Participation in matching markets with distributional constraints

Esteban Peralta*

June 29, 2020

Summary

Many matching markets run clearinghouses that impose quotas to achieve some desirable allocation. Achieving quotas that are not satisfied by any stable allocation requires enforcement ex post, once individuals have been allocated. When participation in the clearinghouse is voluntary and non-participating agents are not bounded by the desired quotas, however, achieving the desired quotas requires, in addition, full participation ex ante. Thus, deterministic clearinghouses can achieve no unstable distributional goal. This paper explores the scope of random clearinghouses as an effective redesign. I show that a random mechanism inducing full participation exists if and only if agents' preferences are sufficiently heterogeneous. The main finding is a polynomial-time algorithm that identifies the class of quotas that *cannot* be achieved by *any* random mechanism.

JEL classification: C62, C78, D47.

Keywords: distributional constraints, matching markets, participation, random assignments.

*Department of Economics and Stephen M. Ross School of Business, University of Michigan, 611 Tappan Ave. (48109), Ann Arbor, MI, USA (eperalta@umich.edu). This paper is a substantially revised version of the first chapter of my dissertation at Yale University. Financial support from Yale and the University of Michigan is greatly appreciated. I am deeply indebted to Larry Samuelson for his time, guidance and support on—and off—path and very grateful to Laura Doval, Simone Galperti, Yuichiro Kamada, and Philipp Strack for detailed comments. I also thank Ian Ball, Marcelo Fernández, Federico Echenique, Fuhito Kojima, Ce Liu, Yusuke Narita, and Ana Reynoso for long discussions about these topics and Dirk Bergemann, Tilman Börgers, Hector Chade, Marina Halac, Ryota Iijima, Navin Kartik, Yuichi Kitamura, Alexey Kushnir, Jacob Leshno, Ce liu, Margaux Luflade, Chiara Margaria, David Miller, Giuseppe Moscarini, David Rahman, Daniel Seidmann, Takuo Sugaya, Fernando Tohmé, and seminar participants at Yale and the University of Michigan for helpful comments and suggestions. Any errors are my own.

1 Introduction

The analysis of matching markets, both applied and theoretical, has largely been guided by the study of stable allocations; namely, allocations that are unblocked in the sense that every agent finds the assignment acceptable, and no group of agents would prefer to be matched together rather than being matched with their respective assignments (Gale & Shapley, 1962). The prevalence of the guidance offered by stability is not surprising, as stable allocations capture a strong sense of fairness and describe how decentralized markets are ultimately organized (Roth & Vande Vate, 1990).

In some markets, however, stability is not the only goal. Two well–known examples are the assignment of medical residents to hospitals and the allocation of students to schools. In these markets, the designer is often concerned with possible distributional imbalances whereas rural regions end up with few doctors and too many students with the same race or socioeconomic background are allocated to the same school. In practice, the clearinghouses in these markets have attempted to achieve their desired distributional goals by imposing distributional constraints; namely, quotas on participating institutions (Abdulkadiroğlu (2005), Kamada & Kojima (2017)).

When a distributional goal is unstable—namely, when the desired distributional constraints are not satisfied by any stable allocation—achieving any desired allocation requires enforcement on participants ex post, once they have been allocated (Roth, 1986).[1] Thus, what distributional goals can be achieved depends on the "degree" of enforcement power vis a vis the set of ex-post blocks that the clearinghouse considers should be tolerated. These ex-post considerations have been the main focus of the literature (see, most notably, Kamada & Kojima (2015) and Kamada & Kojima (2016b)).

When participation in the clearinghouse is voluntary and the desired quotas cannot be enforced on nonparticipants, however, achieving unstable distributional goals requires not only ex-post enforcement, but also the ability to prevent agents from prearranging a matching ex ante, before the clearinghouse opens. But if the desired quotas cannot be enforced on nonparticipants deterministic mechanisms that produce unstable assignments cannot induce full participation (Roth (1984); Roth (2008), Sönmez (1999)). To the extent that a clearinghouse can enforce its allocations but the desired quotas can only be enforced on participants, deterministic mechanisms can therefore achieve no unstable distributional goal.

This paper identifies the scope of random mechanisms as an effective re–design by investigating what distributional goals could be achieved when a clearinghouse exerting *total* ex post enforcement power can randomly al-

[1] This is not necessarily true in dynamic settings. For example, Liu (2018) shows that distributional constraints can be implemented by means of self-enforcing punishments.

locate its participants but the market outside the clearinghouse organizes itself in a decentralized manner, so that non-participants are not bounded by the clearinghouse's desired quotas. I address the following questions. When is the clearinghouse able to induce every group of agents to participate ex ante? How does this ability depend on agents' preferences and the set of unstable allocations the clearinghouse can enforce ex post?

1.1 *Necessary conditions*

Consider a market designer who is evaluating whether a given distributional constraint can be achieved in a medical matching market. The market runs a clearinghouse that randomly allocates its participants, and every agent obtains a cardinal payoff from being matched to any agent on the other side of the market.

The designer assumes that any group of hospitals and doctors deciding whether to form a match among themselves instead of participating in the clearinghouse will jointly opt out if by doing so none of them expects a payoff loss and at least one expects a payoff gain. Both the random assignment proposed by the clearinghouse and the agents' payoffs are commonly known among all doctors and hospitals. The designer recognizes that the program's ability to achieve some unstable distributional goal by means of a random mechanism depends on three ingredients; namely, the program's degree of ex post enforcement on participants, the agents' cardinal payoffs, and whether nonparticipating agents are bounded by the program's desired quotas and have the ability to form random assignments among themselves. To the extent that the designer is clueless about these ingredients, how does she estimate whether the residency program is able to induce participation ex ante, and so achieve some desirable allocation ex post?

The designer might reasonably search for a robust prediction; namely, she might seek to understand what quotas *cannot* be achieved by *any* random mechanism when the program has *total* ex-post enforcement power, nonparticipants are not bounded by the clearinghouse's desired quotas, and any coalition of nonparticipant doctors and hospitals both can only form *deterministic* assignments among its members and would jointly opt out when by doing so none of them expects a payoff loss and at least one of them expects a payoff gain, for *every* cardinal payoff consistent with their preferences.

The designer recognizes that her hypotheses tell her little about the clearinghouse's ability to achieve a given distributional constraint. Indeed, her hypotheses presume both that the program can randomize over the whole set of assignments satisfying its desired constraints and that agents evaluate their ex-ante blocking opportunities extremely "cautiously."[2] How-

[2] While the assumption that nonparticipating coalitions can only form deterministic

ever, she is interested in identifying the scope of *any* random mechanism. Thus, she seeks to understand when she can conclude, with confidence, that a given distributional constraint *cannot be achieved at all.* Put another way, the designer's hypotheses give rise to a random, *ex ante* stability notion that can be used to identify what quotas *cannot* be achieved by *any* random clearinghouse endowed with total ex post enforcement power that is yet part of a larger, decentralized market.[3] It is worth to emphasize that, since this ex-ante stability notion describes the best-case scenario that a clearinghouse that evaluates the use of random mechanisms could face, the notion is not driven by a normative consideration. Instead, it responds to a positive, but theoretical, approach that an uncertain designer could take.

The designer could, of course, search for sharper, albeit less robust, predictions either by looking at more stringent notions of participation or by relaxing the assumption of total ex-post enforcement; i.e., by restricting the analysis to subsets of the market's set of desirable allocations that meet other ex post properties.[4] However, such a search would not only require information about the agents' cardinal payoffs but also a normative discussion about what ex post blocks a clearinghouse should tolerate. This discussion, which amounts to seek the "right" ex-post stability notion in the presence of distributional constraints, has been the main focus of the recent literature (see section 5.5 for a discussion). However, the role of participation in markets imposing distributional constraints has not yet been explored. Instead of taking a stand on these ex-ante and ex-post considerations, this paper presumes that a designer without information about the agents' cardinal payoffs might first want to search for necessary conditions; namely, conditions required by *every* random clearinghouse that achieves a given unstable distributional goal. Indeed, when no random mechanism

assignments is part of the designer's robustness exercise, it might also be a good approximation of the matching arrangements individuals and institutions could form in existing markets. Allowing blocking coalitions to form random assignments would assume that the outcome of such an assignment is enforceable *within* the coalition. While a clearinghouse might have the ability to enforce its assignments, it is not at all clear whether the members of a given coalition can commit to the outcome of a lottery.

[3]Notice that this ex-ante stability notion restricts attention to deterministic blocks, but allows for the same set of blocks that are considered in the absence of distributional constraints. While this seems restrictive, it is at the heart of the analysis. Indeed, when participation in the clearinghouse is voluntary but the desired quotas can be enforced on both participants and non-participants, imposing distributional constraints does not affect participation incentives. Thus, what distributional goals can be achieved depends only on the above ex-post considerations. In other words, requiring nonparticipants to be bounded by the clearinghouse's distributional constraint implies that the assumption of voluntary participation has not bite.

[4]Alternatively, the designer could pursue a different redesign. A natural one would be the use of monetary payments. However, Agarwal (2017) showed that monetary payments seem to be ineffective as a tool to increase the number of doctors in rural regions.

can induce full participation under the designer's hypotheses the desired distributional goal cannot be achieved no matter the degree of ex-post enforcement—or equivalently, no matter the ex-post stability notion satisfying the desired quotas—or notion of ex-ante participation one adopts.[5] Thus, this paper departs from—and thus complements—the ex-post focus of the literature by studying how the presence of voluntary participation affects the class of distributional constraints that can be achieved.

1.2 *Organization and preview of results*

Section 2 defines the main ingredients of the class of two-sided, many-to-one matching markets I will focus on and Section 3 introduces a random stability concept, random constrained stability, that accommodates the presence of distributional constraints and embeds the notion of participation I will consider.

Section 4 characterizes the existence of constrained stable lotteries. I say that agents' preferences are heterogeneous whenever the most preferred set of doctors of every hospital contains some doctor who deems another hospital better. I first show that under a joint condition on individual and distributional constraints, there is a random constrained stable lottery whenever agents' preferences are heterogeneous. Heterogeneity is a very strong condition, but it is not necessary for a random constrained stable lottery to exists. I show, however, that some "degree" of heterogeneity *is* necessary. Indeed, I show that a failure of heterogeneity entails that some group of doctors must be assigned to some hospitals with probability one by every random constrained stable lottery. Thus, a random constrained stable lottery exists if and only if agents' preferences are "sufficiently" heterogeneous with respect to the desired distributional constraints. The main finding of the paper is a polynomial-time algorithm that pin downs precisely the degree of heterogeneity required for a given distributional constraint. Thus, the algorithm determines, for every preference profile and distributional constraint, whether a random constrained stable lottery exists or not, and provides a lower bound on the quotas that can be achieved, at any given profile of preferences, by *any* random mechanism.

Section 5 discusses the assumptions, the implications of the results, their relationship with the literature, and some extensions. Section 6 concludes.

[5] Sections 5.5 and 5.7 discuss the consequences of relaxing the designer's hypotheses. In particular, section 5.5 argues that, in some cases, achieving *any* unstable distributional goal requires *total* ex-post enforcement power. Specifically, I show by example that in some cases there exists a random mechanism inducing full participation only if it puts positive probability on some assignment that is *not* envy-free. Thus, in those cases, achieving the desired distributional goal can only be done via "unfair" assignments.

2 The environment

2.1 *Agents and preferences*

The primitives of the model are the following. There is a set $I = D \cup H$ of agents—where D is a non-empty finite set of doctors and H a non-empty finite set of hospitals—that form a two-sided, many-to-one matching market without transfers. I denote by $i \in I$ a generic agent. A coalition C is a subset of $I \cup \{\emptyset\}$. Each doctor $d \in D$ has a strict preference relation $\succ_d$ over the set of hospitals and being unmatched; namely, over $H \cup \{\emptyset\}$. I denote by $\succsim_d$ the weak order of d and write, for any $h, h' \in H \cup \{\emptyset\}$, $h \succsim_d h'$ 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; namely, over $\mathcal{P}(D)$.[6] I denote by $\succsim_h$ the weak order of h and write, for any $D', D'' \subseteq D$, $D' \succsim_h D''$ if and only if $D' \succ_h D''$ or $D' = D''$. For any $h \in H$ and $d \in D$, hospital h is acceptable to d if and only if $h \succ_d \emptyset$, and doctor d is acceptable to hospital h if and only if $d \succ_h \emptyset$.[7] I assume that every hospital finds every doctor to be acceptable; i.e., $d \succ_h \emptyset$ for every $d \in D$ and every $h \in H$. I denote by $\succ = (\succ_i)_{i \in I}$ a preference profile. Throughout the paper, I assume that $\succ_h$ is responsive for every h:[8]

1. For any $D' \subseteq D$ and any $d \in D \setminus D'$ and $d' \in D'$, $(D' \cup \{d\}) \setminus \{d'\} \succsim_h D'$ if and only if $d \succsim_h d'$.

2. For any $D' \subseteq D$ and $d' \in D'$, $D' \succsim_h D' \setminus \{d'\}$ if and only if $d' \succsim_h \emptyset$.

2.2 *Assignments*

A matching—allocation or assignment—is a mapping μ that satisfies: i) $\mu_d \in H \cup \{\emptyset\}$ for every $d \in D$; ii) $\mu_h \in \mathcal{P}(D)$ for every $h \in H$; and iii) for any $d \in D$ and any $h \in H$, $\mu_d = h$ if and only if $d \in \mu_h$. I denote by $\mathbf{M}$ the set of all matchings and by $\Delta(\mathbf{M})$ the set of all lotteries—or random assignments—over $\mathbf{M}$. I denote by $Supp(\gamma)$ the support of $\gamma \in \Delta(\mathbf{M})$ and by $\gamma(\mu)$ the probability that $\gamma \in \Delta(\mathbf{M})$ assigns to $\mu \in \mathbf{M}$. The distribution that puts probability one to $\mu \in \mathbf{M}$ will be denoted by δ_μ.

Given a preference profile $\succ$ and two distributions $\gamma, \gamma' \in \Delta(\mathbf{M})$, I will say that γ [strictly] first-order stochastically dominates (FOSD) γ' for agent i given $\succ$ iff:

[6] For a set X, $\mathcal{P}(X)$ denotes its power set.

[7] Notice that since hospitals' preferences are defined on subsets of doctors, I should write $\{d\} \succ_h \emptyset$ instead of $d \succ_h \emptyset$. I will, however, slightly abuse the notation and identify singleton sets by their (unique) element.

[8] This is the standard notion of responsiveness (Roth & Sotomayor, 1990). It is often assumed that hospitals' preferences are responsive with respect to their individual capacity. However, this is a difference in notation, not in substance.

$$\sum_{\bar{\mu}:\bar{\mu}_i \succsim_i \mu_i} \gamma(\bar{\mu}) \geq \sum_{\bar{\mu}:\bar{\mu}_i \succsim_i \mu_i} \gamma'(\bar{\mu}) \text{ for every } \mu \in \mathbf{M},$$

[with strict inequality for some $\mu \in \mathbf{M}$].

2.3 *Constraints*

Each hospital h has a capacity $q_h > 0$ that describes the maximum number of doctors it can hire.[9] For any vector of individual capacities $\{q_h\}_{h \in H}$, I denote by:

$$\mathcal{F} := \{\mu \in \mathbf{M} : |\mu_h| \leq q_h \text{ for every } h\}$$

the set of feasible assignments; namely, the set of matchings that respect the capacity of every hospital.

I introduce distributional constraints by assuming the existence of a partition R of H where each cell $r \in R$ denotes a *region*.[10] I will write $r(h)$ for hospital h's region. For each region $r \in R$, there is a regional quota $q_r \geq 0$ that represents the maximum number of doctors who can be assigned to region r. For a given pair $(R, \{q_r\}_{r \in R})$, let $\mu_r := \bigcup_{h \in r} \mu_h$ denote the total number of doctors assigned to region r under $\mu \in \mathbf{M}$. For a given triplet $(R, \{q_r\}_{r \in R}, \{q_h\}_{h \in H})$, let $\mathcal{A}$ denote the set of admissible assignments, namely

$$\mathcal{A} := \{\mu \in \mathbf{M} : |\mu_r| \leq q_r \text{ for every } r \in R \text{ and } |\mu_h| \leq q_h \text{ for every } h \in H\}.$$

The term "admissible" aims to highlight that once the desired distributional constraint has been set, the market should look to produce a matching in this set.[11] Since each triplet $(R, \{q_r\}_{r \in R}, \{q_h\}_{h \in H})$ defines a set of admissible matchings $\mathcal{A}$, I will often use the two interchangeably.

2.4 *States*

A state of nature describes both preferences and constraints. Hence, I will write $\theta = (\succ, \mathcal{A})$ for a generic state and denote by Θ the set of all states of nature. I will write $\succ^\theta$ to denote the preference profile that realizes at θ and $\mathcal{A}(\theta)$ or $(R(\theta), \{q_r(\theta)\}_{r \in R(\theta)}, \{q_h(\theta)\}_{h \in H})$ to denote the realized admissible set at state θ. Moreover, I write $\mathcal{F}(\theta)$ to denote the set of feasible matchings at θ. These dependences will be omitted when no confusion arises. Notice

[9]Since each doctor can work for at most one hospital, one could add to the description a capacity for doctors. I will, however, omit such a capacity.

[10]Thus, $\bigcup_{r \in R} r = H$ and, for every $r, r' \in R$, $r \cap r' = \emptyset$ or $r = r'$.

[11]Some papers, notably Kamada & Kojima (2015), Kamada & Kojima (2018b) and Kamada & Kojima (2016b), call $\mathcal{A}$ the set of "feasible" matchings and define $\mathbf{M}$ in terms of the capacity of the hospitals. This is a difference in terminology, not in substance.

that $\mathcal{A}(\theta) \subseteq \mathcal{F}(\theta)$ for every $\theta \in \Theta$. States are assumed to be commonly known among doctors and hospitals.

2.5 *Stability*

A matching $\mu \in \mathbf{M}$ is blocked by a pair (C, μ'), where C is a coalition and $\mu' \in \mathcal{F}$ if i) $\mu'_i = \emptyset$ for every $i \notin C$; ii) $\mu'_d \in C$ for every $d \in C$; iii) $\mu'_h \subseteq C$ for every $h \in C$; iv) $\mu'_i \succsim_i \mu_i$ for every $i \in C$, $\mu'_i \succsim_i \mu_i$; and v) $\mu'_i \succ_i \mu_i$ for some $i \in C$.[12]

Armed with this blocking notion, one can define the class of stable assignments (Gale & Shapley, 1962): A matching μ is stable if and only if it is both feasible and unblocked. I will let $S(\theta)$ denote the set of stable assignments at state θ.[13]

I denote by $\bar{\Theta}$ the class of states where the distributional constraint is nontrivial in the sense that it is not satisfied by any stable assignment:

$$\bar{\Theta} := \{\theta \in \Theta : S(\theta) \cap \mathcal{A}(\theta) = \emptyset\}.$$

Intuitively, states in $\bar{\Theta}$ describe an attempt to reallocate some doctors with respect to their assignment at some stable assignment.[14]

3 Participation

3.1 *Features*

Participation in the clearinghouse of many centralized markets that impose distributional constraints is voluntary. For example, participation is voluntary for students and charter schools in most school districts (Ekmekci & Yenmez, 2014), and both doctors and hospitals are free to opt out of the U.S. and Japanese residency matching programs (Roth & Shorrer, 2018).[15]

[12]Notice that since coalitions are defined as subsets of $I \cup \{\emptyset\}$, unblocked assignments are individually rational in the following sense: A matching $\mu \in \mathbf{M}$ is individually rational if i) $\mu_d \succsim_d \emptyset$ for every $d \in D$, and ii) $\mu_h \succsim_h \bar{D}$ for every $\bar{D} \subseteq \mu_h$ and every h. Since hospitals have responsive preferences and find every doctor to be acceptable, ii) always holds in this paper.

[13]This stability notion is referred to as "group stability" by Roth & Sotomayor (1990). Whenever hospitals' preferences are responsive, however, the notion is equivalent to "pairwise" stability (Roth & Sotomayor, 1990) (pp. 130); namely, feasible assignments that are unblocked in the following sense. Given a matching μ, a pair (d, h) is a blocking pair if $h \succ_d \mu_d$ and either i) $d \succ_h \emptyset$, and $|\mu_h| < q_h$, or ii) $d \succ_h d'$ for some $d' \in \mu_h$. Moreover, responsiveness implies that a stable assignment exists at every state.

[14]Notice that by the "rural hospitals theorem", restricting attention whether the intersection is nonempty is without loss of generality.

[15]For details about the Japanese residency matching program, see http://www.jrmp.jp/kiyaku-byouin.htm.

In these clearinghouses, however, participation is typically restricted to agents that formed no matching before participating.[16] At the same time, agents opting out from the clearinghouse in these markets are typically free to form any feasible assignment; namely, doctors can work for any hospital not participating and nonparticipating hospitals can hire up to their true capacity. This is typically a feature of markets in which distributional constraints are aggregate—in the sense that regions contain a large number of institutions—but seems to hold also in some markets in which constraints are imposed at the level of the institutions.[17]

This section introduces a notion of participation that builds on these three features.

3.2 *Deterministic assignments*

In markets in which the clearinghouse produces deterministic allocations that can be enforced on participants ex-post but coalitions of agents can opt out, ex-ante, by forming any feasible matching among their members, the

[16] This is the case, for example, for the Japanese residency matching program (I'm grateful to Yusuke Narita for pointing this out). On the other hand, the U.S. national residency matching program (NRMP) imposes an "all in" policy that seeks to restrict the formation of prearrangements (Niederle & Roth, 2003). See http://www.nrmp.org/faq-questions/what-is-the-all-in-policy/.

[17] For example, California and Connecticut require charter schools to maintain a balanced intake of students in terms of race, ethnicity, and socioeconomic background. The Connecticut State statute states that the "State Board of Education shall consider the effect of the proposed charter school on the reduction of racial, ethnic, and economic isolation in the region in which it is to be located." Similarly, the state of California asks charter schools to describe "the means by which the school will achieve a racial and ethnic balance among its pupils that is reflective of the general population residing within the territorial jurisdiction of the school district to which the charter petition is submitted" (see http://www.sccoe.org/depts/esb/charter-schools-office/Documents/faqs/admission.pdf). In practice, however, most states do not establish explicit requirements to meet a diversity balance: By 2006, only 11 states included explicit provisions to enhance racial and ethnic balance. Nine of these have balance provisions that do not require charter schools to meet numeral indices of racial diversity and one of the other two, South Carolina, has a very flexible provision, only requiring the school's intake to differ by no more than 20 percent of that of district-run schools (Gajendragadkar, 2006). In turn, many charter schools show levels of segregation and racial isolation that are higher than those observed in public schools. By 2000, the U.S. Department of Education had identified at least 12 states in which charter schools were significantly more segregated than their surrounding district-run public schools (Parker, 2001). This observation is still valid in some states. Carlson and Seo report that in 2013, only 28 per cent of charter schools complied with the policy of reduced isolation introduced by the State of Connecticut (see http://commons.trincoll.edu/cssp/2015/05/11/reduced-isolation-2/). Since many charter schools do not participate in the clearinghouse for their own district (Ekmekci & Yenmez, 2014), this might suggest that some charter schools do not effectively face the quotas for target characteristics that public schools participating in the clearinghouse are subject to.

standard notion of blocking does not capture ex post blocking arrangements among participants, but rather ex ante participation constraints. Under this interpretation, it follows that no deterministic mechanism can achieve a distributional constraint that is only satisfied by unstable allocations.

Proposition 1. *There is an admissible and unblocked matching at θ if and only if every stable allocation at θ is admissible; i.e., there is an unblocked $\mu \in \mathcal{A}(\theta)$ if and only if $\theta \notin \bar{\Theta}$.*

Notice that the "only if" part hinges on a straightforward application of the rural hospitals theorem. Whenever some stable matching is not admissible, the rural hospitals theorem entails that no stable allocation is admissible. Thus, no admissible and unblocked matching exists.

Proposition 1 entails that clearinghouses that employ deterministic mechanisms can induce full participation if and only if they impose no distributional constraints. Put differently, clearinghouses that employ deterministic mechanisms can only achieve stable allocations. This result is therefore a formal counterpart, in markets that impose distributional constraints, of the observation that a market's ability to induce participation is closely related to its ability to produce stable allocations (Roth (1984), Niederle & Roth (2003)). Notice, importantly, that the result holds regardless of the set of admissible allocations a clearinghouse can enforce. Thus, ex post enforcement is necessary, but never sufficient, to achieve unstable allocations. Put another way, restricting attention to deterministic assignments seems to be a substantial restriction in markets in which participation is voluntary.[18]

3.3 *Random assignments*

Proposition 1 might lead a market designer to ask whether a clearinghouse could induce full participation—and thus achieve some nontrivial set of regional quotas—by randomly allocating its participants.

Imagine that one such designer considers a notion of participation that extends, naturally, the notion of participation in markets that run deterministic clearinghouses; namely, she assumes that a coalition of agents opts out whenever it can form a feasible assignment among its members in such a way that no member expects to be worse off and at least one member expects to be better off.

The designer recognizes that the ability of a random clearinghouse to induce full participation depends on three ingredients: the agents' cardinal

[18] Notice that Proposition 1 also entails that an admissible and unblocked allocation exists if and only if both deferred acceptance algorithms produce an admissible assignment.

payoffs; the set of admissible allocations the clearinghouse can randomize over; and whether nonparticipating coalitions can form a random assignment among its members. While the designer knows that these ingredients are commonly known among doctors and hospitals, she is however clueless about them. In those situations, how does she estimate whether the clearinghouse can induce full participation ex ante and so achieve some desirable allocation ex post?

This paper will presume that instead of taking a stand on the above ingredients, the designer wants to search for a robust prediction; namely, she seeks to understand whether the market's desired distributional goal can be achieved by *some* random assignment when i) the clearinghouse can randomize over the *whole* set of assignments that satisfy the desired quotas; ii) nonparticipating coalitions can only opt out by forming (feasible and) *deterministic* assignments; and iii) a coalition opts out only when doing so makes no member worse off and some member better off for *every* cardinal payoff consistent with their preferences.[19]

The designer of course knows that these three hypotheses tell her little about the clearinghouse's actual ability to achieve a given distributional constraint. Indeed, i) grants the clearinghouse total ex-post enforcement power over the allocation it produces and ii)-iii) jointly entail that agents evaluate ex-ante blocking opportunities very "cautiously".[20] Thus, these hypotheses should be understood as capturing the designer's interest in predicting what distributional goals *cannot* be achieved at all. Indeed, whenever no random clearinghouse can induce full participation under these hypotheses the designer can conclude, with confidence, that the desired distributional goal can be achieved by no random mechanism.[21]

3.3.1 *Random assignments: Blocking*

The ordinal nature of the notion of participation embedded in the designer's working hypotheses ii) and iii) above entails that a coalition of agents does not participate whenever its members can form a feasible and deterministic

[19]While the assumption that nonparticipating coalitions can only form deterministic assignments is part of the designer's robustness exercise, it might also be a good approximation of the matching arrangements individuals and institutions could form in existing markets. Allowing blocking coalitions to form random assignments would assume that the outcome of such an assignment is enforceable *within* the coalition. While a clearinghouse might have the ability to enforce their assignments, it is not at all clear whether the members of a given coalition can commit to the outcome of a lottery.

[20]Thus, this blocking notion resembles the blocking concept considered by Liu *et al.* (2014) for matching markets with incomplete information but no distributional constraints. Alternatively, this blocking notion could be interpreted as capturing agents that evaluate lotteries optimistically (Erdil & Ergin, 2008).

[21]Sections 5.5 and 5.7 offer a brief discussion about the consequences of relaxing these hypotheses.

assignment among themselves that first–order stochastically dominates the lottery for each of them, and strictly for some; i.e.,

Definition 1. $\gamma \in \Delta(\mathbf{M})$ *is blocked* by a pair (C, μ'), where C is a coalition and $\mu' \in \mathcal{F}$, if $\mu'_i = \emptyset$ for every $i \notin C$ and

1. *$\mu'_d \in C$ for every $d \in C$;*

2. *$\mu'_h \subseteq C$ for every $h \in C$;*

3. *for every $i \in C$, $\delta_{\mu'}$ FOSD γ; and*

4. *for some $i \in C$, $\delta_{\mu'}$ strictly FOSD γ.*

Notice that this blocking notion reduces to the standard one whenever the lottery is degenerate (see section 2.5 above). Moreover, notice that a coalition blocks a lottery if and only if it can form a feasible and deterministic assignment among its members where the assignment of every member of the coalition is weakly better than every assignment in the support of the lottery, and strictly better, for some member, than some assignment in the support of the lottery.

3.3.2 *Random assignments: Stability*

Together, the blocking notion described above and the designer's working hypothesis i)—namely that a clearinghouse that imposes a given distributional constraint can randomize over the whole set of admissible assignments—define an ex ante stability notion for centralized markets in which the clearinghouse imposes distributional constraints but cannot enforce participation:

Definition 2. $\gamma \in \Delta(\mathbf{M})$ is ***random constrained stable at*** θ *if:*

1. *Supp$(\gamma) \subseteq \mathcal{A}(\theta)$ and*

2. *γ is unblocked.*

Two comments are in order. First, notice that random constrained stability reduces to stability when we restrict γ to be degenerate and focus on states in $\Theta \setminus \bar{\Theta}$. Second, a random constrained stable lottery should be interpreted as a lottery over admissible assignments that induces full participation. Since random constrained stability builds from the designer's working hypotheses, however, it describes the best scenario for a clearinghouse. As a consequence, the existence of random constrained stable lotteries tells the designer little about the market's ability to induce full participation. Instead, the designer is interested in identifying the class of states in which a random constrained stable lottery does not exist. Indeed, whenever a random constrained stable lottery does not exist the designer can conclude, with confidence, that the desired distributional goal cannot be achieved by any random clearinghouse.

4 Achieving distributional goals

4.1 *Sufficient conditions*

This subsection introduces a joint condition on preferences and constraints that guarantees the existence of random constrained stable lotteries at states in $\bar{\Theta}$. Let $\dot{\succ}_d$ denote d's most preferred hospital and write $\dot{\succ}_h$ for h's most preferred set of doctors that satisfies q_h; i.e.:

$$\dot{\succ}_h := argmax_{\hat{D} \subseteq D : |\hat{D}| \leq q_h} \succ_h.$$

Definition 3. Agents' preferences are ***heterogeneous at*** θ if, for every hospital h such that $\dot{\succ}_d = h$ for some d, there is some $\tilde{d} \in \dot{\succ}_h$ such that $\dot{\succ}_{\tilde{d}} \neq h$.

Notice that heterogeneity is a joint condition on preferences and capacity constraints. To gain a more intuitive grasp, consider the following thought experiment. Imagine that we ask every doctor and hospital to point, respectively, to their most preferred hospital and *feasible* set of doctors. Consider the class of hospitals that are pointed by some doctor. Agents' preferences are heterogeneous whenever the set of doctors pointed by each of these hospitals contains at least one doctor pointing to a different hospital.

The following condition is a joint condition on capacity and distributional constraints:

Definition 4. *A state* θ *is* ***rich*** *if* $q_{r(h)} \geq q_h$ *for every* h.

Richness is a reasonable condition in markets in which the desired distributional goal is aggregate—that is, markets whose regions contain a large number of institutions, and thus the capacity of each institution is small relative to its regional quota. This is the case, for example, for the Japanese medical residency matching program, the assignment of graduate students to universities in China, and the assignment of undergraduate students to universities in Ukraine and the U.K. (see Kamada & Kojima (2015) for details).[22] On the other hand, richness is expected to fail whenever regions contain a single institution. This would be the case, for instance, for affirmative action policies implemented at the level of the schools.

The next result follows:

Proposition 2. *If agents' preferences are heterogeneous in a rich state θ, then there is a random constrained stable lottery at θ.*

The proof is in the appendix, but I go back to the thought experiment described before to provide an intuition. Think of the outside option of a hospital as the best feasible set of doctors pointing at it. Richness ensures that the outside option of, and the set of doctors pointed by, every hospital is admissible. In turn, heterogeneity ensures that no hospital points to its outside option. It follows that there exists a set of admissible assignments in which each assignment matches a different agent to the set of agents she/it points to. Consider any lottery that assigns positive probability (only) to all of these assignments. Since every hospital is assigned to the set of doctors it is pointing to with positive probability, every hospital would only be willing to form a blocking coalition with the set of doctors it is pointing to. However, not every doctor in the set would accept: Heterogeneity implies that the set must contain a doctor who is not pointing to the hospital. Since this doctor is assigned to the hospital she is pointing to with positive probability, she prefers to participate. The following example illustrates this construction:

Example 1. *Consider a market with three doctors and five hospitals and a state $\theta \in \bar{\Theta}$ with three regions, two rural, $r_1 = \{h_1\}$ and $r_3 = \{h_5\}$, and one urban, $r_2 = \{h_2, h_3, h_4\}$. Moreover, every region has a regional quota of one; i.e., $q_r = 1$ for every r. The following figure uses dashed (resp. solid) circles to describe the rural (resp. urban) regions, brackets to denote the capacity of every hospital, and solid arrows to represent the maximal element in the corresponding agents' preferences:*

[22]Within markets that impose aggregate distributional constraints where $q_r \geq 1$ for every $r \in R$, richness is always satisfied whenever every institution can hire at most one agent.

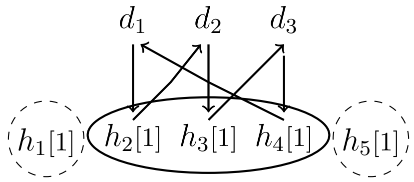

Figure 1: A rich state with heterogeneous preferences.

Notice that the state is rich and that agents' preferences are heterogeneous. Since the urban region has a regional quota of one, the distributional goal involves assigning one doctor to each rural region; i.e., to hospitals one and five. Each of the following admissible assignments assigns some doctor and urban hospital to the agent they point to:

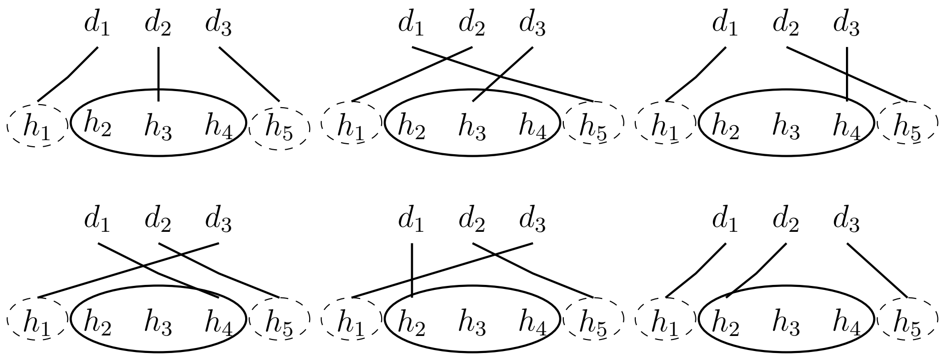

Figure 2: A set of admissible assignments.

It is not hard to check that every lottery over these assignments constitutes a random constrained stable lottery at θ. Thus, each rural region is assigned a doctor with probability one. $\square$

Notice that the strength of Proposition 2 lies in states in which stable assignments are not admissible; namely, states in $\bar{\Theta}$. Since unstable allocations can be achieved by no deterministic clearinghouse, Proposition 2 therefore highlights a stark contrast between deterministic and random clearinghouses. However, Proposition 2 is silent about how the given distributional constraint should be implemented. As a consequence, some random constrained stable lotteries might result in wasteful assignments; namely, assignments in which either some over-demanded regions end up with an intake of doctors that is less than the region's cap or not enough "urban" doctors are allocated to rural hospitals.

4.2 *Necessary conditions*

Proposition 2 seems to shed no light on what is of interest to the designer: The class of states in which a random constrained stable lottery does not exist. Since neither heterogeneity nor richness is sufficient for the existence of random constrained stable lotteries, however, Proposition 2 entails that both are necessary at some states. Thus, analyzing the class of states in which either heterogeneity or richness fail and no random constrained stable lottery exists might shed light on what is required by the existence of random constrained stable lotteries.

4.2.1 *Non-rich states*

It is not hard to see that richness is not necessary for the existence of random constrained stable lotteries. However, a failure of richness can mean that a hospital's outside option is inadmissible—that is, larger than the hospital's regional quota. In those cases, no random constrained stable lottery exists because the hospital finds every doctor to be acceptable (has responsive preferences), and no lottery over admissible assignments could assign to the hospital an intake of doctors equal to or larger than the size of its outside option. This observation always holds true at states in $\bar{\Theta}$ in which every region contains a single hospital, but also more generally as the following example illustrates:

Example 2. *Consider a market with three doctors and four hospitals and a state $\theta \in \bar{\Theta}$ with three regions, two rural, $r_1 = \{h_1\}$ and $r_3 = \{h_4\}$, and one urban, $r_2 = \{h_2, h_3\}$. Moreover, every region has a regional quota of one; i.e., $q_r = 1$ for every r. The following figure uses dashed (resp. solid) circles to describe the rural (resp. urban) regions, brackets to denote the capacity of every hospital, and solid arrows to represent the maximal element in the corresponding agents' preferences:*

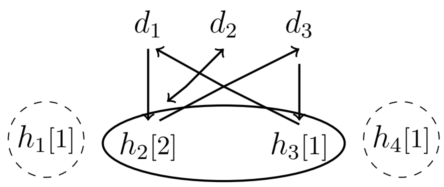

Figure 3: A non-rich state with inadmissible outside options.

Notice that preferences are heterogeneous, as h_2's outside option is $\{d_1, d_2\}$ but h_2 is pointing to $\{d_2, d_3\}$. Since $q_{h_2} = 2 > q_{r(h_2)} = 1$, however,

θ is not rich. Moreover, there is no random constrained stable lottery at θ. The reason is the failure of richness: No admissible assignment can assign two doctors to h_2. □

On the other hand, a failure of richness might entail that a group of doctors must be matched to some hospital with probability one. This would be the case whenever the outside option of a given hospital is admissible, but every assignment the hospital finds strictly better than its outside option is larger than the hospital's regional quota. In those situations, the hospital must be matched to its outside option with probability one by every unblocked lottery. This could be problematic, however, because the "conditional" regional quota in such a hospital's region would be zero, so no other hospital in the region facing a non-empty outside option could be induced to participate. Put differently, this could be problematic because richness could fail with respect to the conditional quotas. This possibility is illustrated in the following example:

Example 3. *Consider a market with three doctors and four hospitals and a state $\theta \in \bar{\Theta}$ with three regions, two rural, $r_1 = \{h_1\}$ and $r_3 = \{h_4\}$, and one urban, $r_2 = \{h_2, h_3\}$. Moreover, the state specifies the following regional quotas: $q_{r_1} = q_{r_3} = 1$ and $q_{r_2} = 2$. The following figure uses dashed (resp. solid) circles to describe the rural (resp. urban) regions, brackets to denote the capacity of every hospital, solid arrows to represent the maximal element in the corresponding agents' preferences, and dashed arrows to denote the second–best element in the corresponding agents' preferences:*

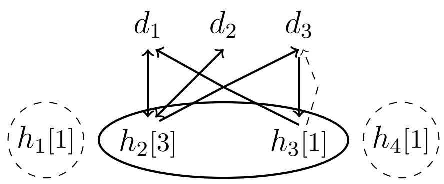

Figure 4: A non-rich state with conditionally inadmissible outside options.

Notice that agents' preferences are heterogeneous, as h_2 is pointing to $\{d_1, d_2, d_3\}$ but d_3 is pointing to h_3. Since $q_r = 2$, however, θ is not rich. Assume that h_2's outside option, $\{d_1, d_2\}$, is h_2's most preferred set of two doctors. While h_2's outside option is admissible, the unique set of doctors that h_2 finds better than its outside option, $\{d_1, d_2, d_3\}$, is not admissible. Thus, every random constrained stable lottery must assign $\{d_1, d_2\}$ to

h_2 with probability one. But then the conditional regional quota is zero. Hence, h_3's outside option, d_3, becomes inadmissible. Notice, then, that no reallocation is possible at this state even though preferences are heterogeneous.[23] □

4.2.2 *Aligned preferences*

It is not hard to see that heterogeneity is not necessary for the existence of random constrained stable lotteries. However, there are rich states in which heterogeneity fails and no random constrained stable lottery exists. Whenever heterogeneity fails, some sets of doctors must be assigned to some hospitals with probability one by every unblocked lottery. Thus, whenever a failure of heterogeneity means that the number of doctors who must be assigned to a given region with probability one is larger than the region's quota, no lottery can induce full participation. This is illustrated in the following example:

Example 4. *Consider a market with three doctors and four hospitals and a state $\theta \in \bar{\Theta}$ with three regions, two rural, $r_1 = \{h_1\}$ and $r_3 = \{h_4\}$, and one urban, $r_2 = \{h_2, h_3\}$. Moreover, every region has a regional quota of one; i.e., $q_r = 1$ for every r. The following figure uses dashed (resp. solid) circles to describe the rural (resp. urban) regions, brackets to denote the capacity of every hospital, and solid arrows to represent the maximal element in the corresponding agents' preferences:*

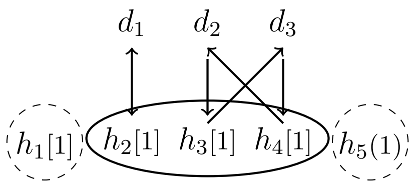

Figure 5: A rich state where preferences are not heterogeneous.

Notice that the state is rich but preferences are not heterogeneous, since d_2 and h_2 are pointing to each other. This implies that d_1 must be matched to h_2 with probability one by every unblocked lottery. But then the outside option of hospitals h_3 and h_4—doctors d_2 and d_3, respectively—is larger

[23]Notice that examples 2 and 3 illustrate that heterogeneity is not sufficient for the existence of random constrained stable lotteries.

than the conditional quota. Hence, no random constrained stable lottery exists at θ. □

A random constrained stable lottery exists in example 4 if the quota of the urban region is increased to two.[24] This suggests that no random clearinghouse can induce full participation in the state described in example 4, because agents' preferences are too aligned *with respect to* the desired quota. Put another way, a failure of heterogeneity does not necessarily mean that no quota can be achieved but it does restrict the extent of the reallocation that can be achieved. As the following example illustrates, however, whenever preferences are "completely aligned", no reallocation is possible.[25]

Example 5. *Consider a market with three doctors and four hospitals and a state $\theta \in \bar{\Theta}$ with three regions, two rural, $r_1 = \{h_1\}$ and $r_3 = \{h_4\}$, and one urban, $r_2 = \{h_2, h_3\}$. Moreover, every region has a regional quota of one; i.e., $q_r = 1$ for every r. The following figure uses dashed (resp. solid) circles to describe the rural (resp. urban) regions, brackets to denote the capacity of every hospital, and solid arrows to represent the maximal element in the corresponding agents' preferences:*

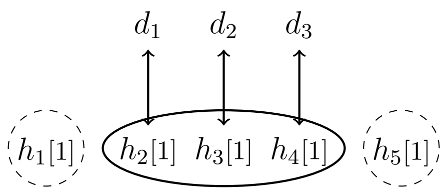

Figure 6: A rich state where preferences are not heterogeneous.

The state is rich, but agents' preferences are completely aligned in the sense that for every $i \in \{1, 2, 3\}$, d_i and h_{i+1} are the maximal element of each other. Hence, d_i must be matched to h_{i+1} with probability one for every $i \in \{1, 2, 3\}$ by every unblocked lottery. As a consequence, no random constrained stable lottery exists at θ. At this state, not even one doctor can be assigned to some rural region. □

[24]This, in fact, proves that heterogeneity is not necessary for the existence of random constrained stable lotteries.

[25]The preferences in this example belong to a domain of preferences for which no reallocation is possible. See section 5.3.

4.2.3 *Algorithm*

Whenever preferences are not heterogeneous and—or a failure of richness means that—a set of doctors must be matched to some hospital with probability one, heterogeneity or richness might in turn fail with respect to the set of unmatched agents and conditional quotas; namely, the regional quotas that take into account the intake of every matched hospital. As a consequence, new sets of doctors might have to be matched to some hospitals with probability one. In other words, failures of heterogeneity or richness might lead to yet further failures and matches. The following procedure keeps track of these successive failures and identifies the set of coalitions that must be matched with probability one by every unblocked lottery.

For $\mu \in \mathbf{M}$, a subset $\bar{D} \subseteq D$ of doctors is **μ-conditionally admissible** for hospital h if

$$|\bar{D}| \leq min\{q_h, q_{r(h)} - |\mu_{r(h)}|\}.$$

For any state $\theta = (\succ, \mathcal{A}) \in \Theta$, consider the following procedure $\mathcal{T}$:

Step 1: : Each doctor d points to her most preferred hospital h and each hospital h points to its most preferred set of doctors that satisfy both q_h and $q_{r(h)}$; if, for some h, a subset of the set of doctors pointing at h satisfies q_h but violates $q_{r(h)}$, then $\mathcal{T}$ **breaks down**; otherwise, if a hospital points to a subset of doctors who point at it, match the hospital to the subset. Let $\mu^1(\theta)$ denote the matching produced in this way and call any doctor and hospital *available* if they are not matched; if $\mu^1(\theta)$ is not admissible, then $\mathcal{T}$ **breaks down**; otherwise, it goes to step 2.

Step $k > 1$: : Each *available* doctor d points to her most preferred *available* hospital and each *available* hospital points to its most preferred $\mu^{k-1}(\theta)$-conditionally admissible set of *available* doctors; if, for some available h, a subset of the set of doctors pointing at h satisfies q_h but is not $\mu^{k-1}(\theta)$-conditionally admissible, then $\mathcal{T}$ **breaks down**; otherwise, if a hospital points to a subset of doctors who point at it, match the hospital to the subset; let $\mu^k(\theta)$ denote the matching that the algorithm produces up to Step k; if $\mu^k(\theta)$ is not admissible, then $\mathcal{T}$ **breaks down**; otherwise, it goes to step $k+1$.

The algorithm **terminates** either if it breaks down, or if $\mu^k(\theta) = \mu^{k-1}(\theta)$ for some $k \geq 1$ (where $\mu^0(\theta) = \emptyset$). In the latter case, I let $\mu(\theta)$ denote this matching. □

Notice that the number of steps $\mathcal{T}$ takes to terminate is bounded above by the number of agents, which is a finite number. It follows that $\mathcal{T}$ is *fast* in the sense that the number of steps it takes to terminate grows linearly with the number of agents.[26] Intuitively, $\mathcal{T}$ describes a procedure that resembles, at each step, the thought experiment described above; i.e., $\mathcal{T}$ asks hospitals to point to their most preferred *conditionally admissible* set of doctors but defines hospitals' outside options as their most preferred *feasible* set of doctors pointing at it. This "asymmetry" captures the key feature triggered by voluntary participation; namely, hospitals cannot be promised a conditionally inadmissible set of doctors by any lottery over admissible assignments, but they are free to form any feasible matching outside the clearinghouse.[27] The following example describes how $\mathcal{T}$ works:

Example 6. *Consider a market with eight doctors and seven hospitals and a state $\theta \in \bar{\Theta}$ with three regions, $r_1 = \{h_1, h_2\}$, $r_2 = \{h_3, h_4\}$ and $r_3 = \{h_5, h_6, h_7\}$ and the following regional quotas: $q_{r_1} = 3$, $q_{r_2} = 2$ and $q_{r_3} = 1$. The following figure uses brackets to denote the capacity of every hospital and solid (resp. dashed) arrows to represent the maximal (resp. second most preferred) element in the corresponding agents' preferences:*

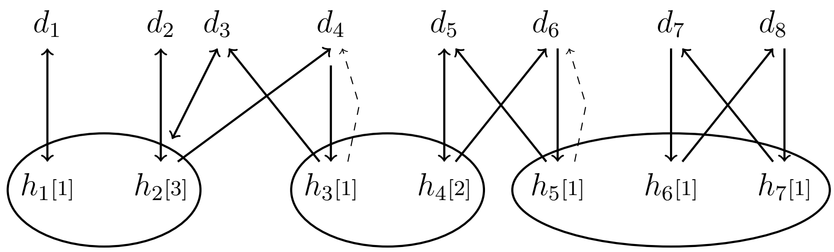

Figure 7: $\mathcal{T}$ at work.

Assume that h_2's most preferred set of two doctors is $\{d_2, d_3\}$, and h_4's most preferred doctor is d_5. The algorithm matches d_1 to h_1 in step 1 and produces $\mu^1(\theta)$, which is admissible. Given $\mu^1(\theta)$, $\mathcal{T}$ then matches h_2 to $\{d_2, d_3\}$ in step 2 and forms $\mu^2(\theta)$, which is admissible. Thus, $\mathcal{T}$ matches h_3 to d_4 in step 3 and forms $\mu^3(\theta)$, which is admissible. Given $\mu^3(\theta)$, $\mathcal{T}$

[26] I am grateful to Larry Samuelson and Philipp Strack for pointing this out.

[27] $\mathcal{T}$ resembles Gale's top trading cycle (TTC) algorithm. Unlike TTC, however, $\mathcal{T}$ is able to deal with the existence of distributional constraints, and this entails some important distinctions. First, the notion of a cycle behind the matches formed by $\mathcal{T}$ is different from that behind TTC, in that $\mathcal{T}$ only looks for cycles involving a single hospital. Second, $\mathcal{T}$ only matches hospitals that point to a subset of the set of doctors pointing at it, and so not every cycle involving a single hospital in the sense of TTC would be matched. These two distinctions imply that $\mathcal{T}$ might terminate matching no set of agents.

then matches h_4 to d_5 in step 4, produces the admissible assignment $\mu^4(\theta)$, and then matches h_5 to d_6 in step 5. In turn, $\mu^5(\theta)$ is admissible. But then both h_6 and h_7's most preferred feasible set of doctors are not $\mu^5(\theta)$-conditionally feasible: At step 6, the conditional cap of region 3 is 0. Hence, $\mathcal{T}$ breaks down. □

The next result is the main finding of the paper:

Proposition 3. *There is a random constrained stable lottery at θ if and only if $\mathcal{T}$ does not break down at θ.*

The proof is in the appendix, but the intuition is as follows: At each step, $\mathcal{T}$ breaks down whenever the outside option of some hospital is conditionally inadmissible because no lottery over admissible assignments can induce the hospital and its outside option to participate. Whenever no successive failure of richness means that some hospital's outside option is inadmissible, each successive failure of heterogeneity or richness entails that some coalitions must be matched by $\mathcal{T}$ and thus increases the lower bound on the quotas that can be achieved. As a consequence, the algorithm breaks down whenever the matching it forms is not admissible at some step.

Whenever the algorithm does not break down, both richness and heterogeneity hold for the set of unmatched agents and conditional quotas. Thus, Proposition 2 implies that a random constrained stable lottery exists for the set of unmatched agents and the vector of conditional regional quotas. Since the matching formed by the algorithm is admissible, a random constrained stable lottery exists for the set of all agents and the original vector of regional quotas.[28]

I end this section with three remarks. First, notice that $\mathcal{T}$ does not produce a random constrained stable lottery whenever one exists. Indeed, Proposition 3 states that the algorithm only checks whether a random constrained stable lottery can be constructed. Thus, Proposition 3 should be interpreted as providing a systematic way to identify the class of problems in which a random constrained stable lottery does not exist. Second, Proposition 3 entails that determining whether a random constrained stable lottery exists can be done fast.[29] Finally, $\mathcal{T}$ provides, for any fixed prefer-

[28]It is not hard to see, then, why $\mathcal{T}$ does not break down at rich states in which preferences are heterogeneous: At rich states, no hospital has a conditionally inadmissible outside option and every hospital points to a conditionally admissible set of doctors. In turn, whenever agents' preferences are heterogeneous, no hospital would point to its outside option. Hence, $\mathcal{T}$ matches no one and terminates—at step 1—without breaking down.

[29]One could, of course, determine whether a random constrained stable lottery exists or not by exhaustively checking whether some subset of admissible assignments is not blocked. However, doing this would require a number of steps that grows exponentially with the number of agents.

ence profile, a lower bound on the quotas that can be achieved: Those that can be imposed without compromising full participation.[30]

5 Discussion

5.1 *Origins*

This paper addresses when random constrained stable lotteries exist, but not how these lotteries might arise. This is due, in large part, to the nature of the exercise that gives rise to random constrained stability; namely, to investigate whether a market's distributional goal cannot be achieved by any random clearinghouse. To the extent that the existence of random constrained stable lotteries tells us little about the market's ability to achieve its distributional goal, the market should *only* be interested in states in which no random constrained stable lottery exists. In this sense, the analysis in this paper should be seen as one that offers a a stability notion that serves to identify what outcomes should not be expected in markets imposing distributional constraints much in the same way that core allocations guide our analysis in markets without distributional and participation constraints.

The analysis above does, however, suggest a natural way to address how random constrained stable lotteries could be obtained. $\mathcal{T}$ could be used to construct a random clearinghouse that delivers a constrained stable lottery at every state in which one such an lottery exists. In fact, $\mathcal{T}$ could be used to construct a random clearinghouse that guarantees full participation at every preference profile. Providing a mechanism that produces random constrained stable lotteries would be of value, but part of a subsequent exercise that seeks to understand whether, and when, random constrained stable lotteries can be implemented.[31] In particular, imple-

[30]Notice that heterogeneity and richness jointly describe the set of states in which the largest number of urban doctors can be reallocated. When richness holds, heterogeneity entails that no set of doctors has to be matched to some hospital with probability one. Hence, regional quotas are bounded, below, by the size of the hospital's largest outside option in each region.

[31]Budish *et al.* (2013) study the allocation of objects in the presence of general constraints and shed light on conditions that guarantee that a random assignment can be implemented; namely, expressed as a convex combination of deterministic assignments that satisfy the desired constraints. Their conditions are always satisfied when distributional constraints take the form assumed in this paper; namely, that of maximum quotas over partitions. While their notion of implementation is different from the sense in which constraints are achieved in this paper (as the latter involves a notion of participation), their result is, to my knowledge, the first attempt to understand what kinds of constraints can, in fact, be implemented. A similar spirit is present in Kamada & Kojima (2018b), who give a condition on the constraint structure that is necessary and

mentation would involve a careful consideration of the agents' incentives to report their preferences truthfully *after* deciding to participate.[32]

5.2 *Ex ante participation*

Random constrained stability assumes a cooperative, core–like notion of participation. The assumption seems reasonable in centralized markets in which the process of matching formation outside the clearinghouse is decentralized (Roth & Vande Vate, 1990). However, there are markets where alternative notions of participation might be more appropriate.

For example, Sönmez (1999) studies the possibility of prearranged matchings between hospitals and doctors in markets without distributional constraints in which participation is not restricted to agents who formed no pre–match (see also Kesten (2012)).[33] He show that this stronger notion of participation entails that not even stable allocations can be achieved by deterministic clearinghouses. A natural question for future research would be whether, in these markets, the formation of pre-matches can be avoided by random clearinghouses.[34]

On the other hand, Afacan (2016) and Ekmekci & Yenmez (2014) examine a Nash-like notion of participation in the assignment of students to schools; they study markets in which only schools opt out of the clearinghouse and can match with students after the clearinghouse produces an assignment. However, both papers consider markets that impose no distributional constraint and examine the ability of stable clearinghouses to induce full participation.

---

sufficient for the existence of a mechanism that is strategy-proof and stable in the sense they propose.

[32] I believe, but have not shown, that if one considers the notion of ordinal Bayesian incentive compatibility (see, e.g., Majumdar (2003)) the construction of a random stable lottery in the proof of Proposition 3 gives rise to a natural random constrained stable mechanism that is incentive compatible. Indeed, no agent matched by $\mathcal{T}$ can obtain a lottery that FOSD the lottery they obtain by reporting truthfully—when every other agent reports truthfully. At the same time, every unmatched doctor would be matched at most twice with positive probability; to her most preferred hospital and, possibly, to some other hospital. Similarly, every unmatched hospital would be matched twice with positive probability; to its best (available) set of doctors, and to some other set of doctors. Thus, no unmatched agent should be able to find a lottery that FOSD the lottery they would obtain when they report truthfully—and every other agent reports truthfully.

[33] Postlewaite (1979) studies a similar notion of manipulation in competitive markets. While Sönmez (1999) argues that some markets, like the Canadian lawyer market, do not restrict participation in this way, many other markets do. See section 3.1 for a brief discussion.

[34] Recall that stable (although possibly inadmissible) clearinghouses always induce full participation in this paper.

Finally, some papers have investigated different ex ante, random stability notions (see Aziz & Klaus (2017)) for a recent survey). However, none of these notions is equivalent to random constrained stability, as none of them capture the presence of distributional constraints.[35]

5.3 *Preference alignment*

The existence of random constrained stable lotteries rules out important domains of preferences known to be sufficient for the existence of a unique stable matching in one-to-one markets without distributional constraints. To my knowledge, the weakest of these conditions is the one proposed by Eeckhout (2000). Imagine that there are n doctors and n hospitals. Then, Eeckhout (2000)'s condition requires that for every $i = 1, 2, ..., n$, the doctor in position i prefers the hospital in position i over hospitals in positions $i + 1$, $i + 2$,...n, and the hospital in position i prefers the doctor in position i over doctors in positions $i + 1$, $i + 2$,...n.

It is not hard to see that no random constrained stable lottery exists at states in which the above condition holds. In fact, at those states no reallocation at all can be achieved.[36] Thus, a natural conjecture is that no random constrained stable lottery exists at states in which there is a unique stable lottery. Interestingly, however, the conjecture is false:

Example 7. *Consider a state with the following preferences:*

[35] Kesten & Ünver (2015) propose an ex ante stability notion that, under strict preferences, is equivalent to the notion of fractional stability of Roth *et al.* (1993) and weaker than the notion proposed by Alkan & Gale (2003). On the other hand, Manjunath (2017) describes two ex ante core notions, one weak and one strong, in which the concept of blocking is defined in terms of stochastic dominance. Under strict preferences, his strong notion is equivalent to the ex ante stability notion of Kesten & Ünver (2015). Afacan (2018) proposes the notion of claimwise stability—a weakening of fractional stability—that is stronger than the weak notion of Manjunath (2017). Since all of these notions reduce to stability when only deterministic matchings are considered, none is equivalent to random constrained stability. More generally, every notion equivalent to, and stronger than, the notion proposed by Kesten & Ünver (2015) induces a lottery over stable allocations. However, lotteries over unstable allocations can be random constrained stable. Finally, Yenmez (2013) considers stability notions that he interprets as capturing participation constraints. The motivation behind his notions is similar to the one in this paper in the sense that it relies on the idea that existing clearinghouses can enforce, ex-post, the allocations they produce. However, he analyzes one-to-one markets with transfers and no distributional constraints.

[36] In turn, no random constrained stable lottery exists at states in which the preference alignment is stronger than that entailed by Eeckhout (2000)'s condition. For example, no random constrained stable lottery exists at states in which preferences are aligned in the sense of Clark (2006) (see also Niederle & Yariv (2009)).

$$\begin{array}{ll} \succ_{d_1}: h_2 \succ h_1 \succ h_3 \succ \emptyset & \succ_{h_1}: d_1 \succ d_3 \succ d_2 \succ \emptyset \\ \succ_{d_2}: h_1 \succ h_2 \succ h_3 \succ \emptyset & \succ_{h_2}: d_2 \succ d_1 \succ d_3 \succ \emptyset \\ \succ_{d_3}: h_2 \succ h_1 \succ h_3 \succ \emptyset & \succ_{h_3}: d_3 \succ d_2 \succ d_1 \succ \emptyset \end{array}$$

Figure 8: A profile of preferences that support a unique stable matching.

When every hospital can hire at most one doctor, these preferences entail the existence of a unique stable matching; namely, the matching μ where $\mu_{d_i} = h_i$ for every $i = 1, 2, 3$. However, these preferences are heterogeneous, so that a random constrained stable lottery exists at every state in which there is, for example, a unique region with a regional quota of one. □

5.4 *Alternative distributional goals*

I have assumed that a market's distributional goal can be captured by a partition of hospitals and a unique vector of maximum, "hard" quotas. While these assumptions are a good approximation of how some markets seek to implement their desired reallocation, they might fail to capture more general distributional goals. For example, school districts typically seek to increase diversity with respect to several characteristics, and thus impose multiple maximum quotas (see, e.g., Goto *et al.* (2017)). Similarly, some markets might pursue distributional goals that do not entail a partition of institutions. For example, a residency matching program might seek to impose maximum quotas on both geographical regions and medical specialties (see, e.g., Kamada & Kojima (2018b)). Perhaps more importantly, maximum quotas are often used to implement some desirable vector of minimum, or floor quotas (see, e.g., Ehlers *et al.* (2014), Fragiadakis & Troyan (2017) and Hafalir *et al.* (2013)). Indeed, a medical matching program might impose a maximum quota on some urban regions only to satisfy some desired minimum quotas on some rural ones. Finally, some papers have investigated markets that might seek to implement "soft bounds"; namely, quotas that take into account the preferences of the participants and allow institutions to violate them to the extent that doing so is not unfair (see Ehlers *et al.* (2014) and Hafalir *et al.* (2013)).

Understanding whether, and when, the results in this paper extend to the cases above would be a natural and interesting next step. I suspect, however, that the main insights offered here have a close counterpart in all of them. The intuition is simple: Whether $\mathcal{T}$ matches a coalition of agents or not makes use of the partitional nature of the model or the assumption that regions are assigned a single maximum quota only to the extent that each hospital has a unique maximum quota. Since $\mathcal{T}$ could be modified to keep track of the minimum (conditional) quotas faced by each hospital at each step, I conjecture that the logic of the results would be valid in these other models.

On the other hand, the results in this paper extend to the case of floor constraints in a natural way whenever the clearinghouse can enforce every admissible assignment. In those cases, every floor quota imposed on a rural region can be implemented by a corresponding vector of maximum quotas on some urban regions. Thus, whether a given vector of minimum quotas can be achieved would amount to asking whether some corresponding vector of maximum quotas can be achieved. Put differently, the results in this paper apply to the case of minimum quotas, because achieving maximum quotas is necessary for achieving floor constraints.[37] Thus, $\mathcal{T}$ could be used to check whether a given distributional goal that involves minimum quotas can be achieved by some random clearinghouse.

While the results in this paper seem to make heavy use of the hard nature of the quotas, Proposition 3 would imply that no upper bound, hard or soft, that entails that the outside option of some hospital is violated can be achieved. To the extent that soft bounds are imposed at the level of the institutions (as in Ehlers *et al.* (2014) and Hafalir *et al.* (2013)), we should then expect difficulties trying to achieve them. Whether the results in this paper extend to the case of aggregate soft quotas is an interesting, but open question. One approach to this question would be to relax the requirement that the clearinghouse must produce an admissible assignment with probability one. While this would result in a weaker stability notion, the existence of preference alignment would nonetheless keep restricting what can be achieved. To see this, notice that in Example 5 no re-distribution can be achieved even if we allow for lotteries that produce an admissible assignment with "high probability".[38]

I end this subsection with a comment about some markets that might be seen as pursuing distributional goals without imposing explicit distributional constraints. For example, some school districts implement voucher programs that seek to reallocate students who attend low-performance public schools (Abdulkadiroğlu *et al.*, 2015), and most clearinghouses in kidney exchange markets seek to implement "long" chains of patient-donor pairs (Ashlagi & Roth, 2014).[39] Since private schools that participate in

[37] Notice, however, the importance of the assumption that the clearinghouse can enforce every admissible assignment in this argument. If a clearinghouse cannot enforce, say, individually irrational assignments, then achieving a given vector of maximum quotas would not entail that the associated floor constraints can be satisfied.

[38] I believe that this line of analysis could be close in spirit to the recent paper by Nguyen & Vohra (2018). They argue that in markets without distributional constraints in which hospitals' capacities can be considered part of the design of the market, the presence of couples does not rule stable solutions.

[39] Many voucher programs also seek to offer low-income families the possibility of

a voucher program effectively face a cap on the number of non-voucher students they can admit, and transplant centers participating in a clearinghouse might effectively face a cap on the number of its patient-donor pairs that are part of the chain implemented by the clearinghouse (Ashlagi & Roth, 2014), these markets could in fact be seen as imposing (some form of) distributional constraints. Whether the analysis and results of this paper have formal counterparts in these markets would of course be an interesting question to investigate (see, however, section 5.6).[40]

5.5 *Ex post enforcement*

When the desired distributional constraints are not satisfied by any stable allocation, achieving the desired distribution of doctors requires ex post enforcement (Roth, 1986).[41]

Random constrained stability builds on the assumption that a clearinghouse's ex-post enforcement power is total. The assumption is grounded on the nature of the designer's robustness exercise, but is also consistent with what it is observed in some existing markets. As Kamada & Kojima (2015) (p.13) point out, this is the case, for example, for the Japanese residency matching program: "*Indeed, in Japan, participants seem to be effectively forced to accept the matching announced by the clearinghouse because a severe punishment is imposed on deviators.*"[42] Another good example might be the U.S. national residency matching program (see Yenmez (2013)).

Since not every clearinghouse might be able to enforce every possible allocation that satisfies its desired distributional constraints, it is only natural

choosing private schools; see Afacan (2019) for a recent discussion.

[40]Participation in voucher programs is voluntary for both students and private schools (Abdulkadiroğlu *et al.*, 2015), and transplant centers are free to opt out of clearinghouses in kidney exchange markets, but participation is not restricted in any way (Ashlagi & Roth, 2014).

[41]As mentioned in the introduction,Liu (2018) shows that this is true only when one restricts attention to static environments. Indeed, the possibility of future punishments might allow a designer to rely on self-enforcement to implement her desired distributional constraints. Interestingly, Liu (2018) obtains results that are very close to those found in this paper. In particular, he shows that, in general, some hospitals must be assigned to the same number of doctors they would be matched to in the static environment.

[42]School districts are not always able to enforce the assignment they produce for students, as some participating students have been found to reject their assignments and look for enrollment in some non-participating school (Ekmekci & Yenmez, 2014). On the other hand, the ability to enforce assignments among participants is also present in markets that impose no distributional constraints. For example, the "commitment clause" allows only unmatched participants to find a (different) match in the U.S. national residency matching program (see http://www.nrmp.org/policies/the-match-commitment/). Interestingly, however, the presence of this clause suggests that the program does not (seek to) produce stable allocations.

to ask how (much) a market's ability to induce full participation depends on its ex-post enforcement power.

Relaxing the clearing house's ability to enforce its allocations would demand a stand on what admissible assignments *can* or *should* be enforced or, equivalently, what ex post objections can or should be tolerated. In turn, this stand should involve both a positive and a normative discussion. On the positive side, the degree of enforcement might depend on some ex-post, normatively appealing, properties.[43] On the normative side, the presence of a given degree of ex post enforcement does not however answer what admissible assignments a clearinghouse should seek to achieve. Searching for the "right" ex post stability notion has been the main focus of the literature (see, e.g., Kamada & Kojima (2015), Kamada & Kojima (2016b), Kamada & Kojima (2017), and Kamada & Kojima (2018a)). Instead, this paper departs from the ex-post viewpoint taken by the literature by investigating how the presence of voluntary participation affects the class of distributional constraints that can be achieved.

I believe, however, that a natural way to weaken the clearinghouse's enforcement power would be to consider unblocked lotteries over envy-free assignments (Wu & Roth, 2016).[44] While a full analysis is outside the scope of this paper, the following lines offer a brief discussion.

Definition 5. Given an assignment $\mu \in \mathbf{M}$, d has *justified envy* towards d' if $\mu_{d'} \neq \emptyset$, $\mu_{d'} \succ_d \mu_d$ and $d \succ_{\mu_{d'}} d'$.

An assignment $\mu \in \mathbf{M}$ is envy-free if no doctor has justified envy. There are two reasons why envy-freeness is a natural "ex-post fairness" notion in the presence of distributional constraints. First, an envy-free assignment exists at every state (e.g., the empty assignment is envy-free). Second, and more importantly, envy-free assignments rule out blocks where a hospital "adds" doctors. Put differently, envy-freeness considers only blocks that do not involve a vacant position of some hospital. Since distributional constraints reduce the intake of hospitals in over-demanded regions, envy-freeness is then a natural ex-post restriction in these markets.[45]

[43] Indeed, legal or institutional restrictions might require that some markets are able to enforce admissible assignments that satisfy, in addition, certain ex post properties. These restrictions would then naturally reduce the subset of admissible assignments a market would focus on.

[44] To my knowledge, every stability notion considered by the literature (in the presence of distributional constraints) satisfies envy-freeness.

[45] I believe that envy-free assignments can be motivated both normatively and positively. Indeed, a clearinghouse imposing a distributional constraint might not be able to prevent the formation of ex-post blocks where the size of the intake of a blocking hospital remains the same. At the same time, it would be natural to imagine that envy-freeness is part of the goals pursued by a clearinghouse regardless of its enforcement power.

The next example illustrates that there are states at which a lottery over assignments satisfying the desired constraints is unblocked *only if* it puts positive probability on assignments that are *not* envy-free. Thus, there is a "maximal domain conflict" between ex-post fairness and a market's distributional goal. Put another way, total ex-post enforcement is necessary for achieving *some* re-distribution at some states.[46]

Example 8. *Consider a market with two doctors and three hospitals and a state $\theta \in \bar{\Theta}$ with two regions, one urban, $r_1 = \{h_1, h_2\}$, and one rural, $r_2 = \{h_3\}$. Both regions have a regional quota of one and the agents' preferences are as follow:*

$$\succ_{d_1}: h_1 \succ h_2 \succ h_3 \succ \emptyset \quad \succ_{h_1}: d_2 \succ d_1 \succ \emptyset$$
$$\succ_{d_2}: h_2 \succ h_1 \succ h_3 \succ \emptyset \quad \succ_{h_2}: d_1 \succ d_2 \succ \emptyset$$
$$\succ_{h_3}: d_1 \succ d_2 \succ \emptyset$$

The following figure uses a dashed (resp. solid) circle to describe the rural (resp. urban) region, brackets to denote the capacity of every hospital, and solid arrows to represent the maximal element in the corresponding agents' preferences:

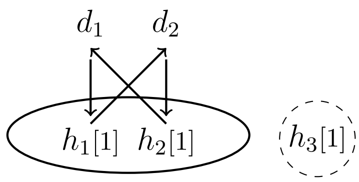

Figure 9: No unblocked lottery over admissible and envy-free assignments.

Notice that agents' preferences are heterogeneous and the state is rich. Thus, a random constrained stable lottery exists. However, every *random constrained stable lottery must match every agent to her/its most preferred agent with positive probability. But then, the support of every random constrained stable lottery must include an assignment where d_1, say, is matched to h_1 and h_2 is unmatched. In this assignment, however, d_2 will have justify envy towards d_1.* □

[46]It it not hard to see that restricting attention to admissible and individually rational assignment might deliver a similar conclusion. Indeed, no unblocked lottery over individual rational assignments exists in Example 8 below if d_1 and d_2 find, respectively, h_2 and h_1 to be unacceptable. Thus, the existence of unblocked lotteries might also rely on the ability of the clearinghouse to enforce, ex-post, individually irrational assignments.

5.6 *Implications*

Many markets employ deterministic clearinghouses but cannot enforce participation. Within the class of these markets in which the clearinghouse has ex post enforcement power and imposes distributional constraints, Proposition 1 could be used to rationalize why some of them fail to achieve their desired distributional goals. For example, some school districts that pursue affirmative action policies fail to induce all charter schools to participate (Ekmekci & Yenmez, 2014), and schools are heavily segregated (Frankenberg & Lee (2003), Parker (2001)). Similarly, not every hospital and resident participates in the Japanese medical residency matching program (Besstremyannaya, 2015), and the distribution of physicians seems to have worsened (Hara *et al.*, 2017). Moreover, high-quality private schools opt out of voucher programs, and some transplant centers withhold part of their endowment of patient-donor pairs when participating in some clearinghouse. Thus, Proposition 1 could also explain why some voucher programs fail to achieve a desirable reallocation of students (Abdulkadiroğlu *et al.*, 2015) and why transplant opportunities are sometimes lost in some kidney exchange markets (Ashlagi & Roth, 2014).[47]

At the same time, Propositions 2 and 3 suggest that the use of random clearinghouses might be an effective redesign.[48] In some markets, however, the use of lotteries should be accompanied by a careful design of both regions and quotas. For example, no random clearinghouse would induce full participation in school districts unless affirmative action policies are implemented at an aggregate (e.g., neighborhood) level. Proposition 3 suggests, in addition, that regional quotas should be part of the allocation to be implemented. After all, a vector of quotas can be achieved only if agents' preferences are not too aligned with respect to them. One could then imagine a random clearinghouse that, eliciting only ordinal information, builds around $\mathcal{T}$ and seeks to achieve no vector of quotas for which the reported preferences would cause $\mathcal{T}$ to breaks down.[49]

[47]Proposition 1 might also explain why some markets that pursue a distributional goal do not run centralized assignment systems—such as the assignment of students to schools in the United Kingdom—and why some centralized markets facing distributional imbalances have not introduced a distributional constraint, such as the assignment of residents to hospitals in the United States.

[48]One could, of course, search for alternative redesigns. A natural one would be the use of monetary payments. As recently shown by Agarwal (2017), however, monetary payments seem to be ineffective as a tool to increase the number of doctors in rural regions.

[49]A natural alternative would be, of course, to form an estimate of the degree of preference alignment and set regional quotas accordingly.

5.7 *Necessary conditions*

Random constrained stability was constructed from hypotheses that arguably describe the best scenario faced by a clearinghouse that evaluates the use of random mechanisms. As I hope it is clear by now, this robustness exercise is not driven by a normative consideration but instead responds to the approach that an uncertain designer could reasonably take.

One could nonetheless ask how the analysis and results would change if alternative hypotheses are considered. Put another way, one might be interested in analyzing some strengthening of random constrained stability. Relaxing the designer's hypotheses would amount to asking whether, and when, full participation can be induced when a clearinghouse cannot randomize over the whole space of admissible assignments or a larger class of ex ante blocking coalitions can be formed. The former could be motivated by either the desire or the need to satisfy other ex post properties, and so it is ultimately part of a market's design (see section 5.5).

Instead, relaxing the notion of participation would require to take a stand on what kind of coalitions agents can form and how these coalitions evaluate lotteries, or equivalently how they evaluate their ex-ante blocking opportunities. I believe that there are two natural ways to strengthen the notion of participation behind random constrained stability. First, one might allow coalitions to form random assignments among its members. While allowing for "random blocks" would deliver interesting possibilities, it is unclear to me whether they would be well grounded in real markets.

Second, one might still restrict blocks to be deterministic but strengthen the way agents evaluate lotteries. One possibility would be to assume that they evaluate their blocking opportunities "incautiously"; namely, that agents would compare any potential blocking assignment with the *worst* assignment in the support of the lottery produced by the clearinghouse. This case would describe the other extreme of the notion of participation assumed in this paper. It is not hard to see, however, that if agents are incautious in this sense then random assignments do not improve upon deterministic ones.

Alternatively, one could strengthen the notion of participation by assuming that the members of a coalition opt out whenever they expect a gain by doing some, for some *fixed* cardinal payoff. While this assumption could be considered the natural stand to take in terms of participation, that would only be so provided the designer has information about the agents' cardinal payoffs.[50]

[50] An older version of the paper analyzes these possibilities in some detail.

6 Conclusions

Many centralized matching markets find unstable allocations undesirable, and so impose distributional constraints; namely, quotas on participating institutions. When participation in the clearinghouses of these markets is voluntary, however, achieving unstable allocations requires not only enforcement ex post among participants but also full participation ex ante. Since unstable and deterministic clearinghouses cannot induce full participation (Sönmez, 1999), this paper investigates whether distributional constraints can be achieved by random clearinghouses by introducing a notion of random stability for centralized markets—random constrained stability—that embeds ex ante participation constraints.

Random constrained stability describes necessary conditions for participation, and so identifies whether a market's distributional goal is achievable at all. I show that a random constrained stable lottery exists if, but only if, agents' preferences are not too aligned. Thus, the presence of distributional constraints might give rise to a novel reason for a market's unraveling. The main finding of the paper is a polynomial-time algorithm that checks whether a random constrained stable lottery exists, and so provides a lower bound on the quotas that can be achieved by some random clearinghouse.

By departing from the ex post analysis conducted in the literature, this paper suggests that voluntary participation might be an important, although somewhat overlooked, feature of most centralized markets. That is, a market's inability to enforce participation might impose substantial restrictions on the class of unstable allocations that can be achieved. To the extent that participation in existing markets is voluntary, but they insist on using deterministic clearinghouses, the results in this paper rationalize why some markets fail to achieve their desired unstable allocations and shed light on whether the use of random mechanisms might be an effective re design.

7 Appendix: Proofs

Since Proposition 2 is a corollary of Proposition 3, this appendix contains the proofs of Proposition 1 and Proposition 3.

Proof of proposition 1

Fix any state $\theta = (\succ, \mathcal{A})$ and assume that there is a unblocked matching $\mu \in \mathcal{A}(\theta)$. It follows that $\mu \in \mathcal{F}(\theta)$, so that μ is stable at θ. Hence, $\theta \notin \bar{\Theta}$. For the other direction, assume that $\theta \notin \bar{\Theta}$. By the Rural Hospitals Theorem, it follows that no stable allocation at θ is admissible. But then, for every matching μ that is unblocked at θ we must have that $\mu \notin \mathcal{A}(\theta)$. $\square$

Proof of Proposition 3

At any state θ such that $\theta \notin \bar{\Theta}$, the existence of random constrained stable lotteries can be determined by running (some version of) deferred acceptance algorithm. Hence, suppose that $\theta \in \bar{\Theta}$.

I first show the "only if" part. Suppose, contrary to hypothesis, that $\mathcal{T}$ breaks down at θ but some random constrained stable lottery exists. Since $\mathcal{T}$ breaks down at θ, there is either:

1. some **Step** $k \geq 1$ such that $\mu^k(\theta)$ is not admissible; or
2. some **Step** $k \geq 1$ and hospital h such that a subset of the set of available doctors pointing at h satisfies $q_h > 0$ but is not $\mu^{k-1}(\theta)$-conditionally admissible.

I start from 1. I show by induction on k that the existence of some random constrained stable lottery entails that $\mu^k_h(\theta) = \mu_h$ for every $\mu \in Supp(\gamma)$ and every $h : \mu^k_h(\theta) \neq \emptyset$. Thus, the assumption that there is some k such that $\mu^k(\theta)$ is not admissible will lead to a contradiction. Take any random constrained stable lottery γ. Suppose that $k = 1$ and consider any $h : \mu^1_h(\theta) \neq \emptyset$. Since $\mu^1_h(\theta) \succsim_h A$ for every $A \subseteq D$ such that $|A| \leq min\{q_h, q_{r(h)}\}$, it follows that:

$$\mu^1_h(\theta) \succ_h \mu_h \text{ for every } \mu \in Supp(\gamma) : \mu_h \neq \mu^1_h(\theta).$$

Moreover, we have that $h \succ_d \tilde{h}$ for every $d \in \mu^1_h(\theta)$ and every $\tilde{h} \neq h$. Hence, for every $d \in \mu^1_h(\theta)$:

$$\mu^1_d(\theta) \succ_d \mu_d \text{ for every } \mu \in Supp(\gamma) : \mu_d \neq \mu^1_d(\theta).$$

If $\mu^1 \notin \mathcal{A}(\theta)$, $Supp(\gamma) \subseteq \mathcal{A}(\theta)$ implies that for every $\mu \in Supp(\gamma)$ there must be some hospital $h : \mu^1_h(\theta) \neq \emptyset$ and some doctor d such that $d \in \mu^1_h(\theta)$ and $d \notin \mu_h$. Consider, then, the following pair (C, μ'), where $C := \{h\} \cup \mu^1_h(\theta)$ and μ' is defined as follows:

$$\mu'_i = \begin{cases} h & i \in C : i \neq h \\ \mu^1_h(\theta) & i = h \\ \emptyset & i \notin C \end{cases}$$

Since $|\mu^1_h(\theta)| \leq min\{q_h, q_{r(h)}\}$, it follows that $\mu' \in \mathcal{F}$. Hence, (C, μ') blocks γ, a contradiction. Hence, for every $\mu \in Supp(\gamma)$ we must have:

$$\mu_h := \mu^1_h(\theta) \text{ for every } h : \mu^1_h(\theta) \neq \emptyset.$$

Suppose that, up to some $k > 1$, we have that for every $\mu \in Supp(\gamma)$ (inductive hypothesis):

$$\mu_h := \mu_h^k(\theta) \text{ for every } h : \mu_h^k(\theta) \neq \emptyset.$$

I now show that this implies that, for every $\mu \in Supp(\gamma)$ we must have:

$$\mu_h := \mu_h^{k+1}(\theta) \text{ for every } h : \mu_h^{k+1}(\theta) \neq \emptyset.$$

Take any $h : \mu_h^{k+1}(\theta) \neq \emptyset$ and notice that $\mu_h^{k+1}(\theta) \succsim_h A$ for every $\mu^k(\theta)$-conditionally admissible set A such that:

$$A \subseteq D \setminus \{d : \mu_d^t(\theta) \neq \emptyset \text{ for some } t \leq k\}.$$

By the inductive hypothesis, we must have:

$$\mu_h^{k+1}(\theta) \succ_h \mu_h \text{ for every } \mu \in Supp(\gamma) : \mu_h \neq \mu_h^{k+1}(\theta).$$

Moreover, we must have $h \succ_d \tilde{h}$ for every $d \in \mu_h^{k+1}(\theta)$ and every:

$$\tilde{h} \neq h : \tilde{h} \in \{h : \mu_h^t(\theta) = \emptyset \text{ for every } t \leq k\}.$$

Hence, for every $d \in \mu_h^{k+1}(\theta)$:

$$\mu_d^{k+1}(\theta) \succ_h \mu_d \text{ for every } \mu \in Supp(\gamma) : \mu_d \neq \mu_d^{k+1}(\theta).$$

By the induction hypothesis, for every $t \leq k$ and every $\mu \in Supp(\gamma)$ we must have $\mu_h = \mu_h^t(\theta)$ for every h such that $\mu_h^t(\theta) \neq \emptyset$. If $\mu^{k+1}(\theta) \notin \mathcal{A}(\theta)$, then $Supp(\gamma) \subseteq \mathcal{A}$ implies that for every $\mu \in Supp(\gamma)$, there must be some hospital h with $\mu_t^{k+1}(\theta) = \emptyset$ for every $t \leq k$ and $\mu_h^{k+1}(\theta) \neq \emptyset$ and a doctor d such that $d \in \mu_h^{k+1}(\theta)$ and $d \notin \mu_h$. Consider, then, the following pair (C, μ'), where $C := \{h\} \cup \mu_h^{k+1}(\theta)$ and μ' is defined as follows:

$$\mu'_i = \begin{cases} h & i \in C : i \neq h \\ \mu_h^{k+1}(\theta) & i = h \\ \emptyset & i \notin C \end{cases}$$

Since $\mu_h^{k+1}(\theta)$ is $\mu^k(\theta)$-conditionally admissible, it follows that $\mu' \in \mathcal{F}$. Hence, (C, μ') blocks γ; a contradiction. Hence, for every $\mu \in Supp(\gamma)$ we must have:

$$\mu_h := \mu_h^{k+1}(\theta) \text{ for every } h : \mu_h^{k+1}(\theta) \neq \emptyset.$$

It thus follows that, for every $\mu \in Supp(\gamma)$ we must have:

$$\mu_h := \mu_h(\theta) \text{ for every } h : \mu_h(\theta) \neq \emptyset.$$

This completes the "only if" part.

I now show 2. Take, then, any **Step** $k \geq 1$ and any hospital h and assume that a subset of the set of doctors pointing at h satisfies $q_h > 0$ but is not μ^{k-1}-conditionally admissible. Hence, h's most preferred subset of the set of available doctors pointing at it satisfies q_h but is not $\mu^{k-1}(\theta)$-conditionally admissible. This follows because $\succ_h$ is responsive and h finds every doctor to be acceptable. Call this set O_h. By 1. we know that, for every $\mu \in Supp(\gamma)$ and every h such that $\mu^t_h(\theta)$ for some $t \leq k-1$, we must have $\mu_h = \mu^t_h(\theta)$. Since $Supp(\gamma) \subseteq \mathcal{A}(\theta)$, it follows that $\mu_h \neq O_h$ for every $\mu \in Supp(\gamma)$. By construction, $O_h \succ_h A$ for every $\mu^{k-1}(\theta)$-conditionally admissible A such that:

$$A \subseteq D \setminus \{d : \mu^t_d(\theta) \neq \emptyset \text{ for some } t \leq k-1\}.$$

Hence, $O_h \succ_h \mu_h$ for every $\mu \in Supp(\gamma)$. Similarly, we must have $h \succ_d \tilde{h}$ for every $d \in O_h$ and every $\tilde{h} : \mu^t_h(\theta) = \emptyset$ for every $t \leq k-1$. Thus, for every $d \in O_h$:

$$h \succ_h \mu_d \text{ for every } \mu \in Supp(\gamma) : \mu_d \neq h.$$

Consider the following pair (C, μ'), where $C := \{h\} \cup O_h$ and μ' is defined as follows:

$$\mu'_i = \begin{cases} h & i \in C : i \neq h \\ O_h & i = h \\ \emptyset & i \notin C \end{cases}$$

Since $\mu' \in \mathcal{F}$, (C, μ') blocks γ; a contradiction.

I now show the "if" part. Suppose that $\mathcal{T}$ does not break down. By construction, $\mathcal{T}$ then produces a sequence of admissible assignments and, by hypothesis, it terminates at any step $k \geq 1$ where $\mu^k(\theta) = \mu^{k-1}(\theta)$. Assume, then, that $\mathcal{T}$ terminates at step $k \geq 1$. By construction, for every $h : \mu^{k-1}_h(\theta) = \emptyset$, h's most preferred subset of the set of available doctors pointing at it satisfying q_h is $\mu^{k-1}(\theta)$-conditionally admissible. Let's denote this set by O_h. This is h's "outside option" at k. On the other hand, by construction no h with a non-empty outside option points to its outside option. Let P_h then denote the set of doctors pointed by h. Similarly, define O_d and P_d for every d such that $\mu^t_d(\theta) = \emptyset$ for every $t \leq k-1$. Notice that, by construction, both O_i and P_i are $\mu^{k-1}(\theta)$-conditionally admissible for every $i : \mu^{k-1}_i(\theta) = \emptyset$.

Recall that $\mu(\theta)$ denotes the output of $\mathcal{T}$ when it terminates without breaking down. By construction, then, $\mu_h(\theta) := \mu^t_h(\theta)$ whenever $\mu^t_h(\theta)$ for some $t \leq k-1$. Notice that $\mu(\theta) \in \mathcal{A}$. Given $\mu(\theta)$, define:

$$A^D := \{d : \mu_d(\theta) = \emptyset\} \text{ and } A^H := \{h : \mu_h(\theta) = \emptyset \text{ and } O_h \neq \emptyset\}$$

to be, respectively, the set of available doctors and the set of available hospitals with a non-empty outside option. I now construct a set of admissible assignments:

1. pick any $d \in A^D$ and consider any assignment μ^d where: i) $\mu^d_d = P_d$ and ii) for every $h : \mu_h(\theta) \neq \emptyset$, $\mu^d_h = \mu_h(\theta)$;

2. pick any $h \in A^H$ and consider any assignment μ^h where: i) $\mu^h_h = P_h$ and ii) for every $h : \mu_h(\theta) \neq \emptyset$, $\mu^d_h = \mu_h(\theta)$.

Notice that, by construction, $\mu^d \in \mathcal{A}$ and $\mu^h \in \mathcal{A}$ for every $d \in A^D$ and every $h \in A^H$. Let F be the set containing all (and only) these admissible assignments. I now show that any lottery over F is a random constrained stable lottery. Denote any such lottery by γ.

It is not hard to see that the argument made in 1. of the "only if" part entails that no $i : \mu_i(\theta) \neq \emptyset$ would block γ. Hence, it is sufficient to show that no coalition of agents in $I \setminus \{i : \mu_i(\theta) \neq \emptyset\}$ would block γ. Take, then, any $d \in A^D$. By construction, d is assigned its most preferred *available* hospital at some matching in F; namely, $\mu^d_d = P_d$. Hence, d would only opt out with P_d or (possibly) some unavailable hospital. The latter is impossible, so that d could only opt out with P_d. Set $P_d = h$ and consider h's incentives to opt out. Suppose that there is a set A of doctors that satisfies q_h and is better than h's best match in F; namely, μ^h. By construction of μ^h, A must contain either an unavailable doctor or a doctor d' such that $P_{d'} \neq h$. No unavailable doctor would opt out with h. Moreover, by construction $\mu^{d'}_{d'} \succ_{d'} h$. Hence, only doctors in h's outside option would opt out with h. Since h's best match in F is strictly better than its outside option, no coalition can block γ. □

8 References

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

Abdulkadiroğlu, Atila, Pathak, Parag A, & Walters, Christopher R. 2015. *School vouchers and student achievement: First-year evidence from the Louisiana Scholarship Program.* National Bureau of Economic Research Cambridge, MA.

Afacan, Mustafa Oğuz. 2016. Enrollment manipulations in school choice. *Journal of Mathematical Economics*, **63**, 119–125.

Afacan, Mustafa Oğuz. 2018. The object allocation problem with random priorities. *Games and Economic Behavior*, **110**, 71–89.

Afacan, Mustafa Oğuz. 2019. School choice with vouchers. *Journal of Economic Theory*, **179**, 57–72.

Agarwal, Nikhi. 2017. Policy Analysis in Matching Markets. *American Economic Review*, **107**(5), 246–250.

Alkan, Ahmet, & Gale, David. 2003. Stable schedule matching under revealed preference. *Pages 3–19 of: ICM Millennium Lectures on Games.* Springer.

Ashlagi, Itai, & Roth, Alvin E. 2014. Free riding and participation in large scale, multi-hospital kidney exchange. *Theoretical Economics*, **9**(3), 817–863.

Aziz, Haris, & Klaus, Bettina. 2017. Random Matching under Priorities: Stability and No Envy Concepts. *arXiv preprint arXiv:1707.01231*.

Besstremyannaya, Galina. 2015. Heterogeneous effect of residency matching and prospective payment on labor returns and hospital scale economies. *Stanford Institute for Economic Policy Research, Discussion Paper*, 015–001.

Budish, Eric, Che, Yeon-Koo, Kojima, Fuhito, & Milgrom, Paul. 2013. Designing random allocation mechanisms: Theory and applications. *The American Economic Review*, **103**(2), 585–623.

Clark, Simon. 2006. The uniqueness of stable matchings. *Contributions in Theoretical Economics*, **6**(1), 1–28.

Eeckhout, Jan. 2000. On the uniqueness of stable marriage matchings. *Economics Letters*, **69**(1), 1–8.

Ehlers, Lars, Hafalir, Isa E, Yenmez, M Bumin, & Yildirim, Muhammed A. 2014. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, **153**, 648–683.

Ekmekci, Mehmet, & Yenmez, M Bumin. 2014. Integrating schools for centralized admissions.

Erdil, Aytek, & Ergin, Haluk. 2008. What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review*, **98**(3), 669–89.

Fragiadakis, Daniel, & Troyan, Peter. 2017. Improving matching under hard distributional constraints. *Theoretical Economics*, **12**(2), 863–908.

Frankenberg, Erica, & Lee, Chungmei. 2003. Charter Schools and Race: A Lost Opportunity for Integrated Education.

Gajendragadkar, Suhrid S. 2006. The constitutionality of racial balancing in charter schools. *Columbia Law Review*, 144–181.

Gale, David, & Shapley, Lloyd S. 1962. College admissions and the stability of marriage. *The American Mathematical Monthly*, **69**(1), 9–15.

Goto, Masahiro, Kojima, Fuhito, Kurata, Ryoji, Tamura, Akihisa, & Yokoo, Makoto. 2017. Designing matching mechanisms under general distributional constraints. *American Economic Journal: Microeconomics*, **9**(2), 226–62.

Hafalir, Isa E, Yenmez, M Bumin, & Yildirim, Muhammed A. 2013. Effective affirmative action in school choice. *Theoretical Economics*, **8**(2), 325–363.

Hara, Koji, Otsubo, Tetsuya, Kunisawa, Susumu, & Imanaka, Yuichi. 2017. Examining sufficiency and equity in the geographic distribution of physicians in Japan: a longitudinal study. *BMJ open*, **7**(3), e013922.

Kamada, Yuichiro, & Kojima, Fuhito. 2015. Efficient Matching under Distributional Constraints: Theory and Applications. *American Economic Review*, **105**(1), 67–99.

Kamada, Yuichiro, & Kojima, Fuhito. 2016b. Stability concepts in matching under distributional constraints. *Journal of Economic Theory*, **168**, 107–142.

Kamada, Yuichiro, & Kojima, Fuhito. 2017. Recent developments in matching with constraints. *American Economic Review*, **107**(5), 200–204.

Kamada, Yuichiro, & Kojima, Fuhito. 2018a. New directions of study in matching with constraints.

Kamada, Yuichiro, & Kojima, Fuhito. 2018b. Stability and strategy-proofness for matching with constraints: A necessary and sufficient condition. *Theoretical Economics*, **13**(2), 761–793.

Kesten, Onur. 2012. On two kinds of manipulation for school choice problems. *Economic Theory*, **51**(3), 677–693.

Kesten, Onur, & Ünver, M Utku. 2015. A theory of school-choice lotteries. *Theoretical Economics*, **10**(2), 543–595.

Liu, Ce. 2018. *Stability in repeated matching markets*. Tech. rept.

Liu, Qingmin, Mailath, George J, Postlewaite, Andrew, & Samuelson, Larry. 2014. Stable matching with incomplete information. *Econometrica*, **82**(2), 541–587.

Majumdar, Dipjyoti. 2003. Ordinally bayesian incentive compatible stable matching.

Manjunath, Vikram. 2017. Stability and the core of probabilistic marriage problems.

Nguyen, Thành, & Vohra, Rakesh. 2018. Near-feasible stable matchings with couples. *American Economic Review*, **108**(11), 3154–69.

Niederle, Muriel, & Roth, Alvin E. 2003. Unraveling reduces mobility in a labor market: Gastroenterology with and without a centralized match. *Journal of political Economy*, **111**(6), 1342–1352.

Niederle, Muriel, & Yariv, Leeat. 2009. *Decentralized matching with aligned preferences*. Tech. rept. National Bureau of Economic Research.

Parker, Wendy. 2001. The color of choice: Race and charter schools. *Tul. L. Rev.*, **75**, 563.

Postlewaite, Andrew. 1979. Manipulation via endowments. *The review of economic studies*, **46**(2), 255–262.

Roth, Alvin, & Sotomayor, Marilda. 1990. *TwoMSided Matching*.

Roth, Alvin E. 1984. The evolution of the labor market for medical interns and residents: a case study in game theory. *Journal of political Economy*, **92**(6), 991–1016.

Roth, Alvin E. 1986. On the allocation of residents to rural hospitals: a general property of two-sided matching markets. *Econometrica: Journal of the Econometric Society*, 425–427.

Roth, Alvin E. 2008. What have we learned from market design? *The Economic Journal*, **118**(527), 285–310.

Roth, Alvin E, & Vande Vate, John H Vande. 1990. Random paths to stability in two-sided matching. *Econometrica: Journal of the Econometric Society*, 1475–1480.

Roth, Alvin E, Rothblum, Uriel G, & Vande Vate, John H. 1993. Stable matchings, optimal assignments, and linear programming. *Mathematics of operations research*, **18**(4), 803–828.

Roth, Benjamin, & Shorrer, Ran I. 2018. Making It Safe to Use Centralized Marketplaces: Dominant Individual Rationality and Applications to Market Design. *Available at SSRN 3073027*.

Sönmez, Tayfun. 1999. Can pre-arranged matches be avoided in two-sided matching markets? *Journal of Economic theory*, **86**(1), 148–156.

Wu, QINGYUN, & Roth, Alvin E. 2016. *The lattice of envy-free matchings.* Tech. rept. Mimeo.

Yenmez, M Bumin. 2013. Incentive-compatible matching mechanisms: consistency with various stability notions. *American Economic Journal: Microeconomics*, **5**(4), 120–41.