residency matching and other applications, thus serving as a tool to make a judgment in practical situations. Third, our results, both positive and negative, are shown under a more general constraint structure than in Kamada and Kojima (2015, 2016), accommodating a wide variety of applications. Lastly, and perhaps most importantly, the focus of the current paper (investigating stability concepts) is different from those in Kamada and Kojima (2015, 2016) (studying a mechanism for a given stability concept).

In another related work, Goto, Kojima, Kurata, Tamura, and Yokoo (2016) study matching under constraints with an engineering-oriented approach. The restriction on constraints they impose, which they call "heredity" (a term in discrete mathematics) is equivalent to the restriction on feasibility in our work. However, all of their and our results are independent of each other because our focus is not engineering-oriented but on conceptual issues about stability notions. The main difference is that Goto, Kojima,

Kurata, Tamura, and Yokoo (2016) study neither of our stability concepts, strong nor weak stability. Hence, none of our negative results about strong stability or our positive results about weak stability appears in their work. Instead, they present a new mechanism and study its properties. Their main result is that their mechanism is strategy-proof for doctors, and there is no analogous result in our paper. On the other hand, the outcome of their mechanism fails to satisfy our fairness requirement, namely the no-justified-envy property. Neither do they provide a characterization of their outcome in terms of standard properties as in our work.

The present paper is a contribution to an active literature on matching problems with various forms of constraints. Examples include Roth (1991) on gender balance in labor markets, Abdulkadiroğlu and Sönmez (2003), Abdulkadiroğlu (2005), Ergin and Sönmez (2006), and Hafalir, Yenmez, and Yildirim (2013) on diversity in schools, Westkamp (2013) on trait-specific college admission, Abraham, Irving, and Manlove (2007) on project-specific quotas in projects-students matching, Biró, Fleiner, Irving, and Manlove (2010) on college admission with varied tuitions and hierarchical constraints, and Budish, Che, Kojima, and Milgrom (2013) on object allocation problems under hierarchical constraints. We take a different approach from these papers. Specifically, instead of restricting the class of markets that the theory can be applied to, such as hierarchical constraints, we define a weaker stability concept for which we show existence and other desirable properties in general environments. In addition, in Section 5.3 we explore the difference and connection with the literature on affirmative action in school choice (Abdulkadiroğlu and Sönmez, 2003; Hafalir, Yenmez, and Yildirim, 2013).

The rest of this paper proceeds as follows. Section 2 introduces the model. Section 3 defines strong stability and finds senses in which strong stability is too strong a condition. Section 4 introduces weak stability and provides several positive results. Section 5 provides discussions, and Section 6 concludes. All proofs are relegated to the Appendix.

2. Model

Let there be a finite set of doctors D and a finite set of hospitals H. Each doctor d has a strict preference relation $\succ_d$ over the set of hospitals and the option of being unmatched (being unmatched is denoted by $\emptyset$). For any $h, h' \in H \cup \{\emptyset\}$, we write $h \succeq_d h'$ if and only if $h \succ_d h'$ or $h = h'$. Each hospital h has a strict preference relation $\succ_h$ over the set of subsets of doctors. For any $D', D'' \subseteq D$, we write $D' \succeq_h D''$ if and only if $D' \succ_h D''$ or $D' = D''$. We denote by $\succ = (\succ_i)_{i \in D \cup H}$ the preference profile of all doctors and hospitals.

Doctor d is said to be **acceptable** to h if $d \succ_h \emptyset$.[4] Similarly, h is acceptable to d if $h \succ_d \emptyset$. It will turn out that only rankings of acceptable partners matter for our analysis, so we often write only acceptable partners to denote preferences. For example,

$$\succ_d \colon h, h'$$

means that hospital h is the most preferred, h' is the second most preferred, and h and h' are the only acceptable hospitals under preferences $\succ_d$ of doctor d.

Each hospital $h \in H$ is endowed with a **capacity** q_h, which is a nonnegative integer. We say that preference relation $\succ_h$ is **responsive with capacity** q_h (Roth, 1985) if

(1) For any $D' \subseteq D$ with $|D'| \leq q_h$, $d \in D \setminus D'$ and $d' \in D'$, $(D' \cup d) \setminus d' \succeq_h D'$ if and only if $d \succeq_h d'$,
(2) For any $D' \subseteq D$ with $|D'| \leq q_h$ and $d' \in D'$, $D' \succeq_h D' \setminus d'$ if and only if $d' \succeq_h \emptyset$, and
(3) $\emptyset \succ_h D'$ for any $D' \subseteq D$ with $|D'| > q_h$.

In words, preference relation $\succ_h$ is responsive with a capacity if the ranking of a doctor (or the option of keeping a position vacant) is independent of her colleagues, and any set of doctors exceeding its capacity is unacceptable. We assume that preferences of each hospital h are responsive with some capacity q_h throughout the paper.

A **matching** μ is a mapping that satisfies (i) $\mu_d \in H \cup \{\emptyset\}$ for all $d \in D$, (ii) $\mu_h \subseteq D$ for all $h \in H$, and (iii) for any $d \in D$ and $h \in H$, $\mu_d = h$ if and only if $d \in \mu_h$. That is, a matching simply specifies which doctor is assigned to which hospital (if any).

A **feasibility constraint** is a map $f : \mathbb{Z}_+^{|H|} \to \{0, 1\}$ such that $f(w) \geq f(w')$ whenever $w \leq w'$ and $f(0) = 1$, where the argument 0 of f is the zero vector and $\mathbb{Z}_+$ is the set of nonnegative integers. The interpretation is that each coordinate in w corresponds to a hospital, and the number in that coordinate represents the number of doctors matched to that hospital. $f(w) = 1$ means that w is **feasible** and $f(w) = 0$ means it is not. If w' is feasible then any w with a weakly fewer doctors in each hospital must be feasible, too. In this model, we say that matching μ is **feasible** if and only if $f(w(\mu)) = 1$, where $w(\mu) := (|\mu_h|)_{h \in H}$ is a vector of nonnegative integers indexed by hospitals whose coordinate corresponding to h is $|\mu_h|$. The feasibility constraint distinguishes the current environment from the standard model. We allow for (though do not require) $f((|q_h|)_{h \in H}) = 0$, that is, it may be infeasible for all the hospitals to fill their capacities.

This general form of feasibility constraints fits various types of feasibility constraints in the real world. For example, it can describe a situation in which the set of hospitals

[4]We denote singleton set $\{x\}$ by x when there is no confusion.

is partitioned into different regions, and for each region there exists an upper bound of the number of doctors that can be matched. It also allows for the case in which such caps are applied to regions that do not form partitions. Furthermore, it is also possible that a cap is imposed on the weighted sum of the numbers of doctors across the hospitals in a given region. Section 5.1 formalizes those ideas using the language of the general feasibility constraint f, and explains that the constraints in the real matching markets that we mentioned in the Introduction can be represented.

Since the feasibility constraint is a primitive of the environment, we consider a constrained efficiency concept. A feasible matching μ is **(constrained) efficient** if there is no feasible matching μ' such that $\mu'_i \succeq_i \mu_i$ for all $i \in D \cup H$ and $\mu'_i \succ_i \mu_i$ for some $i \in D \cup H$.

To accommodate the feasibility constraint, we introduce new stability concepts that generalize the standard notion. For that purpose, we first define two basic concepts. A matching μ is **individually rational** if (i) for each $d \in D$, $\mu_d \succeq_d \emptyset$, and (ii) for each $h \in H$, $d \succeq_h \emptyset$ for all $d \in \mu_h$, and $|\mu_h| \leq q_h$. That is, no agent is matched with an unacceptable partner and each hospital's capacity is respected.

Given matching μ, a pair (d, h) of a doctor and a hospital is called a **blocking pair** if $h \succ_d \mu_d$ and either (i) $|\mu_h| < q_h$ and $d \succ_h \emptyset$, or (ii) $d \succ_h d'$ for some $d' \in \mu_h$. In words, a blocking pair is a pair of a doctor and a hospital who want to be matched with each other (possibly rejecting their partners in the prescribed matching) rather than following the proposed matching.

When the feasibility constraint does not bind (in the sense that $f((|D|+1)_{h\in H}) = 1$), a matching is said to be stable if it is individually rational and there is no blocking pair. Gale and Shapley (1962) show that there exists a stable matching in that setting. In the presence of a binding feasibility constraint, however, there may be no such matching that is feasible. Thus in some cases every feasible and individually rational matching may admit a blocking pair.

A **mechanism** φ is a function that maps preference profiles to matchings. The matching under φ at preference profile $\succ$ is denoted $\varphi(\succ)$ and agent i's match is denoted by $\varphi_i(\succ)$ for each $i \in D \cup H$.

A mechanism φ is said to be **strategy-proof for doctors** if there exist no preference profile $\succ$, doctor $d \in D$, and preferences $\succ'_d$ of doctor d such that

$$\varphi_d(\succ'_d, \succ_{-d}) \succ_d \varphi_d(\succ).^5$$

[5] We do not require strategy-proofness for both sides (i.e., the doctor side and the hospital side) but only consider its weakening. This is because there is no mechanism that produces a stable matching for

3. Strong Stability

The first notion presented below is meant to capture the idea that any blocking pair that will not violate the feasibility constraint should be considered legitimate, so the appropriate stability concept should require that no agents have incentives to form any such blocking pair.

For each $h \in H$, let e_h be an integer vector whose coordinate corresponding to h is one and all other coordinates are zero (and let $e_\emptyset$ be the zero vector).

Definition 1. Fix a feasibility constraint f. A matching μ is **strongly stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $f(w(\mu)+e_h-e_{\mu_d}) = 0$ and (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$.

As stated in the definition, only certain blocking pairs are tolerated under strong stability. Condition (ii) of this definition requires that h likes each of its existing doctors better than doctor d, so the only reason that h is interested in forming a blocking pair is that it wants to fill one of its vacant positions with d. Doing so will change the distribution of doctors across hospitals from $w(\mu)$ to $w(\mu) + e_h - e_{\mu_d}$. Then, condition (i) implies that the new distribution of doctors violates feasibility. In other words, strong stability requires that satisfying any blocking pair will inevitably lead to an infeasible matching. In this sense, strong stability requires that any blocking pair is "caused" by the feasibility constraint. Indeed, this concept reduces to the standard stability concept of Gale and Shapley (1962) if the feasibility constraint does not bind.

As explained above, strong stability appears to be the most natural definition of stability under a feasibility constraint. However, we present two senses in which this concept is too strong a requirement for the market designer to achieve. The first is that a strongly stable matching does not necessarily exist. The following example demonstrates this point.

---

all possible preference profiles and is strategy-proof for both sides even in a market without a feasibility constraint, that is, $f((|D|+1)_{h\in H}) = 1$ (Roth, 1982), which is a special case of our model. One good aspect of having strategy-proofness for doctors is that the matching authority can actually state it as the property of the algorithm to encourage doctors to reveal their true preferences. For example, the current webpage of the Japan Residency Matching Program (last accessed on June 10, 2016, http://www.jrmp.jp/01-ryui.htm) states, as advice for doctors, that "If you list as your first choice a program which is not actually your first choice, the probability that you end up being matched with some hospital does not increase [...] the probability that you are matched with your actual first choice decreases." In the context of student placement in Boston, strategy-proofness for the student side was regarded as a desirable fairness property, in the sense that it provides equal access for children and parents with different degrees of sophistication to strategize (Pathak and Sonmez, 2008).

Example 1 (A strongly stable matching does not necessarily exist)**.** There are two hospitals, h_1 and h_2, and the feasibility constraint is such that $f(1,0) = f(0,1) = 1$ and $f(1,1) = 0$.[6] Each hospital h has a capacity of $q_h = 1$. Suppose that there are two doctors, d_1 and d_2. We assume the following preferences:

$$\succ_{h_1}\colon d_1, d_2, \qquad \succ_{h_2}\colon d_2, d_1,$$
$$\succ_{d_1}\colon h_2, h_1, \qquad \succ_{d_2}\colon h_1, h_2.$$

First, in any strongly stable matching, there is exactly one doctor matched to some hospital. This is because matching two doctors violates the feasibility constraint, while (d_1, h_1) would constitute a blocking pair that is not tolerated in the definition of strong stability if no doctor is matched to any hospital. By symmetry, it suffices to consider the case in which d_1 is matched (and hence d_2 is unmatched). If d_1 is matched with h_1, then (d_1, h_2) is a non-tolerated blocking pair because it violates condition (i) of Definition 1. On the other hand, if d_1 is matched with h_2, (d_2, h_2) is a non-tolerated blocking pair because it violates condition (ii) of Definition 1. Therefore, a strongly stable matching does not exist in this market. □

Even if a strongly stable matching does not always exist, can we try to achieve a weaker desideratum? More specifically, does there exist a mechanism that selects a strongly stable matching whenever there exists one? We show that such a mechanism does not exist if we also require certain incentive compatibility: No mechanism is strategy-proof for doctors and produces a strongly stable matching whenever one exists. This is the second difficulty with strong stability. To see this point, consider the following example.

Example 2 (No mechanism that is strategy-proof for doctors selects a strongly stable matching whenever there exists one)**.** There are two hospitals, h_1 and h_2, and the feasibility constraint is such that $f(1,0) = f(0,1) = 1$ and $f(1,1) = 0$. Each hospital h has a capacity of $q_h = 1$. Suppose that there are two doctors, d_1 and d_2. We assume the following preferences:

$$\succ_{h_1}\colon d_1, d_2, \qquad \succ_{h_2}\colon d_2, d_1,$$
$$\succ_{d_1}\colon h_2, \qquad \succ_{d_2}\colon h_1.$$

In this market, by inspection one can show that there are exactly two strongly stable matchings,

$$\mu = \begin{pmatrix} h_1 & h_2 & \emptyset \\ d_2 & \emptyset & d_1 \end{pmatrix} \quad \text{and} \quad \mu' = \begin{pmatrix} h_1 & h_2 & \emptyset \\ \emptyset & d_1 & d_2 \end{pmatrix}.$$

[6]We note that this is a case of the "partitional regions" setting formalized in item 1 of Section 5.1.

Now, suppose that a mechanism chooses μ under the above preference profile $\succ$. Then d_1 is unmatched. Consider reported preferences $\succ'_{d_1}$ of d_1,

$$\succ'_{d_1}: h_2, h_1.$$

Then μ' is a unique strongly stable matching, so the mechanism chooses μ' at $(\succ'_{d_1}, \succ_{-d_1})$. Doctor d_1 is better off at μ' than at μ since she is matched to h_2 at μ' while she is unmatched at μ. Hence, d_1 can profitably misreport her preferences when the preference profile is $\succ$.

If a mechanism chooses μ' under the above preference profile $\succ$, then by a symmetric argument, doctor d_2 can profitably misreport her preferences when the preference profile is $\succ$. Therefore, there does not exist a mechanism that is strategy-proof for doctors and selects a strongly stable matching whenever one exists. □

The above examples show that a strongly stable matching need not exist, and there exists no mechanism that is strategy-proof for doctors and selects a strongly stable matching whenever one exists. Given these negative findings, we next seek conditions under which these problems do not occur. We begin by formalizing some concepts we use in the analysis.

Definition 2. A set of hospitals H and a feasibility constraint f **guarantee the existence of a strongly stable matching** if, for every D and $\succ$, there exists a strongly stable matching.

As the name suggests, this concept is a restriction on H and f such that a strongly stable matching exists no matter what hospital and doctor preferences are.

Definition 3. A set of hospitals H and a feasibility constraint f satisfy **independence across hospitals** if there exists $\bar{q}_h \in \mathbb{Z}_+ \cup \{\infty\}$ for each $h \in H$ such that $\{w | f(w) = 1\} = \{w | w_h \leq \bar{q}_h, \forall h \in H\}$.

Note that, under the condition in this definition, a profile of weights is feasible if and only if, for each h, the weight w_h for h is at most $\bar{q}_h$. Thus the constraint is placed on each hospital independently, without any reference to the relation between weights for different hospitals.[7] Therefore, independence across hospitals is an extremely strong restriction, requiring that the feasibility constraint reduces to each individual hospital's capacity constraint.

[7]In the context of regional caps, this corresponds to assuming that there be no region with multiple hospitals and a positive and finite cap.

Theorem 1. *Consider a market with H and f. The following three claims are equivalent:*

(1) *H and f satisfy independence across hospitals.*
(2) *H and f guarantee the existence of a strongly stable matching.*
(3) *Under H and f, There exists a mechanism that is strategy-proof for doctors and generates a strongly stable matching whenever one exists.*

This theorem strengthens the negative implications of the earlier examples. It implies that problems pointed out by the previous examples are valid whenever there is a non-trivial feasibility constraint. Therefore, this theorem suggests that the concept of strong stability is not appropriate as the desideratum in our context with a feasibility constraint.

4. Weak Stability

Given the negative result regarding strong stability, this section introduces a more permissive concept and studies its properties.

Definition 4. Fix a feasibility constraint f. A matching μ is **weakly stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $f(w(\mu) + e_h) = 0$ and (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$.

The difference of weak stability from strong stability is that in condition (i) of weak stability, e_{μ_d} is not subtracted in the argument of f. Thus, weak stability checks the feasibility of a "pseudo-matching" in which a blocking doctor is hypothetically added without being removed from the hospital that the doctor was originally matched to.[8] Notice that, as for the case of strong stability, weak stability also reduces to stability of Gale and Shapley (1962) when the feasibility constraint does not bind.

Since weak stability tolerates certain blocking pairs that do not violate the feasibility constraint, we do not necessarily claim that weak stability is the most natural stability concept. However, we show that weak stability satisfies two desirable properties, namely existence and efficiency. Moreover, weak stability is strong enough to exclude certain undesirable outcomes: for example, the real-market mechanisms mentioned in the Introduction (in Japan, China, Ukraine, UK, and Scotland) do not necessarily produce a weakly stable matching.[9] One of the major advantages of weak stability over strong stability is that the existence of a weakly stable matching is guaranteed.[10]

[8]In the context of regional caps, this corresponds to the condition that a blocking pair such that the doctor in the pair moves between two hospitals in the same region is tolerated.

[9]See Kamada and Kojima (2015) for the detail.

[10]Kamada and Kojima (2016) show existence of a matching that satisfies a stronger notion than weak stability by constructing a mechanism that is strategy-proof for doctors in the context of hierarchical

Theorem 2. *A weakly stable matching exists, and any weakly stable matching is (constrained) efficient.*

In our environment with general constraints, there is no easy way to adapt the deferred acceptance algorithm or its modification (such as the flexible deferred acceptance algorithm in Kamada and Kojima (2016)). Hence our proof for existence does not make use of these mechanisms; instead, it is a generalization of Sotomayor's (1996) proof of the existence of a stable matching in one-to-one matching markets without distributional constraints. We define what we call hospital-quasi-stable matching, a generalization of Sotomayor's simple matching, and use that to show existence.

Note that it follows from the efficiency result in Theorem 2 that any notion stronger than weak stability implies efficiency. In particular, strong stability defined in Section 3 and stability defined in Kamada and Kojima (2015) imply efficiency.[11]

As in the above theorem, it is of interest to know what normatively desirable properties are implied by weak stability. To investigate this question, we provide a characterization of weak stability. For that goal, we begin by introducing a few axioms.

A matching μ satisfies the **no-justified-envy** property if there exists no pair of doctors $d, d' \in D$ such that (i) $\mu_{d'} \succ_d \mu_d$ and (ii) $d \succ_{\mu_{d'}} d'$ or $\mu_{d'} = \emptyset$. In the definition, (i) says that d envies d', and (ii) says that the envy is justified. A matching μ is **non-wasteful** if there is no doctor-hospital pair (d, h) such that (i) $h \succ_d \mu_d$ and $d \succ_h \emptyset$, and (ii) $|\mu_h| < q_h$ and $f(w(\mu) + e_h) = 1$.

No-justified-envy and non-wastefulness conditions are standard normative requirements in the literature, although the latter axiom is adapted to the case with a feasibility constraint. More specifically, the requirement $f(w(\mu)+e_h) = 1$ is included in the definition of non-wastefulness because the feasibility constraint regulates what matchings are deemed feasible under distributional constraints. With these concepts, we are now ready to offer a characterization of weak stability.

Proposition 1. *Matching μ is non-wasteful, individually rational, feasible, and satisfies the no-justified-envy property if and only if it is weakly stable.*[12]

---

regional caps. However, Theorem 2 establishes existence under more general constraints, where existence is not guaranteed for the concept of Kamada and Kojima (2016).

[11]The fact that stability in Kamada and Kojima (2015) implies efficiency in the context of partitional regional caps is stated in Theorem 3 of that paper. Hence the efficiency result in Theorem 2 of the present paper is a generalization of that theorem.

[12]It is straightforward to see that these four axioms are independent. See Appendix B.1 for specific examples.

In matching problems without a feasibility constraint, stability is characterized by non-wastefulness, individual rationality, and the no-justified-envy property (Balinski and Sönmez, 1999). Proposition 1 generalizes this characterization to the case with a feasibility constraint. To address additional complications posed by feasibility constraints, our characterization adds feasibility as a new axiom and adapts non-wastefulness to environments with constraints. This proposition makes precise the way in which stability of Gale and Shapley (1962) is generalized to the setting with a feasibility constraint.

5. Discussions

This section provides discussions on several topics. In Section 5.1, we list various types of feasibility constraints and illustrate how they can be analyzed in our framework. Section 5.2 provides a domain restriction result, identifying the necessary and sufficient condition on hospital preference profiles for ensuring the existence of a strongly stable matching. In Section 5.3, we study the relationship between our model and several models of school choice with affirmative action constraints. Section 5.4 studies the existence of a strongly stable matching in a context of large random markets.

5.1. **Examples of Feasibility Constraints.** This paper investigates the question of what the "right" stability concept is, and the generality of feasibility constraint *per se* is not our focus. However, it is desirable to understand the extent to which our results apply. In this section, we present how we can express various distributional constraints using the language of feasibility constraints in our model. In all of the examples in this section, there is a set $R \subseteq 2^H \setminus \{\emptyset\}$ which we call the set of **regions**. For each region $r \in R$, there is a **regional cap** q_r, which is a nonnegative integer.

(1) **Partitional regions:** The set of hospitals H is partitioned into hospitals in different regions, that is, $r \cap r' = \emptyset$ if $r \neq r'$ for $r, r' \in R$, and $H = \cup_{r \in R} r$. A matching is feasible if $|\mu_r| \leq q_r$ for all $r \in R$, where $\mu_r = \cup_{h \in r} \mu_h$. In words, feasibility requires that the regional cap for every region is satisfied. This model is a special case of our general model presented in Section 2. To see this, define function f by

$$f(w) = \begin{cases} 1 & \text{if } \sum_{h \in r} w_h \leq q_r \text{ for every } r \in R, \\ 0 & \text{otherwise.} \end{cases} \tag{5.1}$$

By the construction of f, it is obvious that $f(w(\mu)) = 1$ if and only if μ is feasible in the sense of the model with partitional regions. Moreover, it is obvious that this function satisfies $f(w) \geq f(w')$ whenever $w \leq w'$ and $f(0) = 1$, which are the requirements for f to be a feasibility constraint. Therefore, the model

with partitional regions is a special case of our general model. Examples of constraints given by partitional regions include Japanese and UK medical matching and Chinese graduate school admission (Kamada and Kojima, 2015).

(2) **Hierarchical regions:** Assume that the set of regions R forms a hierarchy (but not necessarily a partition), that is, $r, r' \in R$ implies $r \subseteq r'$ or $r' \subseteq r$ or $r \cap r' = \emptyset$. A matching μ is feasible if $\sum_{h \in r} |\mu_h| \leq q_r$ for every $r \in R$. This model generalizes the one with partitional regions formalized in item 1 above, and is a special case of the general model in Section 2. The feasibility constraint in this case corresponds to function f as in (5.1), with the sole difference being that R can now be a more general hierarchy than a partition. Examples of constraints given by hierarchical regions include Hungarian college admission before 2007 (Biro, Fleiner, Irving, and Manlove, 2010).

(3) **General (non-hierarchical) regions:** Consider a model that is identical to the preceding case, except that we allow R to be an arbitrary subset of $2^H \setminus \{\emptyset\}$, not necessarily a hierarchy. Non-hierarchical regions pose a problem in many respects. For example, Kamada and Kojima (2016) show that, if R is not a hierarchy, there always exist regional caps and regional preferences such that a mechanism that always produces a stable matching that they define and is strategy-proof for doctors does not exist. Nevertheless, this model is again a special case of the general model studied in Section 2, corresponding to the definition of f as in (5.1). Examples of constraints given by non-hierarchical regions include Hungarian college admission after 2007 (Biro, Fleiner, Irving, and Manlove, 2010). Kamada and Kojima (2016) detail the distinction between hierarchical and non-hierarchical regions giving various examples.

(4) **Constraints on weighted sums of doctors:** Let us depart from the preceding models by allowing that doctors in different hospitals may contribute to a regional cap differently. More specifically, let R be an arbitrary subset of $2^H \setminus \{\emptyset\}$. For each $h \in H$ and $r \in R$, s_{hr} is a nonnegative real number, or a "score." A matching μ is feasible if $\sum_{h \in r} s_{hr} |\mu_h| \leq q_r$ for every $r \in R$. The preceding models are special cases of this model in which $s_{hr} = 1$ for every pair of h and r. Compared to the models discussed so far in this section, this model adds additional flexibility, by allowing that assignments in different hospitals count differently toward the constraints.[13] The feasibility constraint in this model can be expressed by function

[13]See the introduction for the motivation for allowing for such flexibility.

f defined by

$$f(w) = \begin{cases} 1 & \text{if } \sum_{h \in r} s_{hr} w_h \leq q_r \text{ for every } r \in R, \\ 0 & \text{otherwise.} \end{cases}$$

This function satisfies $f(w) \geq f(w')$ whenever $w \leq w'$ and $f(0) = 1$, which are the requirements for f to be a feasibility constraint. The feasibility constraint given in Example 3 is an instance of this general form of a constraint. The example of subsidized seats as mentioned in the Introduction fits this case.

5.2. **Domain Restriction for Strong Stability.** In this section, we consider a domain restriction for the existence of a strongly stable matching. More specifically, we identify the necessary and sufficient condition on hospital preference profiles for ensuring the existence of a strongly stable matching.

For this purpose, we first focus on the case in which the feasibility constraint is given through a set of regions that forms a hierarchy, as described in item 2 in Section 5.1 (a more general case is analyzed at the end of this section). Recall that the feasibility constraint is specified by a set of regions R and a profile of nonnegative integers $(q_r)_{r \in R}$, where the feasibility constraint is given by $f(w) = 1$ if and only if $\sum_{h \in r} w_h \leq q_r$ for every $r \in R$.

Given R and hospital preference profile $\succ_H \equiv (\succ_h)_{h \in H}$, we say that $\succ_H$ is **consistent with master lists of** R (Biro, Fleiner, Irving, and Manlove, 2010) if there exists a strict order $\succ_r$ over D for each region $r \in R$ such that, for each hospital $h \in r$, $d \succ_h d' \succ_h \emptyset$ implies $d \succ_r d'$, that is, $\succ_h$ coincides with the restriction of $\succ_r$ to the doctors who are acceptable to h. In particular, under the presumption that all doctors are acceptable to all hospitals, this condition is equivalent to the requirement that all hospital preferences in a region be identical to one another, a condition stronger than imposing the acyclicity condition of Ergin (2002) to hospitals in each region.

With this condition at hand, we are now ready to describe our domain restriction result.

Proposition 2. *Consider sets of doctors D and hospitals H, a hospital preference profile $\succ_H$, and a set of regions R that forms a hierarchy. The following two claims are equivalent:*

(1) *For every $(q_r)_{r \in R}$ and doctor preference profile, there exists a strongly stable matching.*
(2) *The hospital preference profile $\succ_H$ is consistent with master lists of R.*

This proposition can be interpreted as providing a further sense, in addition to Theorem 1, in which strong stability is hard to achieve. In our intended applications such

as medical match where hospitals have preferences over doctors, we expect that there is at least some heterogeneity in preferences. Therefore, in such a context, strongly stable matching is too restrictive to hope for.

On the other hand, there may be other contexts in which hospital preferences are consistent with master lists. An example is Scottish Foundation Allocation Scheme (SFAS), which matches doctors and hospitals in Scotland. In SFAS, the ranking of hospitals over doctors are decided by NHS Education for Scotland, the governing body of SFAS, in order of scores, and these rankings are common across all hospitals (Irving, 2011).[14] Another example is school choice with affirmative action constraints as formalized by Abdulkadiroğlu and Sönmez (2003). In that setting, given a school choice problem with affirmative action constraints, one can construct a hypothetical matching market with constraints in a partitional regional structure described in item 1 in Section 5.1, where consistency with master lists are guaranteed by the construction of the hypothetical market. We explore this connection further in Section 5.3.

When the feasibility constraint is not given by a hierarchy of regions, the conclusion of Proposition 2 does not necessarily hold. To see this point, consider the following example.

Example 3. Let there be three doctors, d_1, d_2, and d_3, and three hospitals, h_1, h_2, and h_3. Let the feasibility constraint f be represented by a constraint on weighted sums of doctors as in item 4 of Section 5.1. More specifically, let $r = \{h_1, h_2\}$ be the only region on which there exists a constraint and the feasibility constraint be given by

$$f(w) = \begin{cases} 1 & \text{if } 2w_{h_1} + w_{h_2} \leq 2, \\ 0 & \text{otherwise.} \end{cases}$$

Let the capacities of hospitals h_1, h_2, and h_3 be $1, 2$, and 1, respectively, and preferences of hospitals and doctors be given as follows:

$$\begin{array}{lll} \succ_{h_1}\colon d_2, d_1, d_3, & \succ_{h_2}\colon d_2, d_1, d_3, & \succ_{h_3}\colon d_3, d_2, \\ \succ_{d_1}\colon h_1, & \succ_{d_2}\colon h_3, h_2, h_1, & \succ_{d_3}\colon h_1, h_2, h_3. \end{array}$$

Note that $\succ_H$ is consistent with master lists in this market. Nevertheless, by inspection one can check that there exists no strongly stable matching in this market.[15] □

[14] We are grateful to a referee for suggesting the SFAS example and encouraging us to consider consistency with master lists.

[15] Appendix B.4 provides the detail.

This example shows that, with general feasibility constraints, even the existence of a master list does not ensure existence of a strongly stable matching. This again fortifies our claim that strong stability is not an appropriate desideratum.

5.3. **Relation with Affirmative Action in School Choice.** This section explores the relationship between our model and two models of school choice with affirmative action constraints, namely the model with type-specific quotas due to Abdulkadiroğlu and Sönmez (2003) and the model with type-specific reserves due to Hafalir, Yenmez, and Yildirim (2013).

First, we introduce a model of school choice with type-specific quotas due to Abdulkadiroğlu and Sönmez (2003), which we will refer to as the **AS model**. There exist a set of schools C and a set of students S. Each student has a strict preference relation over $S \cup \{\emptyset\}$. Each school has a priority order over S and capacity q_c. The set of students is partitioned into two types of students, the set of majority students S^M and the set of minority students S^m (the assumption that there are only two types is made purely for simplicity, and generalization to more than two types is straightforward). Let $T = \{M, m\}$ denote the set of types. For each $c \in C$ and $t \in T$, there is a **type-specific quota** q_c^t, which represents the maximum number of students of type t who can be matched to school c.[16]

In this context, a matching ν is **feasible** if $|\nu_c| \leq q_c$ and $|\nu_c \cap S^t| \leq q_c^t$ for all $c \in C$ and $t \in T$. A matching ν is **AS-stable** if it is feasible and

(1) $\nu_s \succeq_s \emptyset$ for each $s \in S$, and
(2) if $c \succ_s \nu_s$, then either (i) $|\nu_c| = q_c$ and $s' \succeq_c s$ for all $s' \in \nu_c$ or (ii) $|\mu_c \cap S^t| = q_c^t$ and $s' \succeq_c s$ for all $s' \in \nu_c \cap S^t$, where t is the type such that $s \in S^t$.

Abdulkadiroğlu and Sönmez (2003) show that an AS-stable matching always exists in this problem and can be found by a version of the student-proposing deferred acceptance algorithm.

We shall describe how to associate this model with the model of matching with constraints of the present paper, which we refer to as the **KK model** in this subsection. For this purpose, given an AS model, consider the following hypothetical KK model. The set of hospitals is defined as $H \equiv C \times T$, that is, each hospital is identified with a school c and a type $t \in T$. An interpretation is that a hospital $h = (c, t)$ corresponds to school c's seats that can potentially be assigned to type t students. Set the capacity q_h for $h = (c, t)$ as $q_h = q_c^t$. The feasibility constraint is given by partitional regions as in item 1

[16]Hafalir, Yenmez, and Yildirim (2013) consider a case in which there is no type-specific quota for type m. Such a case can be modeled in our framework by setting $q_c^m = q_c$.

of Section 5.1. More specifically, the set of regions are $\{\{(c, M), (c, m)\} : c \in C\}$ and the regional cap for region $\{(c, M), (c, m)\}$ is q_c.

Preferences of each hospital $h = (c, t)$ are identical to the preference of school c in the AS model. Meanwhile, the preferences of each student s in the KK model, which we denote as $\succ_s$ with slight abuse of notation, are constructed as follows: Let t be the type of student s. Then,

(1) In the KK model, $\emptyset \succ_s (c, t')$ for any $c \in C$ if $t' \neq t$.
(2) $(c, t) \succ_s \emptyset$ in the KK model if and only if $c \succ_s \emptyset$ in the AS model.
(3) $(c, t) \succ_s (c', t)$ in the KK model if and only if $c \succ_s c'$ in the AS model.

That is, the preferences of each student in the KK model agree with the preference in the AS model, but only finds school seats of her own type acceptable.[17]

Given a matching in the AS model, we associate with it a matching in the KK model in a natural manner: A student s is matched with a hospital (c, t) in the KK model if and only if s is of type t and matched with c in the AS model.

Proposition 3. *Suppose that a matching is stable in the AS model. Then, the matching associated to it is strongly stable in the KK model.*

The following example shows that the converse of Proposition 3 does not hold.

Example 4. In the AS model, let there be one school c as well as two students s_1 and s_2. Assume that s_1 is a majority type while s_2 is a minority type, $q_c = q_c^M = q_c^m = 1$, and preferences are given as follows:

$$\succ_c\colon s_1, s_2,$$

$$\succ_{s_1}\colon c, \qquad \succ_{s_2}\colon c.$$

Clearly, the matching

$$\nu = \begin{pmatrix} c & \emptyset \\ s_1 & s_2 \end{pmatrix},$$

is the unique AS-stable matching in the AS model. By contrast, in the KK model, both

$$\mu = \begin{pmatrix} (c, M) & (c, m) & \emptyset \\ s_1 & \emptyset & s_2 \end{pmatrix} \quad \text{and} \quad \mu' = \begin{pmatrix} (c, M) & (c, m) & \emptyset \\ \emptyset & s_2 & s_1 \end{pmatrix},$$

[17]We could assume that each hospital (c, t) finds only students of type t acceptable. This assumption may be intuitive given our interpretation that a hospital (c, t) represents the seats that can be assigned to students of type t. An analogous analysis can be done even if we restrict hospitals' preferences in such a manner, but we opted to do otherwise as we find it simpler notationwise. Since our ultimate objective is only to make a formal connection between the AS and KK models, the specific manner in which we do so is not essential, and indeed both methods achieve the same goal.

are strongly stable matchings. Matching μ is the matching associated with ν, while μ' does not have any AS-stable matching in the AS model that is associated with it. □

Next, we introduce a model of school choice with type-specific reserves due to Hafalir, Yenmez, and Yildirim (2013), which we will refer to as the **HYY model**. The HYY model is identical to the AS model except that, instead of type-specific quotas, for each $c \in C$ and $t \in T$, there is a **type-specific reserve** p_c^t, which represents the number of students of type t who should be matched to c as long as there are type-t students who demand a seat at c. We assume $\sum_{t \in T} p_c^t \leq q_c$.

In this context, a matching ν is **feasible** if $|\nu_c| \leq q_c$ for all $c \in C$. A matching ν is **HYY-stable** if it is feasible and

(1) $\nu_s \succeq_s \emptyset$ for each $s \in S$, and
(2) if $c \succ_s \nu_s$ where $s \in S^t$, then either (i) $|\nu_c| = q_c$, $|\nu_c \cap S^t| \geq p_c^t$, and $s' \succeq_c s$ for all $s' \in \nu_c$ or (ii) $|\nu_c| = q_c$, $|\nu_c \cap S^t| \geq p_c^t$, $|\nu_c \cap S^{t'}| \leq p_c^{t'}$ for $t' \neq t$, and $s' \succeq_c s$ for all $s' \in \nu_c \cap S^t$.[18]

Hafalir, Yenmez, and Yildirim (2013) show that an HYY-stable matching always exists in this problem, and it can be found by a version of the student-proposing deferred acceptance algorithm.

Hafalir, Yenmez, and Yildirim (2013) compare their model with the AS model by associating the HYY model with type-specific reserves p_c^t with an AS model with type-specific quotas $q_c^t = q_c - p_c^{t'}$ where $t' \neq t$. Motivated by this approach, given a HYY model, we consider an association that is identical to the association we defined in the case of the AS model, except that the capacity of a hospital $h = (c, t)$ is set as $q_h = q_c - p_c^{t'}$ where $t' \neq t$.

With the above association, the following example shows that even if a matching in the HYY model is HYY-stable, the matching in the KK model that is associated with that matching is not necessarily strongly stable (or even weakly stable).

Example 5. Let there be one school c as well as two students s_1 and s_2. Assume that s_1 is a majority type while s_2 is a minority type, $q_c = p_c^m = 1, p_c^M = 0$, and:

$$\succ_c \colon s_1, s_2,$$

$$\succ_{s_1} \colon c,$$

[18]The definition stated in Hafalir, Yenmez, and Yildirim (2013) contains a minor error, and the one provided here is their intended definition (Hafalir, Yenmez, and Yildirim, 2016). Our definition is also consistent with the definition given in the working paper version (Hafalir, Yenmez, and Yildirim, 2011).

and s_2 prefers $\emptyset$ the most. Clearly, the matching

$$\nu = \begin{pmatrix} c & \emptyset \\ s_1 & s_2 \end{pmatrix},$$

is the unique HYY stable matching in the HYY model. By contrast, in the KK model, the matching associated to ν, i.e.,

$$\mu = \begin{pmatrix} (c, M) & (c, m) & \emptyset \\ s_1 & \emptyset & s_2 \end{pmatrix},$$

is not feasible because $q_{(c,M)} = q_c - p_c^m = 0$. Therefore, matching ν is not strongly stable (or even weakly stable). □

A slight modification of Example 4, with the only change being that we set $p_c^M = p_c^m = 0$, shows that the implication from strong stability to HYY-stability also fails.

Remark 1. There may be more than one plausible manner in which one may associate a matching in the KK model given an HYY model. In particular, one problem with the above association is that even though the reserves in the HYY model do not affect feasibility and they are only used for HYY-stability, the KK model has them affect the feasibility constraint. One possibility to avoid such a feature is to set $q_h = q_c$ for $h = (c, t)$. With this approach, however, no information about the reserves is reflected anywhere in the KK model. Therefore, clearly there is no relationship between HYY-stability and strong stability.

The school choice model with soft bounds due to Ehlers, Hafalir, Yenmez, and Yildirim (2014) is more general than the HYY model, so the above examples also show that there is no logical relationship between their stability concept and strong stability in the KK model.

5.4. **Strong Stability in Large Markets.** One question of interest is whether a strongly stable matching exists in a typical instance of a matching market with constraints. To study this issue, we consider an environment in which preferences of doctors and hospitals are drawn from certain distributions and investigate the probability that a strongly stable matching exists. In particular, we study the asymptotic behavior of the existence probability as the numbers of doctors and hospitals grow.

Our exercise is partly motivated by the recent literature on large matching markets. In that literature, it has been recognized that various impossibilities in classical matching markets can be circumvented approximately in large markets. For example, although a stable matching does not necessarily exist when some doctors are couples (Roth, 1984),

Roth and Peranson (1999) find a stable matching for all years of data from NRMP they analyzed. Motivated by this finding, Kojima, Pathak, and Roth (2013) and then Ashlagi, Braverman, and Hassidim (2014) find conditions under which the probability that a stable matching exists in the presence of couples goes to one as the number of market participants approaches infinity. Che, Kim, and Kojima (2013) and Azevedo and Hatfield (2012) find asymptotic existence in the face of other kinds of preference complementarity.[19]

Given these possibility results in large markets, one may conjecture that the probability that a strongly stable matching exists converges to one as the market size goes to infinity. However, the following analysis shows that this statement does not necessarily hold, at least unless some further restrictions are imposed.

Let there be $n \geq 1$ doctors and $m \geq 1$ hospitals. The feasibility constraint f can be represented by the "partitional regions" model as in item 1 of Section 5.1: There is exactly one region that contains all the hospitals. The regional cap of this region is one. Each hospital has a positive (and arbitrary) capacity.

For each doctor, her preferences are drawn from the uniform distribution over all preferences in which all hospitals are acceptable. Similarly, for each hospital, its preferences are drawn from the uniform distribution over all preferences in which all doctors are acceptable. Preference draws are i.i.d. across doctors and hospitals.[20] Call this model with stochastic preferences the (n, m)-market model.

In this setting, we consider the limit as n varies from 1 to ∞ and m depends on n such that $m \to \infty$ as $n \to \infty$ (one example is the "balanced" markets, i.e., $m = n$, but unbalanced markets, i.e., $m \neq n$, are allowed; see Remark 3 for further discussion). Specifically, we derive the limit of the probability that a strongly stable matching exists, as follows:

Proposition 4. *The probability that there exists a strongly stable matching in the (n, m)-market model converges to $1 - \frac{1}{e} \simeq 0.632$ as n and m go to infinity, where e denotes the base of the natural logarithm.*

[19]See also Nguyen and Vohra (2016) who find the existence of a stable matching with couples in a neighborhood of a given matching problem. Although their result does not necessarily focus on large markets, their result can be interpreted as an approximate existence result in large markets.

[20]Studies such as Immorlica and Mahdian (2005), Kojima and Pathak (2009), Kojima, Pathak, and Roth (2013), Hatfield, Kojima, and Narita (2016), and Arnosti (2016) assume that preferences exhibit "limited acceptability" (Lee, 2011), that is, there is a constant k such that each doctor finds only k hospitals acceptable even in large markets. In contrast, our (n, m)-market model assumes "unlimited acceptability," which follows other studies such as Knuth, Motwani, and Pittel (1990), Lee (2011), Ashlagi, Braverman, and Hassidim (2014), and Ashlagi, Kanoria, and Leshno (2016).

Let us comment on a contrast between this result and asymptotic existence results in matching with couples due to Kojima, Pathak, and Roth (2013) and Ashlagi, Braverman, and Hassidim (2014). In those papers, asymptotic existence is established by showing that the rejection chains (as in Kojima and Pathak (2009)) are likely to be short in large markets. By contrast, the nonexistence of a strongly stable matching appears to be unrelated to arguments based on rejection chains. In fact, Proposition 4 shows that the non-existence probability does not converge to zero.

Proposition 4 demonstrates that enlarging the market size does not necessarily solve the existence problem. More specifically, the probability of the existence of a strongly stable matching does not necessarily converge to one even in a large market limit. This is in a sharp contrast to asymptotic existence results such as Kojima, Pathak, and Roth (2013) and then Ashlagi, Braverman, and Hassidim (2014).

On the other hand, Proposition 4 also shows that the probability of the existence of a strongly stable matching does not necessarily diminish to zero either. In that sense, this result also establishes that the market size does not necessarily intensify the non-existence problem of strong stability.

Remark 2. When studying an asymptotic behavior of a market, one important modeling question is how to model large markets. For example, one could keep the number of hospitals fixed while increasing the size of each hospital, in a manner analogous to earlier papers such as Abdulkadiroğlu, Che, and Yasuda (2015), Che and Kojima (2010), and Azevedo and Leshno (2015). Our modeling approach is to increase the number of hospitals (as well as doctors). This approach is close to earlier studies in the literature such as Roth and Peranson (1999), Immorlica and Mahdian (2005), Kojima and Pathak (2009), Kojima, Pathak, and Roth (2013), Ashlagi, Braverman, and Hassidim (2014), Lee (2011), Hatfield, Kojima, and Narita (2016), and Ashlagi, Kanoria, and Leshno (2016).

Another modeling issue is the treatment of regions. For instance, it is possible to increase the number of regions as the market grows while keeping the number of hospitals in each region fixed. It is even possible to have both the number of regions *and* the number of hospitals in each region increase. In contrast to these possibilities, our setting keeps the number of regions fixed while increasing the size of a region by increasing the number of hospitals in a region.

Of course, what kind of modeling approach is appropriate cannot be decided purely from a theoretical viewpoint, and instead it depends on the intended application. As we are not restricting ourselves to one particular application in this study and instead

focusing on conceptual questions about stability under constraints, such an inquiry is beyond the scope of the present paper.

Remark 3. In the (n, m)-market model of this section, we impose no restriction between the number of doctors n and the number of hospitals m except that the latter goes to infinity as the former goes to infinity. One simple example satisfying this condition is the "balanced" case, i.e., the case where $m = n$, a standard assumption employed by Knuth, Motwani, and Pittel (1990), Roth and Peranson (1999), and Lee (2011), among others. The "unbalanced" case, i.e., $m = n$, may be of interest, however. In a recent study, Ashlagi, Kanoria, and Leshno (2016) study unbalanced two-sided matching markets and show that the behavior of stable matching mechanisms is very different between balanced and unbalanced cases. Our setting allows both balanced and unbalanced cases, suggesting that the existence problem of strongly stable matchings identified in this section does not necessarily depend on whether the market is balanced or not.

6. Conclusion

Matching under constraints is a largely unexplored area of research. In this paper, we addressed foundational issues: What should it mean for a matching to be "stable" in an environment under constraints? Do stable matchings exist or, more precisely, under what conditions do stable matchings exist? Does stability imply efficiency? What normative properties characterize stability?

To answer these questions, we defined two stability concepts, strong and weak stability, both of which reduce to the standard stability concept of Gale and Shapley (1962) in the environments with no binding feasibility constraints. Strong stability is conceptually appealing, but we demonstrated that its existence is not guaranteed unless extremely strong restrictions are imposed on the nature of the constraints. In a sharp contrast, we established that a weakly stable matching always exists, implies efficiency, and is characterized by standard axioms. Discussions were provided to examine various applications and to evaluate the severity of the nonexistence problem.

Before closing, let us mention that our stability concepts can be characterized in terms of a stability notion used in an earlier study. Consider a "partitional regions" model as in item 1 of Section 5.1, where each region has "regional preferences" (Kamada and Kojima, 2016) over the distribution of doctors within the region, which represent certain policy goals by the social planner. Given regional preferences, a stability concept is defined based on those regional preferences. Appendix B.2 of the present paper establishes that a matching is strongly stable if and only if it is stable for every possible regional preference

profile, while a matching is weakly stable if and only if there exists a regional preference profile for which it is stable. In this sense, in the "partitional regions" setting, strong and weak stability concepts can be regarded as two canonical strengthening and weakening of stability by Kamada and Kojima (2016), respectively.

One of the contributions of this paper is to analyze stability concepts when the constraint is not given by bounds on the numbers of matched agents. Analysis of such an environment is practically relevant as we have discussed, but at the same time it is difficult.[21] We give an algorithm to find a weakly stable matching in Appendix B.3, but that algorithm violates strategy-proofness. We have not studied this issue further because our focus is on stability concepts per se rather than mechanisms, but a further analysis on incentives is an important direction for future research.[22]

There still remain many other open questions. One interesting but challenging task would be to develop a general theory when there are floor constraints. With floor constraints, even the existence of an individually rational matching is not guaranteed. Meanwhile, when individual rationality is not a requirement (as is the case in the allocation of soldiers in the military (Sönmez, 2013; Sönmez and Switzer, 2013), for instance), Ehlers, Hafalir, Yenmez, and Yildirim (2014) and Fragiadakis and Troyan (2016) have obtained positive results in the presence of floor constraints. Meanwhile, existing research does not seem to have reached a consensus about the "right" stability concept with floor constraints. This question is beyond the scope of the present paper, and it would be an interesting direction for future research. Another direction would be to consider a restriction that takes into account who is matched, not just how many are matched. Eventually, even a general theory of stable matching under a broad class of potentially complicated constraints might be possible. The analysis of this paper is a step toward these ambitious goals.

[21]Existing works on matching under constraints such as Kamada and Kojima (2015, 2016) and Goto, Iwasaki, Kawasaki, Yasuda, and Yokoo (2014) utilize the existing theory of matching with contracts (Hatfield and Milgrom, 2005) to construct a stable mechanism that is strategy-proof for doctors, but such an approach does not appear promising here. This is because an environment studied in Section **??** may result in violations of key properties in matching with contracts, such as substitutability and the law of aggregate demand (e.g., a decrease of one seat at a hospital can result in two additional seats available at another hospital).

[22]However, some structural properties of weak stability seem to suggest strategy-proofness may be hard to obtain. Namely, we can show that there is no doctor optimal weakly stable matching, and two weakly stable matchings can have different numbers of matched doctors (so in particular the rural hospital theorem fails). The same environments as in Examples 4 and 5 of Kamada and Kojima (2015) can be used to show these results.

We regard this study as a building block for further studies of matching under constraints. We argued that each of our stability concepts has normative appeal and investigated their properties. We established that our stability concepts have various normatively appealing properties. What kind of normative criterion is the most relevant in practice, however, is difficult to predict by introspection alone. For problems without distributional constraints, the standard stability notion has stood the test of time through not only theoretical scrutiny, but also laboratory experiments (Kagel and Roth, 2000) and case studies (Roth, 1984; Roth and Peranson, 1999). We hope that more studies in theoretical as well as empirical and experimental market design will analyze matching under distributional constraints. We envision that such studies will further our understanding of suitable stability concepts and, more generally, normative concepts and practically important issues in matching under constraints, informing researchers and practitioners alike.

References

Abdulkadiroğlu, A. (2005): "College admissions with affirmative action," *International Journal of Game Theory*, 33(4), 535–549.

Abdulkadiroğlu, A., Y.-K. Che, and Y. Yasuda (2015): "Expanding "Choice" in School Choice," *American Economic Journal: Microeconomics*, 7(1), 1–42.

Abdulkadiroğlu, A., and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Abraham, D. J., R. Irving, and D. Manlove (2007): "Two algorithms for the student-project allocation problem," *Journal of Discrete Algorithms*, 5, 73–90.

Arnosti, N. (2016): "Centralized Clearinghouse Design: A Quantity-Quality Tradeoff," *mimeo*.

Ashlagi, I., M. Braverman, and A. Hassidim (2014): "Stability in large matching markets with complementarities," *Operations Research*, 62(4), 713–732.

Ashlagi, I., Y. Kanoria, and J. D. Leshno (2016): "Unbalanced random matching markets: The stark effect of competition," *Journal of Political Economy, forthcoming*.

Azevedo, E. M., and J. W. Hatfield (2012): "Complementarity and multidimensional heterogeneity in matching markets," *Unpublished mimeo*.

Azevedo, E. M., and J. D. Leshno (2015): "A supply and demand framework for two-sided matching markets," *Journal of Political Economy, forthcoming*.

Balinski, M., and T. Sönmez (1999): "A tale of two mechanisms: student placement," *Journal of Economic theory*, 84(1), 73–94.

Biro, P., T. Fleiner, R. Irving, and D. Manlove (2010): "The College Admissions Problem with Lower and Common Quotas," *Theoretical Computer Science*, 411(34-36), 3136–3153.

Biró, P., T. Fleiner, R. W. Irving, and D. F. Manlove (2010): "The College Admissions problem with lower and common quotas," *Theoretical Computer Science*, 411, 3136–3153.

Blum, Y., A. E. Roth, and U. Rothblum (1997): "Vacancy Chains and Equilibration in Senior-Level Labor Markets," *Journal of Economic Theory*, 76, 362–411.

Budish, E., Y.-K. Che, F. Kojima, and P. R. Milgrom (2013): "Designing Random Allocation Mechanisms: Theory and Applications," *American Economic Review*, 103(2), 585–623.

Che, Y.-K., J. Kim, and F. Kojima (2013): "Stable matching in large economies," Discussion paper, mimeo.

Che, Y.-K., and F. Kojima (2010): "Asymptotic Equivalence of Probabilistic Serial and Random Priority Mechanisms," *Econometrica*, 78(5), 1625–1672.

Ehlers, L., I. E. Hafalir, M. B. Yenmez, and M. A. Yildirim (2014): "School Choice with Controlled Choice Constraints: Hard Bounds versus Soft Bounds," *Journal of Economic Theory*, 153, 648–683.

Ergin, H. (2002): "Efficient Resource Allocation on the Basis of Priorities," *Econometrica*, 70, 2489–2498.

Ergin, H., and T. Sönmez (2006): "Games of School Choice under the Boston Mechanism," *Journal of Public Economics*, 90, 215–237.

Fragiadakis, D., and P. Troyan (2016): "Improving Matching under Hard Distributional Constraints," forthcoming, *Theoretical Economics*.

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

Goto, M., A. Iwasaki, Y. Kawasaki, Y. Yasuda, and M. Yokoo (2014): "Improving Fairness and Efficiency in Matching Markets with Regional Caps: Priority-list based Deferred Acceptance Mechanism," mimeo (the latest version is available at http://mpra.ub.uni-muenchen.de/53409/).

Goto, M., F. Kojima, R. Kurata, A. Tamura, and M. Yokoo (2016): "Designing Matching Mechanisms under General Distributional Constraints," *American Economic Journal: Microeconomics, forthcoming*.

Hafalir, I. E., M. B. Yenmez, and M. A. Yildirim (2011): "Effective affirmative action in school choice," mimeo.

——— (2013): "Effective affirmative action in school choice," *Theoretical Economics*, 8(2), 325–363.

——— (2016): Personal Communication.

Hatfield, J. W., F. Kojima, and Y. Narita (2016): "Imiproving schools through school choice: A market design approach," *forthcoming, Journal of Economic Theory*.

Hatfield, J. W., and P. Milgrom (2005): "Matching with Contracts," *American Economic Review*, 95, 913–935.

Immorlica, N., and M. Mahdian (2005): "Marriage, Honesty, and Stability," *SODA*, pp. 53–62.

Irving, R. (2011): "Matching Practices for Entry-labor Markets – Scotland," http://www.matching-in-practice.eu/the-scottish-foundation-allocation-scheme-sfas/, last accessed on July 7th, 2016.

Kagel, J. H., and A. E. Roth (2000): "The dynamics of reorganization in matching markets: A laboratory experiment motivated by a natural experiment," *The Quarterly Journal of Economics*, 115(1), 201–235.

Kamada, Y., and F. Kojima (2015): "Efficient Matching under Distributional Constraints: Theory and Applications," *American Economic Review*, (1), 67–99.

——— (2016): "Stability and Strategy-Proofness for Matching with Constraints: A Necessary and Sufficient Condition," mimeo.

Knuth, D. E., R. Motwani, and B. Pittel (1990): "Stable Husbands," *Random Structures and Algorithms*, 1, 1–14.

Kojima, F., and P. Pathak (2009): "Incentives and stability in large two-sided matching markets," *American Economic Review*, 99(3), 608–627.

Kojima, F., P. A. Pathak, and A. E. Roth (2013): "Matching with Couples: Stability and Incentives in Large Markets," *The Quarterly Journal of Economics*, 128(4), 1585–1632.

Lee, S. (2011): "Incentive compatibility of large centralized matching markets," *mimeo*.

Nguyen, T., and R. Vohra (2016): "Near feasible stable matchings with couples," mimeo.

Pathak, P., and T. Sonmez (2008): "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism," *The American Economic Review*, 98(4), 1636–1652.

Roth, A. E. (1982): "The Economics of Matching: Stability and Incentives," *Mathematics of Operations Research*, 7, 617–628.

——— (1984): "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy*, 92, 991–1016.

——— (1985): "The College Admission Problem is not Equivalent to the Marriage Problem," *Journal of Economic Theory*, 36, 277–288.

——— (1991): "A Natural Experiment in the Organization of Entry Level Labor Markets: Regional Markets for New Physicians and Surgeons in the U.K.," *American Economic Review*, 81, 415–440.

Roth, A. E., and E. Peranson (1999): "The Redesign of the Matching Market for American Physicians: Some Engineering Aspects of Economic Design," *American Economic Review*, 89, 748–780.

Sönmez, T. (2013): "Bidding for army career specialties: Improving the ROTC branching mechanism," *Journal of Political Economy*, 121(1), 186–219.

Sönmez, T., and T. B. Switzer (2013): "Matching With (Branch-of-Choice) Contracts at the United States Military Academy," *Econometrica*, 81(2), 451–488.

Sotomayor, M. A. O. (1996): "A Non-constructive Elementary Proof of the Existence of Stable Marriages," *Games and Economic Behavior*, 13, 135–137.

Westkamp, A. (2013): "An Analysis of the German University Admissions System," *Economic Theory*, 53(3), 561–589.

Appendix A. Proofs

A.1. **Proof of Theorem 1.**

We begin by offering an equivalent formulation of independence across hospitals.

Claim 1. *H and f satisfy independence across hospitals if and only if, for every pair of distinct hospitals h_1 and h_2, and every w, $f(w+e_{h_1})=f(w+e_{h_2})=1$ imply $f(w+e_{h_1}+e_{h_2})=1$.*

Proof. The "only if" direction is obvious, so let us show the "if" direction in the following. Define $\bar{q}_h := \sup\{t \in \mathbb{Z}_+ | f(te_h)=1\}$. By the definition of $\bar{q}_h$ and the assumption that f is a feasibility constraint, it is clear that $f(w)=0$ unless $w_h \leq \bar{q}_h$ for all $h \in H$.[23] Therefore, $\{w \in \mathbb{Z}_+^{|H|} | f(w)=1\} \subseteq \{w \in \mathbb{Z}_+^{|H|} | w_h \leq \bar{q}_h, \forall h \in H\}$.

Thus it remains to show $\{w \in \mathbb{Z}_+^{|H|} | f(w)=1\} \supseteq \{w \in \mathbb{Z}_+^{|H|} | w_h \leq \bar{q}_h, \forall h \in H\}$. That is, we show that every $(q'_h)_{h\in H} =: q' \in \mathbb{Z}_+^{|H|}$ such that $q' \leq (\bar{q}_h)_{h\in H}$ satisfies $f(q')=1$. To show this, assume for contradiction that there exists $q' \in \mathbb{Z}_+^{|H|}$ with $q' \leq \bar{q}$ such that $f(q')=0$. Since $f(0)=1$, there exists $w \leq q'$ such that $f(w)=0$ and $f(w-e_h)=1$ for each hospital h with $w_h>0$. This is because if there does not exist such w then for any $w \leq q'$ such that $f(w)=0$, we can find h such that $f(w-e_h)=0$. But this means that $f(0)=0$, a contradiction. Now suppose that the only w's with the aforementioned property are the ones that can be expressed by $w=te_h$ for some hospital h and integer $t \in \{0,1,\ldots,q'_h\}$. But this is a contradiction because $f(w)=0$ by assumption, while $f(w)=1$ must hold because (i) $f(q'_h e_h)=1$ by the definitions of q'_h and $\bar{q}_h$ and the assumption that f is a feasibility constraint, (ii) $w \leq q'_h e_h$, and (iii) again f is a feasibility constraint. If there is a w with the aforementioned property such that there are at least two hospitals h and h' with $w_h, w_{h'}>0$, then we have $f(w)=0$, $f(w-e_h)=1$ and $f(w-e_{h'})=1$, a contradiction. □

Proof of "(2) → (1)": We prove this claim by contraposition. Fix a market with $H=\{h_1,h_2,h_3,\ldots,h_{|H|}\}$ and f such that there exist h_1,h_2 and w such that $f(w+e_{h_1})=f(w+e_{h_2})=1, f(w+e_{h_1}+e_{h_2})=0$. Consider a market in which $D=\{d_1,d_2\} \cup$

[23] To see this, assume $w_h > \bar{q}_h$ for some h. Then by the definition of $\bar{q}_h$, $f(w_h e_h)=0$. Because $w_h e_h \leq w$, this and the assumption that f is a feasibility constraint imply that $f(w)=0$.

$(\bigcup_{h\in H}\{d_1^h, d_2^h, \dots, d_{w_h}^h\})$ such that $d_j^h \neq d_k^{h'}$ if $j \neq k$ or $h \neq h'$, with the following preferences:

$$\begin{aligned}
\succ_{d_i^h} &: h, \quad \text{for } h \in H, i = 1, \dots, w_h \\
\succ_{d_1} &: h_3, h_4, \dots, h_{|H|}, h_2, h_1 \\
\succ_{d_2} &: h_3, h_4, \dots, h_{|H|}, h_1, h_2 \\
\succ_{h_1} &: d_1^{h_1}, d_2^{h_1}, \dots, d_{w_{h_1}}^{h_1}, d_1, d_2; \quad q_{h_1} = w_{h_1} + 1, \\
\succ_{h_2} &: d_1^{h_2}, d_2^{h_2}, \dots, d_{w_{h_2}}^{h_2}, d_2, d_1; \quad q_{h_2} = w_{h_2} + 1, \\
\succ_{h} &: d_1^{h}, d_2^{h}, \dots, d_{w_h}^{h}, d_1, d_2; \quad q_h = w_h, \text{ for any } h \neq h_1, h_2.
\end{aligned}$$

In this market, we first show that in any strongly stable matching μ, $d_i^h \in \mu_h$ for all $h \in H, i = 1, \dots, w_h$. To show this, consider the following cases.

- Suppose that a doctor $d \in \{d_1, d_2\}$ is matched to some hospital $h \notin \{h_1, h_2\}$. Then, since $q_h = w_h$, there exists a doctor d_i^h who is unmatched in μ. Then (d_i^h, h) is a blocking pair such that $d_i^h \succ_h d$, violating strong stability.
- Suppose that no doctor $d \in \{d_1, d_2\}$ is matched to any hospital $h \notin \{h_1, h_2\}$. Consider the following cases.
  - Suppose that $d_i^h \notin \mu_h$ for some d_i^h and $h \notin \{h_1, h_2\}$. Then $|\mu_h| < q_h$ and $h \succ_d \mu_d$ for each $d \in \{d_1, d_2\}$ because $\mu_d \in \{h_1, h_2, \emptyset\}$ by assumption of the present case. Furthermore, satisfying block (d, h) results in a feasible matching by the assumption $f(w + e_{h_1}) = f(w + e_{h_2}) = 1$. Therefore μ is not strongly stable.
  - Suppose $d_i^h \in \mu_h$ for all $h \notin \{h_1, h_2\}, i = 1, \dots, w_h$. If $d_i^h \notin \mu_h$ for some d_i^h and $h \in \{h_1, h_2\}$, then either $\{d_1, d_2\} \subseteq \mu_h$ or $|\mu_h| < q_h$. In the former case, (d_i^h, h) is a blocking pair such that $d_i^h \succ_h d_1$, violating strong stability. In the latter case, (d_i^h, h) is a blocking pair, and satisfying block (d_i^h, h) results in a feasible matching by the assumption $f(w + e_{h_1}) = f(w + e_{h_2}) = 1$. Therefore μ is not strongly stable.

Therefore we have established that in any strongly stable matching μ, $d_i^h \in \mu_h$ for all $h \in H, i = 1, \dots, w_h$.

Under this restriction, we consider the following exhaustive cases:

- If $\mu_{d_1} = \mu_{d_2} = \emptyset$, then μ is not strongly stable because (d_1, h_1) is a blocking pair and the matching satisfying this blocking pair is feasible because $f(w + e_{h_1}) = 1$, so μ is not strongly stable.

- If $\mu_{d_1} = h_1$ and $\mu_{d_2} = \emptyset$, then (d_1, h_2) is a blocking pair and the matching satisfying this blocking pair is feasible, so μ is not strongly stable. The case in which $\mu_{d_2} = h_2$ is symmetric.
- If $\mu_{d_1} = h_2$ and $\mu_{d_2} = \emptyset$, then (d_2, h_2) is a blocking pair such that $d_2 \succ_{h_2} d_1$, so μ is not strongly stable. The case in which $\mu_{d_2} = h_1$ is symmetric.
- If $\mu_{d_1} \neq \emptyset \neq \mu_{d_2}$, then μ violates either feasibility or individual rationality (the capacity of at least one hospital is violated), so it is not strongly stable.

Proof of "(1) $\rightarrow$ (2)": Suppose that there exist no h_1, h_2, and w such that $f(w+e_{h_1}) = f(w+e_{h_2}) = 1, f(w+e_{h_1}+e_{h_2}) = 0$. Order hospitals arbitrarily as $H = \{h_1, h_2, \ldots, h_H\}$. Now, for each $h \in H$, define $\bar{q}_h$ by $\bar{q}_h := \sup\{t \in \mathbb{Z}_+ | f(te_h) = 1\}$.

Consider a stable matching (in the standard matching problem without a feasibility constraint) where the capacity for each hospital h is $\min\{q_h, \bar{q}_h\} < \infty$. Such a stable matching exists because h's preferences are responsive with capacity $\min\{q_h, \bar{q}_h\}$, too. By Claim 1, the definition of strong stability in the new market is identical to the definition of stability (in the standard market without a feasibility constraint) with the capacity for each hospital h being $\min\{q_h, \bar{q}_h\}$. Thus, there exists a strongly stable matching in the new market. Hence there exits a strongly stable matching in the original market.

Proof of "(1) $\rightarrow$ (3)" The proof is straightforward because if (1) holds, then the doctor-proposing deferred acceptance algorithm is well-defined, it is strategy-proof for doctors, and it finds a strongly stable matching for any D and $\succ$.

Proof of "(3) $\rightarrow$ (1)" To prove this claim by contraposition, assume that H and f violate independence across hospitals. Let $H = \{h_1, h_2, h_3, \ldots, h_{|H|}\}$ and fix h_1, h_2 and w such that $f(w+e_{h_1}) = f(w+e_{h_2}) = 1, f(w+e_{h_1}+e_{h_2}) = 0$ (note that such h_1, h_2, and w exist by Claim 1). Consider a market in which $D = \{d_1, d_2\} \cup (\bigcup_{h \in H} \{d_1^h, d_2^h, \ldots, d_{w_h}^h\})$

with the following preferences:

$$\begin{aligned}
d_i^h &: h, \quad \text{for } h \in H, i = 1, \dots, w_h \\
d_1 &: h_3, h_4, \dots, h_{|H|}, h_2, \\
d_2 &: h_3, h_4, \dots, h_{|H|}, h_1, \\
h_1 &: d_1^{h_1}, d_2^{h_1}, \dots, d_{w_{h_1}}^{h_1}, d_1, d_2; \quad q_{h_1} = w_{h_1} + 1, \\
h_2 &: d_1^{h_2}, d_2^{h_2}, \dots, d_{w_{h_2}}^{h_2}, d_2, d_1; \quad q_{h_2} = w_{h_2} + 1, \\
h &: d_1^h, d_2^h, \dots, d_{w_h}^h, d_1, d_2; \quad q_h = w_h, \text{ for any } h \neq h_1, h_2.
\end{aligned}$$

In this market, we first show that in any strongly stable matching μ, $d_i^h \in \mu_h$ for all $h \in H, i = 1, \dots, w_h$. To show this, consider the following cases.

- Suppose that a doctor $d \in \{d_1, d_2\}$ is matched to some hospital $h \notin \{h_1, h_2\}$. Then, since $q_h = w_h$, there exists a doctor d_i^h who is unmatched in μ. Then (d_i^h, h) is a blocking pair such that $d_i^h \succ_h d$, violating strong stability.
- Suppose that no doctor $d \in \{d_1, d_2\}$ is matched to any hospital $h \notin \{h_1, h_2\}$. Consider the following cases.
  - Suppose that $d_i^h \notin \mu_h$ for some d_i^h and $h \notin \{h_1, h_2\}$. Then $|\mu_h| < q_h$ and $h \succ_d \mu_d$ for each $d \in \{d_1, d_2\}$ because $\mu_d \in \{h_1, h_2, \emptyset\}$ by assumption of the present case. Furthermore, satisfying block (d, h) results in a feasible matching by the assumption $f(w + e_{h_1}) = f(w + e_{h_2}) = 1$. Therefore μ is not strongly stable.
  - Suppose $d_i^h \in \mu_h$ for all $h \notin \{h_1, h_2\}, i = 1, \dots, w_h$. If $d_i^h \notin \mu_h$ for some d_i^h and $h \in \{h_1, h_2\}$, then $|\mu_h| < q_h$. Then (d_i^h, h) is a blocking pair, and satisfying block (d_i^h, h) results in a feasible matching by the assumption $f(w + e_{h_1}) = f(w + e_{h_2}) = 1$. Therefore μ is not strongly stable.

Therefore we have established that in any strongly stable matching μ, $d_i^h \in \mu_h$ for all $h \in H, i = 1, \dots, w_h$.

Under this restriction, we consider the following exhaustive cases:

- If $\mu_{d_1} = \mu_{d_2} = \emptyset$, then μ is not strongly stable because (d_2, h_1) is a blocking pair and the matching satisfying this blocking pair is feasible because $f(w + e_{h_1}) = 1$, so μ is not strongly stable.
- If $\mu_{d_1} \neq \emptyset \neq \mu_{d_2}$, then μ violates either feasibility or individual rationality (the capacity of at least one hospital is violated), so it is not strongly stable.

- Thus there are only two strongly stable matchings μ and μ', where

$$\mu_d = \begin{cases} \emptyset & d = d_1 \\ h_1 & d = d_2 \\ h & d = d_i^h \end{cases}, \qquad \mu'_d = \begin{cases} h_2 & d = d_1 \\ \emptyset & d = d_2 \\ h & d = d_i^h \end{cases}.$$

Now, suppose that a mechanism chooses μ under the above preference profile $\succ$. Then d_1 is unmatched. Consider reported preferences $\succ'_{d_1}$ of d_1,

$$\succ'_{d_1}: h_3, h_4, \ldots, h_{|H|}, h_2, h_1.$$

Then it can be verified, by an argument analogous to the above, that μ' is a unique strongly stable matching, so the mechanism chooses μ' at $(\succ'_{d_1}, \succ_{-d_1})$. Doctor d_1 is better off at μ' than at μ since she is matched to h_2 at μ' while she is unmatched at μ. Hence, d_1 can profitably misreport her preferences when her true preferences are $\succ_{d_1}$.

If a mechanism chooses μ' under the above preference profile $\succ$, then by a symmetric argument, doctor d_2 can profitably misreport her preferences when her true preferences are $\succ_{d_2}$. Therefore there does not exist a mechanism that is strategy-proof for doctors and selects a strongly stable matching whenever one exists.

A.2. **Proof of Theorem 2.**

Proof of Existence:

The proof is a generalization of a proof due to Sotomayor (1996) who focuses on the standard one-to-one matching problem without a feasibility constraint.

Define a matching μ to be hospital-quasi-stable[24] if it is feasible and individually rational, and every blocking pair (d, h) of this matching is such that $d' \succ_h d$ for all doctors $d' \in \mu_h$. The set of hospital-quasi-stable matchings is nonempty because a matching with $\mu_i = \emptyset$ for all $i \in D \cup H$ is hospital-quasi-stable.

Now, take a doctor-efficient matching μ in the set of hospital-quasi-stable matchings, i.e., a matching μ that is hospital-quasi-stable such that there is no other hospital-quasi-stable matching $\tilde{\mu}$ with $\tilde{\mu} \succeq_d \mu$ for all d.[25] We will show that μ is weakly stable. Suppose that it is not. Then, by feasibility and individual rationality of μ, we only need to show that if (d, h) is a blocking pair then (i) $f(w(\mu) + e_h) = 0$ and (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$. By the definition of hospital-quasi-stable matching, if (d, h) is a blocking pair then condition (ii) is satisfied. We will show that condition (i) is satisfied as well.

[24]The concept of hospital-quasi-stability is a generalization of the concept that was originally defined for the one-to-one matching without a feasibility constraint. The term "hospital-quasi-stable matching" is due to Blum, Roth, and Rothblum (1997). Sotomayor (1996) uses the term "simple matching."

[25]Such a matching exists because the set of hospital-quasi-stable matchings is finite.

If there is no blocking pair, then this condition is vacuously satisfied, so μ is weakly stable. So suppose that the set of blocking pairs is nonempty. For each h, consider the set of doctors B_h such that $d' \in B_h$ if and only if (d', h) is a blocking pair. There exists h such that we can take $d \in B_h$ with $d \succ_h d'$ for all $d' \in B_h \setminus \{d\}$ (note that such h and d exist by the assumption that the set of blocking pairs is nonempty). Now consider a matching μ' such that $\mu'_d = h$ and $\mu'_{d'} = \mu_{d'}$ for all $d' \in D \setminus \{d\}$. We first show that (1) μ' is hospital-quasi-stable if condition (i) in the definition of weak stability is violated, and then show that (2) μ' is a Pareto improvement over μ for doctors, contradicting our starting assumption that μ is doctor-efficient.

(1) Since (d, h) is a blocking pair, μ' is individually rational.[26] If condition (i) is violated, then $f(w(\mu)+e_h) = 1$ and hence $f(w(\mu')) = 1$ because $w(\mu') \leq w(\mu)+e_h$ by the definition of μ', and thus μ' is feasible. Suppose that under μ', there exists a blocking pair (d^*, h^*) such that $d^* \succ_{h^*} d'$ for some doctor $d' \in \mu'_{h^*}$. Given the definition of μ', it must be either (a) $d^* = d$, (b) $h^* = h$, or (c) $h^* = \mu_d$.

(a) In this case, it holds that $d \succ_{h^*} d'$ for some doctor $d' \in \mu'_{h^*}$. But this means that under μ, (d, h^*) can form a blocking pair such that $d \succ_{h^*} d'$ for some doctor $d' \in \mu_{h^*}$ because $\mu_{\tilde{h}} = \mu'_{\tilde{h}}$ for all $\tilde{h} \in H \setminus \{h, \mu_d\}$ and $h^* \neq h, \mu_d$ by the definitions of μ' and blocking pair. This contradicts the assumption that μ is hospital-quasi-stable.

(b) In this case, it holds that $d^* \succ_h d'$ for some doctor $d' \in \mu'_h$. But this means that under μ, (d^*, h) can form a blocking pair such that $d^* \succ_h d'$ for some doctor $d' \in \mu_h$ because $d \succ_h d''$ for all $d'' \in B_h \setminus \{d\}$. But this contradicts the assumption that μ is hospital-quasi-stable.

(c) In this case, it holds that $d^* \succ_{\mu_d} d'$ for some doctor $d' \in \mu'_{\mu_d}$. But this means that under μ, (d^*, μ_d) can form a blocking pair such that $d^* \succ_{\mu_d} d'$ for some doctor $d' \in \mu_{\mu_d}$ because $\mu'_{\mu_d} \subset \mu_{\mu_d}$. This contradicts the assumption that μ is hospital-quasi-stable.

(2) Since $\mu'_d \succ_d \mu_d$ and $\mu'_{d'} \succeq_{d'} \mu_{d'}$ for all d', μ_d is not doctor-efficient, contradicting our starting assumption that it is doctor-efficient.

Proof of Efficiency:

Let μ be a weakly stable matching and assume, for contradiction, that μ is not efficient. Then there exists a feasible matching μ' that Pareto dominates μ, that is, there is a feasible matching μ' such that $\mu'_i \succeq_i \mu_i$ for all $i \in D \cup H$, with at least one being strict. Noting

[26]Note that responsiveness implies that μ' is individually rational for the hospital μ_d.

that matching is bilateral and that preferences are strict, this implies that there exists a doctor $d \in D$ with $\mu'_d \succ_d \mu_d$. Since μ is a weakly stable matching, $\mu_d \succeq_d \emptyset$ and hence $\mu'_d \neq \emptyset$, so $\mu'_d \in H$. Denote $h = \mu'_d$. Since μ is a weakly stable matching, $h \succ_d \mu_d$ implies one of the following (cases (1) and (2) correspond to a situation in which (d, h) is not a blocking pair of μ. Case (3) covers, by the definition of weak stability, the case in which (d, h) blocks μ):

(1) $\emptyset \succ_h d$.
(2) $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$.
(3) $f(w(\mu) + e_h) = 0$ and $d' \succ_h d$ for all $d' \in \mu_h$.

Suppose $\emptyset \succ_h d$. Then, if $|\mu_h| = q_h$, then there is a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h d'$ for some $d' \in \mu_h$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is weakly stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ. If $|\mu_h| < q_h$, then there should be a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h \emptyset$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is weakly stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ.

Suppose $|\mu_h| = q_h$ and $d' \succ_h d$ for all $d' \in \mu_h$. Then there should be a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h d'$ for some $d' \in \mu_h$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is weakly stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ.

Suppose $f(w(\mu) + e_h) = 0$ and $d' \succ_h d$ for all $d' \in \mu_h$. Then, if $|\mu'_h| \leq |\mu_h|$, then there should be a doctor $d'' \in \mu'_h \setminus \mu_h$ such that $d'' \succ_h d'$ for some $d' \in \mu_h$ (otherwise, by responsiveness of the preference of h, it follows that $\mu_h \succ_h \mu'_h$). Then, since μ is weakly stable, $\mu_{d''} \succ_{d''} h = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ. If $|\mu'_h| > |\mu_h|$, then since $f(w(\mu) + e_h) = 0$, there exists a hospital $h' \in H$ with $|\mu'_{h'}| < |\mu_{h'}|$. This, since $\mu'_{h'} \succ_{h'} \mu_{h'}$ as μ' Pareto dominates μ, implies that there should be a doctor $d'' \in \mu'_{h'} \setminus \mu_{h'}$ such that $d'' \succ_{h'} d'$ for some $d' \in \mu_{h'}$ (otherwise, by responsiveness of the preference of h', it follows that $\mu_{h'} \succ_{h'} \mu'_{h'}$). Then, since μ is weakly stable, $\mu_{d''} \succ_{d''} h' = \mu'_{d''}$, contradicting the assumption that μ' Pareto dominates μ.

A.3. **Proof of Proposition 1.**

The "only if" part: Fix a matching μ that is feasible, non-wasteful and individually rational, and satisfies the no-justified-envy property. We will show that μ is weakly stable. We only need to show that any blocking pair (d, h) satisfies the conditions (i) and (ii) in the definition of weak stability. We show this by contradiction.

So suppose that (d, h) is a blocking pair. That is, $h \succ_d \mu_d$ and either (a) $|\mu_h| < q_h$ and $d \succ_h \emptyset$, or (b) $d \succ_h d'$ for some $d' \in \mu_h$.

If μ is not weakly stable, then either (a') $f(w(\mu) + e_h) = 1$ or (b') $d \succ_h d'$ for some doctor $d' \in \mu_h$.[27] If (b) or equivalently (b') hold, then the no-justified-envy property is violated because for d and the doctor d' such that $d \succ_h d'$ for some $d' \in \mu_h$, conditions (i) and (ii) in the definition of no-justified-envy hold ((i) follows because (d, h) is a blocking pair; (ii) directly follows). But if (a) and (a') hold, then the collection of preference relations in these conditions directly implies violation of non-wastefulness.

The "if" part: To show the "if" part, suppose that μ is weakly stable. By definition it implies that μ is feasible and individually rational.

To show that μ is non-wasteful assume, for contradiction, that μ does not satisfy non-wastefulness. Then there exists a doctor-hospital pair (d, h) such that (i) $h \succ_d \mu_d$ and $d \succ_h \emptyset$, and (ii) $|\mu_h| < q_h$ and $f(w(\mu) + e_h) = 1$. Because $h \succ_d \mu_d$, $d \succ_h \emptyset$, and $|\mu_h| < q_h$, the pair (d, h) is a blocking pair. Moreover, h satisfies $f(w(\mu) + e_h) = 1$, implying that μ is not weakly stable.

To show that μ satisfies the no-justified-envy property assume, for contradiction, that μ does not satisfy the condition. Then there exists a pair of doctors $d, d' \in D$ such that (i) $\mu_{d'} \succ_d \mu_d$ and (ii) $d \succ_{\mu_{d'}} d'$ or $\mu_{d'} = \emptyset$. This implies that $(d, \mu_{d'})$ is a blocking pair because $\mu_{d'} \neq \emptyset$ by individual rationality for d under μ, a weakly stable matching. Moreover, condition (ii) of weak stability is violated for this pair, and hence μ is not weakly stable.

A.4. **Proof of Proposition 2.**

We first define a condition called no preference cycle, and then state and prove two lemmas that show that no preference cycle is equivalent to consistency with master lists. We use those results to prove the direction "(1) $\to$ (2)."

Given R and hospital preference profile $\succ_H \equiv (\succ_h)_{h \in H}$, we say that $\succ_H$ **has no preference cycle** in R if there exist no $r \in R$, distinct hospitals $h_1, \ldots, h_n \in r$, and distinct doctors $d_1, \ldots, d_n \in D$ such that $d_{i+1} \succ_{h_i} d_i \succ_{h_i} \emptyset$ for every $i \in \{1, \ldots, n\}$ (with the convention that $d_{n+1} \equiv d_1$, and $h_{n+1} \equiv h_1$).

Lemma 1. *Suppose that $\succ_H$ has no preference cycle. Then, there do not exist a sequence of (not necessarily distinct) doctors $d_1, \ldots, d_n \in D$ and that of (not necessarily distinct) hospitals $h_1, \ldots, h_n \in H$ such that $d_{i+1} \succ_{h_i} d_i \succ_{h_i} \emptyset$ for every $i \in \{1, \ldots, n\}$ where $d_{n+1} \equiv d_1$ and $h_{n+1} \equiv h_1$.*

[27] This follows from the assumption that h has strict preferences.

Proof. We show this lemma by proving its contraposition. Fix a pair of sequences satisfying the property stated in the statement of the lemma, and denote them by $d_1, \dots, d_n \in D$ and $h_1, \dots, h_n \in H$.

For each $i \in \{1, \dots, n\}$, let $d_i^{A(0)} = d_i$ and $h_i^{A(0)} = h_i$. Consider the following algorithm.

Step $A(k)$:

If there do not exist i and j with $i < j$ such that $d_i^{A(k-1)} = d_j^{A(k-1)}$, then define $B(0) := A(k-1)$, and go to Step $B(1)$. Otherwise, take an arbitrary pair of indices i and j with $i < j$ such that $d_i^{A(k-1)} = d_j^{A(k-1)}$.

Construct a new pair of sequences $(d_1^{A(k)}, \dots, d_{n_{A(k)}}^{A(k)}) \in D^{n_{A(k)}}$ and $(h_1^{A(k)}, \dots, h_{n_{A(k)}}^{A(k)}) \in H^{n_{A(k)}}$ with $n_{A(k)} = n_{A(k-1)} - (j-i)$ such that $d_l^{A(k)} = d_l^{A(k-1)}$ and $h_l^{A(k)} = h_l^{A(k-1)}$ for $l = 1, \dots, i$, and $d_l^{A(k)} = d_{l+(j-i)}^{A(k-1)}$ and $h_l^{A(k)} = h_{l+(j-i)}^{A(k-1)}$ for $l = i+1, \dots, n_{A(k-1)} - (j-i)$. Go to Step $A(k+1)$.

Step $B(k)$:

If there do not exist i and j with $i < j$ such that $h_i^{B(k-1)} = h_j^{B(k-1)}$, then, define $C := B(k-1)$, and stop the algorithm. Otherwise, take an arbitrary pair of indices i and j with $i < j$ such that $h_i^{B(k-1)} = h_j^{B(k-1)}$.

(i) Suppose that $d_{j+1} \succ_{h_i} d_i$ holds. Construct a new pair of sequences $(d_1^{B(k)}, \dots, d_{n_{B(k)}}^{B(k)}) \in D^{n_{B(k)}}$ and $(h_1^{B(k)}, \dots, h_{n_{B(k)}}^{B(k)}) \in H^{n_{B(k)}}$ with $n_{B(k)} = n_{B(k-1)} - (j-i)$ such that $d_l^{B(k)} = d_l^{B(k-1)}$ and $h_l^{B(k)} = h_l^{B(k-1)}$ for $l = 1, \dots, i$, and $d_l^{B(k)} = d_{l+(j-i)}^{B(k-1)}$ and $h_l^{B(k)} = h_{l+(j-i)}^{B(k-1)}$ for $l = i+1, \dots, n_{B(k-1)} - (j-i)$.

(ii) Suppose that $d_{j+1} \succ_{h_i} d_i$ does not hold. Construct a new pair of sequences $(d_1^{B(k)}, \dots, d_{n_{B(k)}}^{B(k)}) \in D^{n_{B(k)}}$ and $(h_1^{B(k)}, \dots, h_{n_{B(k)}}^{B(k)}) \in H^{n_{B(k)}}$ with $n_{B(k)} = j - i$ such that $d_l^{B(k)} = d_{i+l}^{B(k-1)}$ and $h_l^{B(k)} = h_{i+l}^{B(k-1)}$ for $l = 1, \dots, (j-i)$.

Go to Step $B(k+1)$.

This algorithm ends in finite steps for the following reasons. First, Step $B(1)$ is reached in finite steps because, for every $k \in \mathbb{N}$, if Step $A(k)$ is reached, then $n_{A(k)} < n_{A(k-1)}$ holds. Once Step $B(1)$ is reached, again $n_{B(k)} < n_{B(k-1)}$ holds by definition for every $k \in \mathbb{N}$, so the algorithm ends in finite steps.

Next, for each $i, j \in \{1, \dots, n_C\}$, $d_i^C \neq d_j^C$ and $h_i^C \neq h_j^C$ hold by the definition of the algorithm.

Finally, it is straightforward that if $d_{i+1}^{A(k-1)} \succ_{h_i^{A(k-1)}} d_i^{A(k-1)} \succ_{h_i^{A(k-1)}} \emptyset$ for every $i \in \{1, \dots, n_{A(k-1)}\}$ such that $d_{n_{A(k-1)}+1}^{A(k-1)} \equiv d_1^{A(k-1)}$ and $h_{n_{A(k-1)}+1}^{A(k-1)} \equiv h_1^{A(k-1)}$, then $d_{i+1}^{A(k)} \succ_{h_i^{A(k)}} d_i^{A(k)} \succ_{h_i^{A(k)}} \emptyset$ for every $i \in \{1, \dots, n_{A(k)}\}$ such that $d_{n_{A(k)}+1}^{A(k)} \equiv d_1^{A(k)}$ and $h_{n_{A(k)}+1}^{A(k)} \equiv h_1^{A(k)}$. Also, it is true that if $d_{i+1}^{B(k-1)} \succ_{h_i^{B(k-1)}} d_i^{B(k-1)} \succ_{h_i^{B(k-1)}} \emptyset$ for every $i \in \{1, \dots, n_{B(k-1)}\}$

such that $d^{B(k-1)}_{n_{B(k-1)}+1} \equiv d^{B(k-1)}_1$ and $h^{B(k-1)}_{n_{B(k-1)}+1} \equiv h^{B(k-1)}_1$, then $d^{B(k)}_{i+1} \succ_{h^{B(k)}_i} d^{B(k)}_i \succ_{h^{B(k)}_i} \emptyset$ for every $i \in \{1, \dots, n_{B(k)}\}$ such that $d^{B(k)}_{n_{B(k)}+1} \equiv d^{B(k)}_1$ and $h^{B(k)}_{n_{B(k)}+1} \equiv h^{B(k)}_1$. This follows because for each $h \in H$, $\succ_h$ is a strict order, so if there exist $d, d'd'', d''' \in D$ such that $d \succ_h d'$ and $d'' \succ_h d'''$, then either $d \succ_h d'''$ or $d'' \succ_h d'$ holds. Hence, for each $k = 1, \dots,$ in (ii) of Step $B(k)$, it must be the case that $d_{i+1} \succ_{h_i} d_j$.

This completes the proof. □

Lemma 2. *For a set of regions R and a hospital preference profile $\succ_H$, the following two claims are equivalent.*

(1) *$\succ_H$ is consistent with master lists of R.*
(2) *$\succ_H$ has no preference cycle in R.*

Proof. **Proof of "(1) → (2)":** Suppose for contradiction that $\succ_H$ is consistent with master lists of R while $\succ_H$ has a preference cycle in R. Then there exists a region $r \in R$, distinct hospitals $h_1, \dots, h_n \in r$, and distinct doctors $d_1, \dots, d_n \in D$ such that $d_{i+1} \succ_{h_i} d_i \succ_{h_i} \emptyset$ for every $i \in \{1, \dots, n\}$. Because $\succ_H$ is consistent with master lists in R, this implies $d_{i+1} \succ_r d_i$ for all $i \in \{1, \dots, n\}$. By transitivity of $\succ_r$, we obtain $d_1 \succ_r d_1$, a contradiction to the assumption that $\succ_r$ is a strict order over D and hence satisfies antisymmetry.

Proof of "(2) → (1)": Assign indices $1, \dots, n$ to all the doctors arbitrarily, $D = \{d_1, \dots, d_n\}$. For each r, construct a binary relation $\succ_r$ through the following procedure: Let $D^0 = D$. For $k = 1, \dots, n$,

> Step k: Let $d^k = \min_{l \in \{1,\dots,n\}} \{d_l \in D^{k-1} |\ \nexists (d, h) \in D^{k-1} \times r$ s.t. $d \succ_h d_l \succ_h \emptyset\}$. Let $D^k = D^{k-1} \setminus \{d^1\}$.

Define $\succ_r$ by, for all $k, l \in \{1, \dots, n\}$, letting $d^k \succ_r d^l$ if and only if $k < l$.

Note that $|D^k| = n - k$ for each $k = 0, \dots, n$ because, for each k, the set $\{d_l \in D^{k-1} |\ / \exists (d, h) \in D^{k-1} \times r$ s.t. $d \succ d_l\}$ is nonempty. To see this, suppose for contradiction that it is empty for some k. Then, for each $d \in D^{k-1}$, there exists $d' \in D^{k-1}$ and $h \in r$ such that $d' \succ_h d$. This implies that there exist a sequence of (not necessarily distinct) doctors $d_1, \dots, d_m \in D$ and that of (not necessarily distinct) hospitals $h_1, \dots, h_m \in H$ such that $d_{i+1} \succ_{h_i} d_i \succ_{h_i} \emptyset$ for every $i \in \{1, \dots, m\}$ where $d_{m+1} \equiv d_1$ and $h_{m+1} \equiv h_1$. Then, by Lemma 1, $\succ_H$ has no preference cycle in R, a contradiction. Completeness, transitivity, and antisymmetry of $\succ_r$ is obvious from the construction of $\succ_r$. It is also clear from construction of $\succ_r$ that for each hospital $h \in r$, $d \succ_h d' \succ_h \emptyset$ implies $d \succ_r d'$. Therefore, $\succ_H$ is consistent with master lists in R, completing the proof. □

Proof of "(1) → (2)" of Proposition 2: Suppose for contradiction that (1) holds while (2) does not. Then Lemma 2 implies that there exist a region r as well as sequences of distinct hospitals $h_1, \ldots, h_n$ and distinct doctors $d_1, \ldots, d_n$ such that $d_{i+1} \succ_{h_i} d_i \succ_{h_i} \emptyset$ for every $i \in \{1, \ldots, n\}$. Now, let $q_r = 1$, $q_{r'} \geq |D|$ for all $r' \in R \setminus \{r\}$, $\succ_{d_i}$: h_i, h_{i-1}, for each $i = \{1, \ldots, n\}$ where $h_0 \equiv h_n$, and every doctor $d \in D \setminus \{d_1, \ldots, d_n\}$ prefers $\emptyset$ the most.

First, note that any strongly stable matching must have exactly one hospital $h \in r$ matched to one doctor while every other $h' \in r \setminus \{h\}$ is unmatched because (1) if no doctor is matched to hospitals in r, then any pair of a hospital in r and a doctor who find each other acceptable (note such a pair exists by the present preference specification) forms a blocking pair and (unless the present matching is itself infeasible) violates condition (i) of Definition 1, and (2) if more than one doctor is matched to hospitals in r, then the matching is infeasible. Therefore, in the remainder of the proof, consider a matching under which there is exactly one hospital $h_i \in r$ that is matched with one doctor while all other hospitals in r are unmatched. Consider the following (exhaustive) cases.

(1) Suppose that h_i is matched to d_i. Then (d_{i+1}, h_i) is a blocking pair because d_{i+1} is currently unmatched by our earlier argument and hence prefers h_i to her present outcome; h_i prefers d_{i+1} to d_i by assumption. Furthermore, blocking pair (d_{i+1}, h_i) violates condition (ii) of Definition 1, which shows that the matching is not strongly stable.

(2) Suppose that h_i is matched to d_{i+1}. Then (d_{i+1}, h_{i+1}) is a blocking pair because h_{i+1} is currently unmatched by our earlier argument and prefers d_{i+1} to $\emptyset$ by assumption; d_{i+1} prefers h_{i+1} to h_i by assumption. Finally, blocking pair (d_{i+1}, h_{i+1}) violates condition (i) of Definition 1 (unless the present matching is itself infeasible), which shows that the matching is not strongly stable.

(3) Suppose that h_i is matched to a doctor $d \in D \setminus \{d_i, d_{i+1}\}$. Then this matching is not individually rational and hence it is not strongly stable.

Proof of "(2) → (1)" of Proposition 2: When $\succ_H$ has no preference cycle in R, by inspection strong stability is implied by the stability concept due to Biro, Fleiner, Irving, and Manlove (2010). Biro, Fleiner, Irving, and Manlove (2010) show the existence of a stable matching in their setting, which implies the existence of a stable matching.

A.5. **Proof of Proposition 3.**

Let ν be a stable matching in the AS model and μ be a matching in the KK model that is associated to ν.

To show that μ is strongly stable in the KK model, note first that feasibility and individual rationality of μ are obvious by feasibility and individual rationality of ν in the AS model (which are implied by stability of ν in the AS model).

Suppose that a student s, whose type is t, blocks matching μ with hospital (c,t) in the KK model. This implies $c \succ_s \nu_s$ in the AS model. Then, because ν is stable in the AS model, it follows that either (i) $|\nu_c| = q_c$ and $s' \succeq_c s$ for all $s' \in \nu_c$ or (ii) $|\nu_c \cap S^t| = q_c^t$ and $s' \succeq_c s$ for all $s' \in \nu_c \cap S^t$. In the former case, $|\mu_r| = q_r$ for $r = \{(c,m),(c,M)\}$, $s' \succeq_{(c,t)} s$ for all $s' \in \mu_{(c,t)}$, and $\mu_s \notin r$. Therefore, μ is strongly stable. In the latter case, $|\mu_{(c,t)}| = q_{(c,t)}$, and $s' \succeq_{(c,t)} s$ for all $s' \in \mu_{(c,t)}$, so $(s,(c,t))$ is not a block. This is a contradiction, completing the proof.

A.6. **Proof of Proposition 4.**

In the (n,m)-market model, we first establish the following characterization of the existence of a strongly stable matching.

Lemma 3. *A matching is strongly stable if and only if exactly one doctor-hospital pair (d,h) is matched in that matching and d prefers h the most and h prefers d the most.*

Proof. The "if" direction is obvious, because if the matching satisfies the above condition, then every blocking pair satisfies conditions (i) and (ii) in Definition 1.

To show the "only if" direction, first note that any strongly stable matching must have exactly one pair of a doctor and a hospital who are matched because (1) if no doctor is matched, then any doctor-hospital pair forms a blocking pair and violates condition (i) of Definition 1, and (2) if more than one doctor is matched, then the matching is infeasible. Therefore, in the remainder of the proof, consider a matching under which there is exactly one doctor-hospital pair (d,h) who are matched to each other. Consider the following two cases.

(1) Suppose that d is not the first choice of h. Let d' be the first choice of h. Then (d',h) is a blocking pair because d' is currently unmatched by assumption and prefers h to $\emptyset$ given the assumption that all hospitals are acceptable to all doctors; h prefers d' to d by assumption. Furthermore, blocking pair (d',h) violates condition (ii) of Definition 1, which shows that the matching is not strongly stable.

(2) Suppose that h is not the first choice of d. Let h' be the first choice of d. Then (d,h') is a blocking pair because h' is currently unmatched by assumption and prefers d to $\emptyset$ given the assumption that all doctors are acceptable to all doctors; d prefers h' to h by assumption. Finally, blocking pair (d,h') violates condition (i) of Definition 1, which shows that the matching is not strongly stable.

□

Recall that m is assumed to depend on n. Denote by p_n the probability that, in the (n, m)-market model there exists no doctor-hospital pair (d, h) such that d prefers h most and h prefers d most. By Lemma 3, the probability that there exists at least one strongly stable matching is equal to $1 - p_n$. Recall that $m \to \infty$ as $n \to \infty$.

Lemma 4.

$$\lim_{n\to\infty} p_n = \frac{1}{e}.$$

Proof. Assign indices from 1 to n to the doctors (with exactly one index assigned to each doctor and vice versa). Similarly, assign indices from 1 to m to the hospitals. For each index i, let Y_i be (the index of) the hospital that doctor (with the index) i prefers the most. That is, Y_i is a random variable with $\Pr(Y_i = j) = 1/m$ for every i, j that is independently distributed across i. Then, the conditional probability, given a realization of $(Y_i)_{i=1}^n$, that there exists no doctor-hospital pair who prefer each other the most is

$$\left(1 - \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=1\}}}{n}\right) \times \left(1 - \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=2\}}}{n}\right) \times \cdots \times \left(1 - \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=m\}}}{n}\right).$$

Hence, by linearity of the expectation operator, we have

(A.1)

$$\begin{aligned} p_n &= \mathbb{E}\left[\left(1 - \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=1\}}}{n}\right) \times \left(1 - \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=2\}}}{n}\right) \times \cdots \times \left(1 - \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=m\}}}{n}\right)\right] \\ &= \mathbb{E}\left[1 - \sum_{j=1}^m a_j + \sum_{j<j'} a_j a_{j'} - \cdots + (-1)^m a_1 a_2 \ldots a_m\right] \\ &= 1 - \mathbb{E}\left[\sum_{j=1}^m a_j\right] + \mathbb{E}\left[\sum_{j<j'} a_j a_{j'}\right] - \cdots + (-1)^m \mathbb{E}\left[a_1 a_2 \ldots a_m\right], \end{aligned}$$

where

$$a_j = \frac{\sum_{i=1}^n \mathbb{1}_{\{Y_i=j\}}}{n},$$

for each $j \in \{1, \ldots, m\}$. By symmetry with respect to $j \in \{1, \ldots, m\}$, for each $k \in \{1, \ldots, m\}$, the $(k+1)$-st term of Equation (A.1) can be expressed as the sum of $\binom{m}{k}$ terms of the form

$$(-1)^k \frac{1}{n^k} \mathbb{E}\left(\sum_{i=1}^n \mathbb{1}_{\{Y_i=j_1\}} \sum_{i=1}^n \mathbb{1}_{\{Y_i=j_2\}} \cdots \sum_{i=1}^n \mathbb{1}_{\{Y_i=j_k\}}\right),$$

where $j_\ell \neq j_{\ell'}$ for all $\ell \neq \ell'$, all of which have the same value as one another. Note that

$$\begin{aligned}
\mathbb{E}\left(\sum_{i=1}^{n} \mathbb{1}_{\{Y_i=j_1\}} \sum_{i=1}^{n} \mathbb{1}_{\{Y_i=j_2\}} \cdots \sum_{i=1}^{n} \mathbb{1}_{\{Y_i=j_k\}}\right) &= \sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\cdots\sum_{i_k=1}^{n} \mathbb{E}\left(\mathbb{1}_{\{Y_{i_1}=j_1\}}\mathbb{1}_{\{Y_{i_2}=j_2\}}\cdots\mathbb{1}_{\{Y_{i_k}=j_k\}}\right) \\
&= \sum_{\substack{1\le i_\ell\le n,\forall \ell \\ i_\ell\neq i_{\ell'},\forall \ell\neq\ell'}} \mathbb{E}\left(\mathbb{1}_{\{Y_{i_1}=j_1\}}\mathbb{1}_{\{Y_{i_2}=j_2\}}\cdots\mathbb{1}_{\{Y_{i_k}=j_k\}}\right) \\
&= \sum_{\substack{1\le i_\ell\le n,\forall \ell \\ i_\ell\neq i_{\ell'},\forall \ell\neq\ell'}} \mathbb{E}\left(\mathbb{1}_{\{Y_{i_1}=j_1\}}\right)\mathbb{E}\left(\mathbb{1}_{\{Y_{i_2}=j_2\}}\right)\ldots\mathbb{E}\left(\mathbb{1}_{\{Y_{i_k}=j_k\}}\right) \\
&= \frac{n(n-1)\ldots(n-k+1)}{m^k},
\end{aligned}$$

where the first equality follows from linearity of the expectation operator, the second equality follows from the fact that $\mathbb{1}_{\{Y_{i_1}=j_1\}}\mathbb{1}_{\{Y_{i_2}=j_2\}}\cdots\mathbb{1}_{\{Y_{i_k}=j_k\}} = 0$ if there exit ℓ, ℓ' with $\ell \neq \ell'$ such that $i_\ell = i_{\ell'}$ (because, for each i, $\mathbb{1}_{\{Y_i=j\}} = 0$ for all but one j), the third equality follows from independence of random variables $Y_{i_1}, Y_{i_2}, \ldots, Y_{i_k}$, and the last equality follows because $\mathbb{E}(\mathbb{1}_{\{Y_i=j\}}) = 1/m$ for all $i \in \{1, \ldots, n\}$ and $j \in \{1, \ldots, m\}$ and there are $n(n-1)\ldots(n-k+1)$ terms to be added together in the summation. Therefore, the $(k+1)$-st term of Equation (A.1) is equal to

$$\begin{aligned}
&\binom{m}{k}(-1)^k\frac{1}{n^k}\frac{n(n-1)\ldots(n-k+1)}{m^k} \\
&= (-1)^k\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{k-1}{m}\right).
\end{aligned}$$

Therefore,

$$p_n = 1 + \sum_{k=1}^{m}(-1)^k\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{k-1}{m}\right).$$

In the right hand side of this equation, note that the term

$$\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{k-1}{m}\right)$$

lies in $[0, 1]$ and is nonincreasing in k. Thus, for any nonnegative integers m and ℓ with $m \geq 2\ell + 1$, we have

$$\begin{aligned}
&1 + \sum_{k=1}^{2\ell+1}(-1)^k\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{k-1}{m}\right) \le p_n \\
&\le 1 + \sum_{k=1}^{2\ell}(-1)^k\frac{1}{k!}\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)\ldots\left(1-\frac{k-1}{n}\right)\left(1-\frac{1}{m}\right)\left(1-\frac{2}{m}\right)\ldots\left(1-\frac{k-1}{m}\right).
\end{aligned}$$

Thus, taking $n \to \infty$ and recalling the assumption that $m \to \infty$ as $n \to \infty$,

$$\sum_{k=0}^{2\ell+1}(-1)^k\frac{1}{k!} \leq \liminf p_n \leq \limsup p_n \leq \sum_{k=0}^{2\ell}(-1)^k\frac{1}{k!},$$

for any nonnegative integer ℓ. By taking $\ell \to \infty$, we obtain

$$\frac{1}{e} \leq \liminf p_n \leq \limsup p_n \leq \frac{1}{e},$$

and hence, $\lim_{n\to\infty} p_n = \frac{1}{e}$. □

Lemmas 3 and 4 lead to the desired conclusion, completing the proof.

Appendix B. Additional Discussions

B.1. **Independence of the Axioms for Proposition 1.** Proposition 1 characterizes weak stability by non-wastefulness, individual rationality, feasibility, and the no-justified-envy property. This section provides examples to show that these four axioms are independent.

- Non-wastefulness: The empty matching, that is, a matching μ such that $\mu_d = \emptyset$ for every $d \in D$, violates non-wastefulness as long as there are a doctor d and a hospital h such that $h \succ_d \emptyset$, $d \succ_h \emptyset$, and $f(e_h) = 1$. It is straightforward to see that the empty matching satisfies individual rationality, feasibility, and the no-justified-envy property.
- Individual rationality: Let $D = \{d\}, H = \{h\}$, $f(e_h) = 1$, and $\emptyset \succ_d h$ and $\emptyset \succ_h d$. Then the matching μ such that $\mu_d = h$ violates individual rationality. This matching μ satisfies non-wastefulness, feasibility, and the no-justified-envy property.
- Feasibility: Let $D = \{d\}, H = \{h\}$, $R = \{H\}$, $q_H = 0$, and $h \succ_d \emptyset$ and $d \succ_h \emptyset$ and $q_h = 1$. Then the matching μ such that $\mu_d = h$ violates feasibility. This matching μ satisfies non-wastefulness, individual rationality, and the no-justified-envy property.
- No-justified-envy: Let $D = \{d_1, d_2\}, H = \{h_1, h_2\}$, f is such that $f(1,1) = 1$, $\succ_h$: d_1, d_2 and $q_h = 1$ for each $h \in H$, and $\succ_d$: h_1, h_2 for each $d \in D$. Consider matching μ defined by

$$\mu = \begin{pmatrix} h_1 & h_2 \\ d_2 & d_1 \end{pmatrix}.$$

Because $\mu_{d_2} = h_1 \succ_{d_1} h_2 = \mu_{d_1}$ and $d_1 \succ_{h_1} d_2$, matching μ violates the no-justified-envy property. Meanwhile, μ satisfies non-wastefulness, individual rationality, and feasibility.

B.2. **Characterization of Strong and Weak Stability Using Stability in Kamada and Kojima (2016).** Here we relate strong and weak stability with the stability notion developed in Kamada and Kojima (2016). They develop a stability concept that is weaker than strong stability while stronger than weak stability. In doing so, they devise a notion of regional preferences and use them to define stability. The idea is that, under the matching μ, a blocking pair (d, h) involving a move of a doctor d is considered legitimate only if it leads to Pareto improvement among the regions that contain both μ_d and h, in light of their regional preferences. An important feature of this notion is that it depends on the regional preferences at hand. Although this is helpful in explicitly taking into account the policy goal regarding allocation of doctors, it also means that we need additional information to define stability.

In this section we try to relate such a stability notion with strong and weak stability in our paper, which do not use such additional information. We find that strong stability corresponds to stability for *all* regional preferences, while weak stability corresponds to stability for *some* regional preferences.

In what follows, we consider the "partitional regions" setting formalized in item 1 of Section 5.1. To make the paper self-contained, we provide formal definitions of regional preferences and stability.

Let **regional preferences** $\succeq_r$ be a weak ordering over nonnegative-valued integer vectors $W_r := \{w = (w_h)_{h \in r} | w_h \in \mathbb{Z}_+\}$. That is, $\succeq_r$ is a binary relation that is complete and transitive (but not necessarily antisymmetric). We write $w \succ_r w'$ if and only if $w \succeq_r w'$ holds but $w' \succeq_r w$ does not. Vectors such as w and w' are interpreted to be supplies of acceptable doctors to the hospitals in region r, but they only specify how many acceptable doctors apply to each hospital and no information is given as to who these doctors are.

For each $h \in H$, let $r(h)$ denote the region r such that $h \in r$. Given a profile of regional preferences $(\succeq_r)_{r \in R}$, under partitional regional caps, stability defined in Kamada and Kojima (2016) reduces to the following.

Definition 5. A matching μ is **stable** if it is feasible, individually rational, and if (d, h) is a blocking pair then (i) $|\mu_{r(h)}| = q_{r(h)}$, (ii) $d' \succ_h d$ for all doctors $d' \in \mu_h$, and

(iii) either $\mu_d \notin r(h)$ or $w \succeq_{r(h)} w'$,

where $w_{h'} = |\mu_{h'}|$ for all $h' \in r(h)$ and $w'_h = w_h + 1$, $w'_{\mu_d} = w_{\mu_d} - 1$ and $w'_{h'} = w_{h'}$ for all other $h' \in r(h)$.

Under partitional regional caps, strong stability reduces to replacing (i) in Definition 1 with "$|\mu_{r(h)}| = q_{r(h)}$ and $\mu_d \notin r(h)$," and weak stability reduces to replacing (i) in Definition 4 with "$|\mu_{r(h)}| = q_{r(h)}$."

Proposition 5. (1) *μ is strongly stable if and only if μ is stable for all possible regional preferences.*

(2) *μ is weakly stable if and only if there exists a profile of regional preferences $(\succeq_r)_{r\in R}$ under which μ is stable.*

Proof. **Part 1:** By the definition of stability, the "only if" part is obvious. We prove the "if" part. Since the only difference of strong stability and stability is that $\mu_d \notin r(h)$ in the definition of strong stability is replaced with condition (iii) in the definition of stability, it suffices to show that for any matching μ, if there is a blocking pair (d, h) such that $\mu_d \in r(h)$, we can find a profile of regional preferences $(\succeq_r)_{r\in R}$ such that $w' \succ_{r(h)} w$ where $w_{h'} = |\mu_{h'}|$ for all $h' \in r(h)$ and $w'_h = w_h + 1$, $w'_{\mu_d} = w_{\mu_d} - 1$ and $w'_{h'} = w_{h'}$ for all other $h' \in r(h)$.

To this end, fix a blocking pair (d, h) such that $\mu_d \in r(h)$ and define w and w' in the above manner. For each region $r \in R$, consider an ordering of hospitals in it, $(h^r_1, \ldots, h^r_{|r|})$, such that $h^{r(h)}_1 = h$. Define a profile of regional preferences $(\succeq_r)_{r\in R}$ by the rule: For each weight vector $\tilde{w}$ for region r, let $n(\tilde{w}) = \sum_{h'\in r} \tilde{w}_{h'}$ be the number of doctors matched to r under $\tilde{w}$. For any pair of weight vectors $\tilde{w}$ and $\hat{w}$ for region r:

(1) If $n(\tilde{w}) > q_r \geq n(\hat{w})$, then $\hat{w} \succ_r \tilde{w}$. If $n(\tilde{w}), n(\hat{w}) > q_r$, then $\hat{w} \succeq_r \tilde{w}$ if and only if $n(\tilde{w}) \geq n(\hat{w})$.

(2) Suppose $n(\tilde{w}), n(\hat{w}) \leq q_r$. Then, if $\tilde{w}_h > q_h$ for some h while $\hat{w}_{h'} \leq q_{h'}$ for all $h' \in r$ then $\hat{w} \succ \tilde{w}$. If $\tilde{w}_h > q_h$ for some h and $\hat{w}_{h'} > q_{h'}$ for some h' then $\hat{w} \succeq \tilde{w}$ if and only if $n(\tilde{w}) \geq n(\hat{w})$.

(3) If $q_r \geq \max\{n(\tilde{w}), n(\hat{w})\}$ and $q_{h'} \geq \max\{\tilde{w}_{h'}, \hat{w}_{h'}\}$ for all h', and there exists i such that $\tilde{w}_{h^r_j} = \hat{w}_{h^r_j}$ for all $j < i$ and $\tilde{w}_{h^r_i} > \hat{w}_{h^r_i}$, then $\tilde{w} \succ_r \hat{w}$.

With this specification, it is straightforward to see that $w' \succ_{r(h)} w$, which completes the proof.

Part 2: By the definition of stability, the "if" part is obvious. We prove the "only if" part. Consider a profile of regional preferences such that for each region r and for each pair of vectors $\tilde{w}$ and $\hat{w}$ for region r, regional preference $\succeq_r$ is specified by (1) and (2) for

Part 1 and, if $q_r \geq \max\{n(\tilde{w}), n(\hat{w})\}$, $q_{h'} \geq \max\{\tilde{w}_{h'}\hat{w}_{h'}\}$ for all h', and $n(\hat{w}) \geq n(\tilde{w})$, then $\hat{w} \succeq_r \tilde{w}$. In this case, condition (iii) of stability automatically holds. Since the rest of the conditions in the definition of stability is identical to the definition of weak stability, this implies that these two notions are equivalent under the specified regional preferences. □

Remark 4. The "if" part of each of the statements would be stronger if we restrict ourselves to a narrower class of regional preferences. In fact, the results of the proposition hold even if we restrict ourselves to a class of regional preferences that Kamada and Kojima (2016) restricts to. We also note that the regional preferences constructed in the proofs of both Parts 1 and 2 satisfy the conditions assumed in that paper.[28]

B.3. **Construction of a Weakly Stable Matching.** Theorem 2 shows existence of weakly stable matchings, and their proof in Appendix A.2 does not explicitly construct a weakly stable matching. However, we can use the same idea as in that proof to define an algorithm that finds a weakly stable matching for any given preference profile.

To do this, for any matching μ and a preference profile, for each h, define the set of doctors $B_h(\mu)$ to be such that $d' \in B_h(\mu)$ if and only if (d', h) is a blocking pair under μ.

Algorithm.

Step 0: Fix an order of hospitals, $\{h_1, \ldots, h_H\}$. Define μ by $\mu_d = \emptyset$ for all $d \in D$. Go to Step 1.

Step 1: If $f(w(\mu) + e_h) = 0$ for every h with $B_h(\mu) \neq \emptyset$, stop the algorithm and output μ. Otherwise, go to Step 2.

Step 2: Take h_i such that $B_{h_j}(\mu) = \emptyset$ for all $j < i$, $B_{h_i}(\mu) \neq \emptyset$, and $f(w(\mu) + e_{h_i}) = 1$. Take $d \in B_{h_i}(\mu)$ with $d \succ_h d'$ for all $d' \in B_{h_i}(\mu) \setminus \{d\}$, and define a matching μ' such that $\mu'_d = h_i$ and $\mu'_{d'} = \mu_{d'}$ for all $d' \in D \setminus \{d\}$. Let $\mu = \mu'$. Go back to Step 1.

This algorithm ends in finite steps because in Step 2, $\mu'_d \succeq_d \mu_d$ hold for all $d \in D$. The proof that this algorithm produces a weakly stable matching is omitted because it is essentially the replication of the proof for Theorem 2 that we presented in Appendix A.2.

B.4. **A Detailed Analysis of Example 3.** The following observations show that there exists no strongly stable matching in the market presented in Example 3. Suppose for contradiction that there exists a strongly stable matching μ, and consider the following (exhaustive) cases.

[28]Specifically, they satisfy feasibility with respect to both hospital capacities and the regional cap, the acceptance condition, consistency, and substitutability.

(1) Suppose that h_1 is matched to d_2 at μ. Then feasibility of μ implies that h_1 is not matched to any other doctor and h_2 is unmatched. Then, however, (d_2, h_2) is a blocking pair that violates condition (i) of Definition 1, showing that μ is not strongly stable.

(2) Suppose that h_1 is matched to d_1 at μ. Then, feasibility of μ implies h_2 is unmatched. Then h_3 should be matched with d_3 because otherwise (d_3, h_3) is a blocking pair that violates condition (i) of Definition 1. This and feasibility of μ imply that d_2 is unmatched. This implies, however, that (d_2, h_1) is a blocking pair that violates condition (ii) of Definition 1, which shows μ is not strongly stable.

(3) Suppose that h_1 is matched to d_3 at μ. Then, individual rationality of μ implies d_1 is unmatched while feasibility of μ implies h_2 is unmatched. Thus, (d_1, h_1) is a blocking pair which violates condition (ii) of Definition 1, showing that μ is not strongly stable.

(4) Suppose that h_1 is unmatched at μ. Then, d_3 should be matched to h_2 at μ because otherwise (d_3, h_2) is a blocking pair that violates condition (i) of Definition 1. This implies that d_2 must be matched to h_3, because otherwise (d_2, h_3) is a blocking pair that violates condition (i) of Definition 1. This in turn implies that h_2 is matched to only one doctor, namely d_3, while h_1 is unmatched at μ. This implies that (d_3, h_1) is a blocking pair which violates condition (i) of Definition 1, showing that μ is not strongly stable.