THREE ESSAYS IN APPLIED MARKET DESIGN

BY

JUAN FUNG

DISSEIRATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Economics
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2016

Urbana, Illinois

Doctoral Committee:

Professor Steven R. Williams, Chair
Professor Daniel Bernhardt
Professor Daniel McMillen
Professor Martin Perry
Abstract

The market design approach to economics recognizes that markets do not arise naturally but are rather an amalgamation of various rules and norms. From this perspective, an economist can reverse engineer the rules that are consequential to a functioning market and evaluate the effects of those rules on market outcomes, with an eye toward potentially re-engineering certain components to achieve some objective. In this thesis, I present three market settings inspired by real-world applications, viewed broadly from the lens of market design.

The first chapter, joint work with Chia-Ling Hsu, explores the consequences of various market details on equilibrium outcomes. Specifically, we consider a situation in which a matching problem between two sets of agents is solved by a platform serving as an intermediary. For instance, an artist who wants to find donors and a backer who wants to find artists to support can find each other on a crowdfunding platform like Kickstarter. Existing models of platform markets restrict agent heterogeneity and so the matching problem is secondary. However, it is possible that different artists target different types of backers and even likely that backers differ in their preferences for artists. In this chapter, we introduce agent heterogeneity by proposing a matching model of platform markets. In such markets, stability eliminates the possibility of an individual or group of agents switching in equilibrium, thus ensuring successful coordination. The model allows exploration into the properties of equilibrium with heterogeneous agents, offering a new approach to studying platform markets.

In the second chapter, I empirically quantify the value of public school choice. Traditionally, public school assignment is determined by a family’s residence in the district. An alternative policy is to allow families to apply to any school in the district. Such school choice programs provide families with more options, but it is unclear how much families value these options over
having a guaranteed school. In this chapter, I exploit a natural experiment in Champaign-Urbana, IL: in 1998, Champaign school district adopted school choice while the neighboring district of Urbana did not. Using variation in housing prices in each district, before and after the policy change, I estimate the marginal willingness to pay for school choice relative to residence-based assignment. I find that, on average, households are willing to pay between 5-7% more for school choice relative to residence-based assignment. The results are robust to regularization and alternative model specifications.

The third chapter, joint work with Blake Riley, is motivated by decentralized matching: the process by which agents find matches on their own. We show that, without revealing information to a centralized matchmaker and without coordination, agents can find stable matches on their own. Existing work on uncoordinated matching, based on the random better reply dynamics of Roth and Vande Vate (1990), shows that agents do find stable matches but that in the worst case it could take exponentially long. We introduce a new process that, in various numerical experiments, appears to converge in polynomial time. The key to our proposal process is mitigating a major bottleneck in uncoordinated matching: the possibility that an agent is single for a very long time before finding a match. In the worst case, our process converges in $O(n^3)$ time in moderate sized balanced markets with $n$ agents on each side. We also consider unbalanced markets, in which there are more agents on one side of the market. While convergence to stability is not guaranteed in polynomial time, we show numerically that typical outcomes of our proposal process are more egalitarian than stable outcomes. This chapter thus sheds some light on the value of centralizing a matching market, as opposed to allowing the market to clear on its own.

The common thread in all three chapters is that, while markets should not be taken as given, it is important to evaluate the relative importance of particular design elements. The first chapter characterizes equilibrium outcomes under various designs; the second considers the relative value of two particular designs; and the third questions the value of designing at all. In the spirit of market design, each application is driven by actual markets and a variety of methodologies.
To my wife, Kaitlin Straker, for her love and support.
Acknowledgments

I am indebted to Chia-Ling Hsu, Kwanghyun Kim, Blake Riley, Tom Sahajdack, Rachel Shafer, and many more colleagues. I also thank Dan Bernhardt, Song-Hyun Hong, Jorge Lemus, Dan McMillen, and Marty Perry for listening and for feedback. I am especially grateful to Steve Williams for many wonderful conversations, support, and friendship.
# Table of Contents

List of Tables ........................................ vii
List of Figures ....................................... viii

Chapter 1  Matching in Platform Markets  
with Chia-Ling Hsu ...................................... 1  
1.1 Introduction ........................................ 1  
1.2 A matching model of platform markets .............. 6  
1.3 Main results: Properties of equilibrium allocations . 14  
1.4 Concluding remarks ................................ 20

Chapter 2  Sorting Under Public School Choice .......... 21  
2.1 Introduction ........................................ 21  
2.2 Public school assignment policies in Champaign-Urbana, IL 24  
2.3 Empirical framework ................................ 27  
2.4 Main results ........................................ 30  
2.5 Conclusion ......................................... 35

Chapter 3  Uncoordinated One-to-One Matching Markets  
with Blake Riley ........................................ 36  
3.1 Introduction ........................................ 36  
3.2 Model .............................................. 38  
3.3 Computational results ................................ 44  
3.4 Conclusion ......................................... 49

Appendix A ............................................. 51  
A.1 Proofs ............................................. 51

Appendix B ............................................. 56  
B.1 Figures ............................................ 56  
B.2 Tables ............................................. 61

Appendix C ............................................. 70  
C.1 Figures ............................................ 70  
C.2 Tables ............................................. 78

References ............................................. 79
List of Tables

B.1 Descriptive statistics by school district: housing . . . . . . . . . 62
B.2 Descriptive statistics by school district: Census tract level . . . 63
B.3 Descriptive statistics by school district: schools . . . . . . . . . 64
B.4 Number of observations by “treatment” and “control” . . . . . . 65
B.5 Estimates for DD hedonic model . . . . . . . . . . . . . . . . . . . 66
B.6 Estimates for mixed-effects DD hedonic model . . . . . . . . . . 67
B.7 Estimates for regularized DD hedonic model . . . . . . . . . . . 69

C.1 An instance of hard preferences in balanced markets, repro- 
produced from Ackermann et al. (2011). . . . . . . . . . . . . . . 78
## List of Figures

1.1 Artists and backers choose platforms ........................................ 2
1.2 Artists and backers choose contracts through a platform ............ 4
1.3 Access versus Interaction: $b_2$ has access to $a_3$ but only interacts with $a_1, a_2$. ......................................................... 14

B.1 Champaign and Urbana attendance areas, circa 1998. Points represent schools. ................................................................. 56
B.2 Champaign attendance areas, circa 1989-1998. Points represent schools corresponding to attendance area. .................. 57
B.3 Urbana attendance areas, circa 1989-2002. Points represent schools corresponding to attendance area. .................. 58
B.4 Timeline of assignment policy changes ..................................... 58
B.5 Distribution of log sale price, before (left) and after (right) the policy change. ................................................................. 59
B.6 Median log sale price over time. ................................................. 60

C.1 Number of proposals to reach stability in Random Best Attainable process in balanced random environment .......... 70
C.2 Number of proposals to reach stability in Random Best Attainable process in almost balanced random environment .... 71
C.3 Number of proposals to reach stability in Random Best Attainable process in hard environment .............................. 71
C.4 Number of proposals to reach stability in Random Singles First process in balanced random environment ............... 72
C.5 Number of proposals to reach stability in Random Singles First process in almost balanced random environment .... 72
C.6 Number of proposals to reach stability in Random Singles First process in hard environment .............................. 73
C.7 Proportion single after multiple of proposal in almost balanced environment for $n = 50$. ......................................................... 73
C.8 Proportion of single agents after multiple of proposal in balanced environment for $n = 50$. ......................................................... 74
C.9 Proportion of single agents after multiple of proposal in almost balanced environment for $n = 1000$. ........................................ 74
C.10 Average ranks for $n$ men and $n + 1$ women after $5n^2$ proposals. 75
C.11 Men’s average rank relative to the WOSM with \( n \) men and 
\( n + 1 \) women after \( 5n^2 \) proposals. . . . . . . . . . . . . . . . . . 75
C.12 Total average rank relative to the WOSM with \( n \) men and 
\( n + 1 \) women after \( 5n^2 \) proposals. . . . . . . . . . . . . . . . . . 76
C.13 Men’s average rank relative to the WOSM with \( n \) men and 
1.5\( n \) women after \( 5n^2 \) proposals. . . . . . . . . . . . . . . . . . 76
C.14 Total average rank relative to the WOSM with \( n \) men and 
1.5\( n \) women after \( 5n^2 \) proposals. . . . . . . . . . . . . . . . . . 77
1.1 Introduction

Imagine an artist who would like to raise money for a large public exhibition. How should the artist go about finding backers? He might consider outreach through local galleries, filing grant applications, or perhaps, if it’s the 16th century, asking his local monarch. Such fundraising is challenging even for an established artist. On the other hand, imagine a patron of the arts who would like to fund a promising new artist. How does this individual find an artist to fund? For the very few “in the know,” there is a well-informed art community. However, the art community may not be available to an “outsider,” nor is it likely to be aware of every new artist.

The preceding scenario presents a matching problem: artists and backers would like to get together. A solution to this matching problem has emerged in the form of crowdfunding. Platforms such as Kickstarter and IndieGoGo allow artists to present their fundraising campaign to a wide audience of potential backers, eliciting donations in exchange for rewards. For example, the artist might offer each backer an autographed sketch of the exhibition. If the artist’s campaign is successful, the crowdfunding platform receives a percentage of the funds raised. In this way, platforms solve the matching problem between artists and backers.

Existing models of platform markets, however, ignore the matching aspect. The standard approach focuses on pricing strategies by monopoly platforms or price competition between competing platforms and the resulting allocations of agents across platforms; see Rysman (2009) and Wright (2004) for good overviews. In these models, agent heterogeneity is limited and thus

\footnote{Other examples of such markets include Airbnb, matching hosts and travelers, and Coursera, matching college instructors and students.}
the matching problem is trivial. Platforms set prices and agents choose platforms, as illustrated in Figure 1.1. Such models have provided much insight into platform behavior while taking agent behavior as given. In many platform markets, this is a close approximation to reality (e.g., ride-sharing services like Uber, online marketplaces like eBay, and the modern advertising business as exemplified by Facebook and Google). If, however, agents are heterogeneous in their preferences for agents on the other side of the market, as in the matching between artists and backers, then it is natural to ask how heterogeneity affects market outcomes.

Figure 1.1: Artists and backers choose platforms

We model platform markets as a matching problem in order to study agent heterogeneity. In doing so, we address three important questions. The first is how to model matching in platform markets. Second, in a matching model of platform markets, what is a reasonable notion of equilibrium? With an equilibrium concept in hand, we then explore how agent heterogeneity affects properties of equilibrium.

Our model of matching in platform markets has three important features. The key feature is that we allow for arbitrary heterogeneity in preferences for platforms and for other agents. An artist may want to raise money from as many backers as humanly possible, or he might be targeting a very specific segment of like-minded patrons. Potential donors are even more likely to be heterogeneous in how they value artists. In contrast, existing models of platform markets restrict heterogeneity in preferences for other agents. Rochet and Tirole (2003) model agents who differ in how they value the number of agents from the other side, but the identities of the agents do

A second feature of our model is that we allow for an arbitrary number of platforms, but there is no price competition. Thus, we focus on agent behavior while taking platform behavior as given. In many matching problems, prices are of second-order importance, in the sense that prices are not sufficient to clear the market. For instance, Kickstarter takes a 5% cut of a successful campaign while IndieGoGo only takes 4%. Nevertheless, Kickstarter is focused on creative projects while IndieGoGo allows for a broader range of campaigns, such as charities and entrepreneurs. To an artist attempting to reach patrons of the arts, this distinction is likely to matter more than the price. Rochet and Tirole (2006) model the monopoly case, while Caillaud and Jullien (2003), Rochet and Tirole (2003), and Armstrong (2006) consider both monopoly and duopoly cases. Weyl (2010) also models a monopoly platform, but the platform chooses participation rates rather than prices. White and Weyl (2012) extend this to the oligopoly case. In all of these models, the focus is on platform behavior.\footnote{Exceptions include Lee (2014), who models a bargaining game between platforms and one side of the market, and Ambrus and Argenziano (2009), who model a dynamic game between platforms and agents. In all of these, heterogeneity is limited to how agents value the number on the other side.}

Third, we assume single-homing: each agent chooses a single platform. An artist must have his campaign approved on Kickstarter and, once approved, must actively engage with potential backers in order to reach his target. In addition to these costs, an artist is more likely to succeed on a single platform than by campaigning on many platforms. Backers, on the other hand, do not pay the platform to contribute. However, once a backer is signed on to Kickstarter, it is easier both to find additional artists through the platform and to contribute to those artists as all payment information is already in the system. The homing decision in existing models is typically exogenous, as in the classic models of Caillaud and Jullien (2003), Rochet and Tirole (2003) and Armstrong (2006).\footnote{Roson (2005) and Jeitschko and Tremblay (2015) endogenize the homing decision.}

We build a model with these features as a special case of the matching with contracts framework of Hatfield and Milgrom (2005).\footnote{In particular, our model is one of unitary many-to-many matching with contracts; see,}
contracts allows for a rich set of potential matching relationships. Agents not only choose one another but must specify the terms of their relationship. For example, Alice the Artist sets a funding goal of $10,000, promising Bob the Backer a signed sketch for his $100 contribution, and sets a 90 day deadline to complete her project. A contract summarizes this match. Alice can choose to offer this contract through Kickstarter or through IndieGoGo. At the same time, backers choose which artists they want to support and the platform through which they will donate. This is illustrated in Figure 1.2, where the links represent the contracts.

The natural next question for the model is how to define equilibrium. In equilibrium, Alice and her potential backers must coordinate successfully on a single platform through their choices of contracts. The central solution concept for matching markets is stability: no individual or group of agents can find a better match. This property is desirable for successful coordination in platform markets as it eliminates the possibility of individuals or groups of agents switching platforms. Moreover, stable allocations are guaranteed to exist under general conditions.\(^5\) We thus adopt stability as our equilibrium concept and show that the structure of stable matching can provide insight into qualitative features of market structure in platform markets.

Figure 1.2: Artists and backers choose contracts through a platform

First, we consider properties of equilibrium when contract terms can be ordered, like membership fees. In particular, contracts offer either a high

\(^5\)We assume the standard substitutes condition for preferences, which is roughly that contracts are not complementary for agents. However, this condition can be relaxed; see Hatfield and Kominers (2015); Yenmez (2015).
fee or a low fee, with the low fee preferred to the high fee all else equal. A well-known property of stable matching is the “opposition of interests”: there is a stable match that all artists prefer least, and all backers prefer most, from the set of stable matchings. At this matching, we show that more artists pay the high fee, and more backers pay the low fee, than at any other stable matching. This result parallels the “seesaw” (or divide-and-conquer) principle from the literature on platform markets (Caillaud and Jullien (2003); Rochet and Tirole (2006)). In that literature, the result is driven by a platform’s desire to subsidize one side in order to attract the other. In our model, the result arises from the opposition of interests at stability: one side compromises in order to match with the other.

Second, we consider conditions under which the market “tips” in favor of a dominant platform. Market tipping concerns the equilibrium balance of market power across platforms. The competition for agents often results in one or a handful of platforms dominating the market.\textsuperscript{6} We derive results on market tipping using the structure of stable matching. In particular, we introduce a condition that allows platforms to be differentiated with respect to contract terms and thus capture the market. For simplicity, suppose all backers prefer to fund musical projects over other types of art projects. Under this condition, agents match assortatively with respect to contract terms. All backers will sort into one platform to fund musical projects, and all musicians will sort into the same platform.

Third, we consider the effect of restricting contract terms on outcomes. In particular, suppose all artists on Kickstarter must not only pay the same percentage of funds raised, but must also provide the same rewards to backers. Moreover, suppose all backers continue to donate on Kickstarter for free but must also donate the same amount. While the example is a bit extreme, it is not difficult to imagine regulation in platform markets requiring such “non-discriminatory” contract terms. Unsurprisingly, this restriction results in coordination failure: a stable allocation is not guaranteed to exist. With fully flexible contract terms, Alice the Artist and Bob the Backer can easily negotiate on a common platform. This essentially removes the platform from the intermediary role as in Figure 1.2.\textsuperscript{7} With constrained contract

\textsuperscript{6}See Rysman (2009) for a discussion. Ellison and Fudenberg (2003) and Weyl and White (2014) argue that market tipping is a “knife-edge” outcome rather than a generic property of equilibrium.

\textsuperscript{7} We adopt the bipartite structure represented by Figure 1.2 as a direct analogue to a
terms, Alice and Bob cannot negotiate on a common platform because their contract terms will be constrained by others joining the platform. In essence, this returns the platform to the intermediary role as in Figure 1.1, breaking the two-sided structure. We thus propose a constrained version of stability and show existence under the additional assumption of lexicographic preferences. 8 Lexicographic preferences restore the bipartite structure in Figure 1.2: finding backers is of first-order importance to Alice, making negotiations over contract terms of second-order importance. 9 Our proof is constructive, based on Hatfield and Milgrom (2005)’s cumulative offer algorithm. 10

Our matching model offers a complementary approach to analyzing platform markets that allows for fully heterogeneous agents. By focusing on agents rather than platforms, we can derive qualitative properties of the equilibrium market structure while abstracting away from price competition. As Weyl (2010) points out, there is a gulf between the standard platform markets literature and the “matching design” literature, even though “[t]hese literatures have much in common.” It is our hope that this paper provides a first step towards a synthesis of the two approaches.

The paper is structured as follows. Section 1.2 presents the matching model of platform markets. The main results are derived in section 1.3, with proofs in Appendix A.1. Section 2.5 offers some concluding remarks.

### 1.2 A matching model of platform markets

In this section, we construct our matching model of platform markets. We propose stable matching as the natural equilibrium concept: at a stable match, no individual or group of agents can find a better match. Stability

---

8In our paper, artists are lexicographic in the sense that preferences over backers are of first-order importance, followed by preferences over platforms and contract terms. However, if preferences over platforms are of first-order importance, for instance, our results still hold, so long as all artists are lexicographic in this way.

9Our result is in the spirit of Dutta and Massó (1997), who show that when artists have preferences over the other artists in a match, the two-sided structure is broken and a matching is not guaranteed to exist. If, however, artists are lexicographic in the sense that preferences over backers are of first-order importance, then a stable matching exists.

10See Sönmez (2013) and Sönmez and Switzer (2013) for recent applications of the cumulative offer algorithm to real-world market design problems.
guarantees agents do not switch platforms in equilibrium and is thus essential for successful coordination in platform markets.

The following notation is used for agents and platforms. Let $S = A \cup B$ denote the finite set of agents in the market, and $I$ the finite set of platforms. Let $a$, $b$, and $s$ denote generic agents in the sets $A$, $B$, and $S$, respectively, and let $i$ denote a generic platform in the set $I$. Agents from $A$ can only interact with agents from $B$ through a platform in $I$.

We assume that each agent can join at most one platform. In the language of two-sided markets, each side is single-homing. Moreover, an agent can choose not to participate in any platform. In this case, the agent does not interact with anyone on the other side.

In addition to choosing matching partners and platforms, agents must agree to contract terms. When an agent $s \in S$ joins a platform, she may be required to agree to terms in a contract. The terms we consider here are rather broad. For instance, they may include Alice’s funding target, the percentage of funds that Alice pays to Kickstarter, and the rewards Alice promises Bob for a $100 donation. We call $t_s$ the contract terms, or simply terms, imposed on agent $s$. Let $T \equiv T_A \times T_B$ denote the finite set of possible terms, where $T_A$ are the terms for side $A$ and $T_B$ are the terms for side $B$. The sets $T_A$ and $T_B$ may be identical, disjoint, or have elements in common. We assume that the set of available terms is the same for all platforms.

### 1.2.1 Platforms as objects of choice

In this subsection, we discuss the role that platforms play in the matching between agents. Platforms are objects of choice and agents must agree on a common platform in order to interact (by the single-homing assumption).

We treat platforms as objects of choice: platforms do not have preferences. Consider an agent’s choice of platform. When an agent chooses a platform directly, she indirectly chooses agents on the other side. We make this implicit choice explicit by allowing agents to choose each other. Moreover, the agents must choose the same platform. Thus, each agent chooses a bundle of agents from the other side of the market where the bundle is indexed by a platform and contract terms (e.g., membership fees).

An agent’s choices are embedded in a contract. A contract $x$ associates
a pair of agents from each side of the market with a platform and, possibly, contract terms. Formally, let $X$ be the set of all contracts. A typical contract $x \in X$ between agents $a \in A$ and $b \in B$ takes the form

$$x = (a, b, i, t_a, t_b) \in X,$$

where $X \equiv S \times I \times T$.

Thus, we are in a many-to-many matching with contracts setting. The following notation is standard in matching with contracts. For a contract $x \in X$, let $x^A \in A$ denote the side $A$ agent, $x^B \in B$ denote the side $B$ agent, $x^I \in I$ denote the platform associated with contract $x$, and $x^T = (x^T_A, x^T_B) \in T_A \times T_B$ denote the pair of terms associated with contract $x$. Agents may sign any number of contracts in general but may not sign multiple contracts with the same agent. Kominers (2012) calls such environments unitary. Moreover, agents are free to choose an outside option instead of joining any platform, represented by the null contract, $\emptyset$.

The following notation will also be useful. For a given set of contracts $Z \subseteq X$, let $Z^a \subseteq Z$ denote the contracts in $Z$ associated with agent $a \in A$. Analogous definitions hold for $Z^b \subseteq Z$ and $Z^i \subseteq Z$. Also, let $Z^A \subseteq A$ denote the set of side $A$ agents associated with $Z$. Analogous definitions hold for $Z^B \subseteq B$, $Z^I \subseteq I$, and $Z^T = Z^T_A \times Z^T_B \subseteq T_A \times T_B$. Finally, we write $(Z^a)_B \subseteq B$ to indicate the side $B$ agents with contracts in $Z^a$. A similar definition applies for $(Z^a)_I \subseteq I$, $(Z^a)_T \subseteq T_A \times T_B$, and so on.

We define feasibility with respect to the unitary and single-homing conditions. Consider an agent $a \in A$. If $a$ signs multiple contracts then by single-homing, the platforms must be the same across all those contracts. Moreover, each contract must represent a partnership with a unique agent on the other side.

Formally, a set of contracts $Z \subseteq X$ is feasible if

1. $\forall a \in A$: (i) $z_B = z'_B \Rightarrow z = z', \forall z, z' \in Z^a$, and (ii) $(Z^a)_I = \{i_a\}$

2. $\forall b \in B$: (i) $z_A = z'_A \Rightarrow z = z', \forall z, z' \in Z^b$, and (ii) $(Z^b)_I = \{i_b\}$

In other words, $Z$ is feasible if each agent with contracts in $Z$ holds contracts that are (i) unitary and (ii) single-homing.

To illustrate feasibility, consider the following simple example. Let agent $a \in A$ and suppose $B = \{b_1, b_2\}$, $I = \{i_1, i_2\}$, and the terms are $T_A = \{t_L, t_H\}$.
and $T_B = \emptyset$. The set $Z = \{(a, b_1, i_1, t_L), (a, b_1, i_1, t_H)\}$ is not unitary for any agent but satisfies single-homing for both agents, while the set $Z' = \{(a, b_1, i_1, t_L), (a, b_2, i_2, t_L)\}$ is unitary for both agents but violates single-homing for $a$. On the other hand, the set $Z^* = \{(a, b_1, i_2, t_L), (a, b_2, i_2, t_L)\}$ is feasible.

### 1.2.2 Preferences of agents

In this subsection, we discuss the assumptions on agent preferences used throughout the paper. They key assumption is substitutability, which is central to the existence of stable allocations.

Our construction of preferences is standard. Let $P_s$ denote agent $s$'s preference relation over sets of contracts. For ease of analysis, we assume this ordering is strict. Note that, for an agent $s$ choosing between two contracts, $x$ and $x'$, that only differ with respect to the terms for the agent on the other side, $s$ is likely to be indifferent. In order to have $P_s$ be strict, we assume $s$ breaks ties between $x, x'$ arbitrarily. Let $P \equiv (P_s)_{s \in S}$ denote the profile of strict preferences over sets contracts.

Rather than working with the preference relation directly, it is standard to focus on an agent’s choices. For a given set of available contracts $Z \subseteq X$, agent $s$’s choice set is defined as

$$C_s(Z) \equiv \{Z' \subseteq Z : Z' P_s Z'', \forall Z'' \subseteq Z \text{ feasible and } Z' \text{ feasible}\}$$

In other words, when agent $s \in A \cup B$ is offered a set of contracts $Z$, she chooses the feasible subset from $Z$ that she most prefers with respect to her preferences, $P_s$.

It is also useful to define an agent’s non-choices. Agent $s$’s rejected set with respect to $Z \subseteq X$ is

$$R_s(Z) \equiv Z \setminus C_s(Z).$$

Let $C_A(Z) \equiv \cup_{a \in A} C_a(Z)$ and $C_B(Z) \equiv \cup_{b \in B} C_b(Z)$ denote the set of contracts from $Z$ chosen by all $a \in A$ and all $b \in B$, respectively. The rejected sets for each side are then $R_A(Z) \equiv Z \setminus C_A(Z)$ and $R_B(Z) \equiv Z \setminus C_B(Z)$.

In order to guarantee existence of a stable allocation, it is standard in
the matching literature to assume preferences satisfy a substitutes condition (Kelso and Crawford (1982)). Preferences over contracts for agent \( s \in S \) are substitutable if

\[ Z \subseteq Z' \subseteq X \Rightarrow R_s(Z) \subseteq R_s(Z'). \]

In other words, new offers are not influenced by previously rejected offers. An equivalent definition is that \( z \in C_s(Y \cup \{x, z\}) \Rightarrow z \in C_s(Y \cup \{z\}), \forall x, z, \) and \( Y \subseteq X \). In other words, the choice of a new contract, \( z \), does not depend on the availability of another contract, \( x \).

In many matching problems, it is natural for agents to exhibit some sort of monotonicity with respect to match offers. Following Hatfield and Milgrom (2005) for many-to-one matching markets and Hatfield and Kominers (2015) for many-to-many matching markets, we say that preferences satisfy the law of aggregate demand, if for agent \( s \in S \), and \( X' \subseteq X'' \subseteq X \), we have \( |C_s(X')| \leq |C_s(X'')| \). In other words, more contracts are better. Intuitively, if more contracts become available, agent \( s \) demands weakly more contracts.

The law of aggregate demand has a natural interpretation in two-sided markets: that agents benefit purely by having more agents to interact with. In section 1.3.2, we introduce an extension of the law of aggregate demand that is defined only with respect to the subscription terms.

We introduce a class of preferences that is particularly relevant to platform markets: the class of lexicographic preferences. We say that preferences are lexicographic if agents consider the desirability of contracts in a particular order with respect to the components of the contracts: the agents from the other side, the choice of platform, and the subscription terms. In particular, we consider an environment where each agent first considers the set of agents on the other side, followed by the platform, and finally the terms.\(^{11}\)

**Definition 1.1.** Consider an agent \( a \in A \). Preferences of \( a \) are lexicographic if whenever any of the following conditions hold,

1. \( B_1 = (Y^a_1)_B = (\tilde{Y}^a_1)_B \neq (Y^a_2)_B = (\tilde{Y}^a_2)_B = B_2; \)
2. \( B_1 = B_2 \) and \( i_1 = (Y^a_1)_I = (\tilde{Y}^a_1)_I \neq (Y^a_2)_I = (\tilde{Y}^a_2)_I = i_2, \)
3. \((B_1, i_1) = (B_2, i_2)\) and \( t_1 = (Y^a_1)_T = (\tilde{Y}^a_1)_T \neq (Y^a_2)_T = (\tilde{Y}^a_2)_T = t_2. \)

\(^{11}\)Our results also hold if we define the lexicographic ordering with platforms as most important, followed by the agents from the other side, and finally the contract terms.
for any sets $Y_1, Y_2, \tilde{Y}_1, \tilde{Y}_2 \subseteq X$, then

$$C_a(Y_1 \cup Y_2) = Y_1 \quad \Rightarrow \quad C_a(\tilde{Y}_1 \cup \tilde{Y}_2) = \tilde{Y}_1. \quad (1.1)$$

Moreover, for any $\hat{Y} \subseteq X$ such that $(\hat{Y}^a)_{B,I} = (\hat{B}, \hat{i})$,

$$C_a(\hat{Y})_{B,I} = (\hat{B}, \hat{i}) \Rightarrow C_a(Y)_{B,I} = (\hat{B}, \hat{i}), \quad \forall Y \subseteq X \text{ such that } (Y^a)_{B,I} = (\hat{B}, \hat{i}). \quad (1.2)$$

The first equation requires that an agent evaluates any set of contracts lexicographically, first in terms of the set of agents on the other side, followed by the platform, and finally by the contract terms. The second equation requires that if an agent finds a $(\hat{B}, \hat{i})$-combination acceptable at some set of contracts $\hat{Y}$, then the agent finds this combination acceptable at any set of contracts $Y$; that is, at any $t \in T$.

Note that it is possible that $C_a(Y_1 \cup Y_2) = \emptyset$ in equation (1.1). In particular, we do not require that every $(B', i', t')$ combination be acceptable to $a$, only that any $t' \in T$ is acceptable.

Lexicographic preferences may be relevant in some settings. Consider the example of crowdfunding. It is easy to imagine backers whose first-order concern is finding musicians they like, with the crowdfunding platform and match terms secondary. On the other hand, a rock band might be primarily concerned with reaching its core fans than with the platform through which it reaches those fans. Lexicographic preferences play an important role in section 1.3.3 when stable allocations have an additional constraint.

1.2.3 Stable allocations

A desirable property for successful coordination in equilibrium is that agents are not constantly switching, be it their platform, subscription terms, or match partners. Stability captures this very notion: no individual or group of agents mutually prefers another allocation. Thus, it is a reasonable criterion for allocations in platform markets.

Formally, an allocation $Y \subseteq X$ is simply a feasible set of contracts. That is, each agent $s \in S$ signs unitary contracts associated with the same platform. Preferences over allocations correspond directly to the underlying
preferences over sets of contracts.

An allocation is stable if there is no unilateral or multilateral switching by agents. Our notion of stability follows from the notion of weak setwise stability in the matching with contracts literature (Klaus and Walzl (2009), Hatfield and Kominers (2015)). Formally, an allocation \( Y \) is \textbf{stable} if the following holds:

1. **Individual rationality**: \( C_s(Y) = Y^s \), for \( s \in S \).

2. **No blocking set of contracts**: \( \nexists Y' \subseteq X \) such that \( Y^s \subseteq C_s(Y \cup Y') \), for all \( s \in (Y)_S \).

Individual rationality rules out unilateral switching in equilibrium. Individual rationality requires that an agent \( s \) voluntarily participates in allocation \( Y \) by accepting all available contracts in \( Y \). Intuitively, an agent \( s \) who drops some or all of the contracts in \( Y^s \) is dissatisfied with some of the agents on the other side or his platform.

Blocking sets capture multilateral switching. A set \( Y' \) is a blocking set for \( Y \) if some or all of the contracts in \( Y \) are replaced by \( Y' \) for some group of agents. Intuitively, agents on opposite sides of a platform mutually agree to switch to another platform, to update their subscription terms, or to swap match partners. This cannot occur at a stable allocation.

Substitutes guarantees existence of stable allocations, and the set of stable allocations has a unique structure:

**Result 1.1** (Hatfield and Kominers (2015)). \textit{When preferences are substitutable, the set of stable allocations is nonempty and forms a lattice.}

The lattice structure implies that there exists a partial ordering over stable allocations, where stable allocations more preferred by one side are less preferred by the other side. Moreover, the extreme points of the lattice are optimal in this sense.

In particular, there exists a stable allocation that side \( A \) agents prefer to all other stable allocations, the \textbf{\( A \)-optimal stable allocation}, and this allocation is the least preferred stable allocation for side \( B \) agents, the \textbf{\( B \)-pessimal stable allocation}. Similarly, the most preferred stable allocation for side \( B \) agents, the \textbf{\( B \)-optimal stable allocation}, is the least preferred stable allocation for side \( A \) agents, the \textbf{\( A \)-pessimal stable allocation}. 

1.2.4 Access versus interaction at stable allocations

It is worth noting that stability does not require that all agents who join the same platform sign contracts. Thus, Bob is free to join Kickstarter and donate to Alice without also having to donate to Carol the Composer. Nevertheless, Bob gains access to Carol through his choice of platform without having to commit to a relationship. This is obviously a desirable feature for successful coordination and it is consistent with existing models of platform markets; see Rochet and Tirole (2006); Weyl (2010). In fact, as the following example illustrates, if we do require that all agents who join the same platform must also interact then a stable allocation may not exist. This is not surprising since such a requirement essentially forces agents into a match, and this cannot be stable in general.

**Example 1.1.** Let $A = \{a_1, a_2, a_3\}, B = \{b_1, b_2\}, I = \{i\}, T = \emptyset$, so that contracts only specify a platform. Consider the following contracts,

\[
\begin{align*}
    x_1 &= (a_1, b_1, i) & x_2 &= (a_1, b_2, i) \\
    y_1 &= (a_2, b_1, i) & y_2 &= (a_2, b_2, i) \\
    z_1 &= (a_3, b_1, i) & z_2 &= (a_3, b_2, i)
\end{align*}
\]

and suppose preferences over sets of these contracts are as follows:

\[
\begin{align*}
    P_{a_1} : \{x_1, x_2\}, \{x_1\}, \{x_2\} \\
    P_{a_2} : \{y_1, y_2\}, \{y_1\}, \{y_2\} \\
    P_{a_3} : \{z_1, z_2\}, \{z_1\}, \{z_2\} \\
    P_{b_1} : \{x_1, y_1, z_1\}, \{x_1, y_1\}, \{x_1, z_1\}, \{y_1, z_1\}, \{x_1\}, \{y_1\}, \{z_1\} \\
    P_{b_2} : \{x_2, y_2\}, \{x_2\}, \{y_2\}
\end{align*}
\]

Note that preferences are substitutable.

The $A$-optimal stable allocation is $Y = \{x_1, y_1, z_1, x_2, y_2\}$. Note that at $Y$, $b_1$ is linked to $a_1, a_2$, and $a_3$ through $i$, while $b_2$ only links to $a_1$ and $a_2$. However, $b_2$ has access to $a_3$ via platform $i$. Thus at a stable allocation, agents may share a platform without actually interacting. See Figure 1.3.

Now, suppose we require full interaction at stable allocations. Then $b_2$ does not want to interact with $a_3$, and so $a_3$ cannot be part of an allocation. However, $a_3$ and $b_1$ mutually prefer interacting to not interacting, blocking full-interaction stability.
1.3 Main results: Properties of equilibrium allocations

In this section, we present the main results. We provide conditions under which qualitative features of the equilibrium platform market structure are reflected in properties of stable matching. Our results suggest that agent heterogeneity may be an important component of platform market equilibrium.

1.3.1 The seesaw principle reflected in stable matching

In this subsection, we show that the opposition of interests property of stable allocations implies an analogue of the seesaw principle in platform markets.

The seesaw principle in platform markets is the observation that platforms may treat each side differently with respect to pricing. Rochet and Tirole (2006) observe a tendency for all agents on one side to pay the low price while agents on the other side pay the high price. Consider the example of crowdfunding. Suppose the only relevant contract terms for artists are the fees paid to the platform, a high fee and a low fee, and suppose the only relevant contract terms for backers are the cost to contribute, a high cost and a low cost. The platform might allow backers to donate at the

\footnote{We can imagine a setting in which the platform is more of a matchmaker, with agents arranging terms of their relationship outside of the match. For instance, suppose backers are indifferent between a fixed number of possible donations and prizes, which are chosen...}
low cost, creating a critical mass of potential donors and thus attracting artists who will happily pay the high fee. In the platform markets literature, this outcome arises as a result of the platform’s incentives. The intuition is that platforms may subsidize one side in order to attract the other and thus capture a larger overall market share. This result is sometimes called the “seesaw” (or divide-and-conquer) principle.

In Theorem 1.2, we show that a similar phenomenon arises even if we ignore the platform’s incentives. Suppose the set of subscription terms is \( T_A = T_B = \{t_L, t_H\} \), where \( t_L < t_H \). We may interpret such ordered terms simply as “fees” and assume that all agents prefer low fees \( t_L \) to high fees \( t_H \), all else equal. The important point is that the contract terms are ordered. Theorem 1.2 states that when preferences are substitutable, the number of side-A agents who sign the low fee \( t_L \) at the \( A \)-optimal stable allocation is weakly larger than in any other stable allocation. The opposite holds for the side-B agents: the number of side-B agents who sign the low fee \( t_L \) at the \( A \)-optimal stable allocation is weakly smaller than in any other stable allocation. The result is a direct consequence of the opposition of interests property of stable allocations: the \( A \)-optimal stable allocation is the \( B \)-pessimal stable allocation.

**Theorem 1.2.** Suppose preferences are substitutable. Then at the \( A \)-optimal stable allocation,

1. the number of side-A agents that sign \( t_L \)-contracts is weakly higher than at any stable allocation, and

2. the number of side-B agents that sign \( t_H \)-contracts is weakly higher than at any stable allocation.

By symmetry, the \( B \)-optimal stable allocation has a similar property. The number of side-A agents that sign \( t_H \)-contracts is weakly higher at the \( B \)-optimal stable allocation than at any other stable allocation, and so on. The proof, in Appendix A.1.1, exploits the lattice structure and in particular side optimality.

In the example of crowdfunding, the result can be interpreted as saying that at the backer-optimal stable allocation more backers join for free and more artists pay the high membership fee than at any other stable allocation. at random by the platform after the agents have been matched.
While our environment is too simple to reflect reality, this property seems to be reflected in practice, as donating is always free for backers.

1.3.2 Market tipping under stable allocations

In this section, we ask whether platform markets tip in equilibrium. Market tipping occurs when agents concentrate on a few platforms, shifting the balance of market power. We present a condition under which the market tips toward platforms that are differentiated in the terms they offer.

In some situations, an agent may prefer to join a platform that draws agents from the other side with similar interests or attributes. Consider the example of crowdfunding platforms. While IndieGoGo welcomes artists, entrepreneurs, and charities, Kickstarter’s focus is on creative projects. Kickstarter’s definition of creative, however, still encompasses a broad range of categories, from art and music to games and technology.\(^{13}\) A musician looking for backers could also turn to ArtistShare, which is focused exclusively on musicians.\(^{14}\) If musicians prefer to target music lovers, and music lovers are only interested in funding musicians, then it is possible that all musicians and music lovers will join a single platform. In our model, this occurs exactly when all agents prefer to join platforms that are differentiated with respect to the contract terms they offer. Thus, it is not necessarily the case that a single platform dominates the entire market but rather that platforms become niches which capture particular market segments and dominate within that segment, such as crowdfunding for music.

Formally, let \( T = T_A \times T_B \) be general sets of terms as defined in the beginning of Section 1.2. Thus, we do not assume contract terms are ordered. However, we will now assume preferences satisfy the following condition.

**Definition 1.2.** We say that preferences satisfy the law of aggregate demand for similar terms, if for any agent \( a \) and contract term \( t_a \) and two sets of contracts \( Y \) and \( Y' \), we have the following.

\[
|\{x \in Y : x_{TB} = t_a\}| \geq |\{x \in Y' : x_{TB} = t_a\}| \Rightarrow Y R_a Y';
\]
\[
|\{x \in Y : x_{TB} = t_a\}| > |\{x \in Y' : x_{TB} = t_a\}| \Rightarrow Y P_a Y'.
\]

\(^{13}\)See https://www.indiegogo.com/explore and https://www.kickstarter.com/discover.

\(^{14}\)See http://www.artistshare.com/v4/About.
Roughly, the condition states that agents match assortatively with respect to contract terms: musicians prefer $Y$ to $Y'$ because it has more donors willing to fund musical projects. We can define an analogous condition for backers. Under this condition, we have the following result:

**Theorem 1.3.** Suppose preferences satisfy substitutes and the law of aggregate demand for similar terms. Let $Y$ be a stable allocation and consider a term $t \in Y_T$. Then all agents with term $t$ are associated with one platform.

The proof is in Appendix A.1.2. Substitutes is needed for existence, while the second condition guarantees agents sort into differentiated platforms in equilibrium.

### 1.3.3 Coordination failure under term-constrained stable allocations

In this subsection, we consider an environment in which agents must be treated symmetrically with respect to contract terms. Under this constraint, agents have limited flexibility in expressing a match relationship and thus stable allocations may not exist. We provide a remedy to this problem under the additional condition of lexicographic preferences.

Note that our model allows two artists to join the same crowdfunding platform under different contract terms. This is not problematic when thinking, for instance, of the rewards Alice offers potential backers versus the rewards Carol offers. If, however, the only relevant contract term in the model is the fee Kickstarter charges artists, then it is difficult to justify Alice and Carol paying different fees for the same service. In the literature on platform markets, this would be a case of price discrimination. However, since we abstract away from price competition we can imagine a scenario in which a regulator requires that Kickstarter offers all artists the same contract terms.

The example has two important features that distinguish it from the previous examples. First, there is an additional restriction on an allocation that requires all artists joining the same platform to offer the same campaign and pay the same fee. This restriction is likely to result in coordination failure: it is more difficult for artists to entice backers under limited contract terms. Thus, a stable allocation is not guaranteed to exist. This occurs because the constraint on contract terms means artists and backers rely more heavily on
the platform as an intermediary, since they lack the flexibility to express their match relationship. With the platform in the intermediary role, we lose the bipartite structure in Figure 1.2. To the extent that matching each artist-backer pair is subject to a common object, the platform, we have something closer to a one-sided (or stable roommates) matching problem. In such problems, stable matchings are not guaranteed to exist. See Roth and Sotomayor (1990) for a treatment, as well as Gusfield and Irving (1989) for an in-depth analysis of the connection to two-sided matching problems.\footnote{To the extent that the platform is in the center of the graph, as in Figure 1.1, our problem resembles a three-sided matching problem. However, platforms do not have preferences over either set of agents in our model. In three-sided matching problems, Alkan (1988) shows that a stable matching may not exist.}

Second, note that an artists’ preferences naturally express a lexicographic order on several components. The most important component in their preferences might be the campaign, embodied in the contract terms. The second most important component could be the platform, followed by the identities of the backers on the other side. Note that the lexicographic ordering could be different. For instance, artists may be most drawn to platforms that allow them to focus on music lovers, agreeing to whatever terms the platform sets. In this case, agents are most important, followed by the platform and then contract terms. If artists have lexicographic preferences, we recover the two-sided structure: finding backers is the most important consideration, making it easier to find a match.

Let $T = T_A \times T_B$ be the set of contract terms as in the previous subsection. We define the notion of a term-constrained stable allocation as follows.

**Definition 1.3.** A set of contracts $Y \subseteq X$ is **term-constrained feasible** if for $x, x' \in Y$, if $x_I = x'_I$ then $x_{T_A, T_B} = x'_{T_A, T_B}$. Let $\tilde{X}$ denote the term-constrained feasible set. An allocation $Y$ is **term-constrained stable**, if the following holds:

1. **Feasibility:** $Y \in \tilde{X}$.

2. **Individual rationality:** $C_s(Y) = Y_s$, for $s \in A \cup B$.

3. **No feasible blocking set of contracts:** $\exists Y' \in \tilde{X}$, such that $Y'^s = C_s(Y \cup Y')$, for all $s \in (Y')_s$.

The requirement of individual rationality is the same, but now we impose feasibility on both the allocation $Y$ and any potential blocking set $Y'$. 

---

15To the extent that the platform is in the center of the graph, as in Figure 1.1, our problem resembles a three-sided matching problem. However, platforms do not have preferences over either set of agents in our model. In three-sided matching problems, Alkan (1988) shows that a stable matching may not exist.
The following result states that a term-constrained stable allocation may not exist in our general environment.

**Theorem 1.4.** When preferences are substitutable, the set of term-constrained stable allocations may be empty.

The proof is by the following example.

Let \(A = \{a_1, a_2\}, B = \{b_1, b_2\}, I = \{i\}, T_A = \{t_1, t_2\}, \) and \(T_B = \emptyset.\)

Suppose preferences over contracts are as follows:\(^\text{16}\)

\[
\begin{align*}
P_{a_1} : & \{(a_1, b_1, t_1)\}, \{(a_1, b_2, t_1)\} \\
P_{a_2} : & \{(a_2, b_2, t_2)\}, \{(a_2, b_1, t_2)\} \\
P_{b_1} : & \{(a_2, b_1, t_2)\}, \{(a_1, b_1, t_1)\} \\
P_{b_2} : & \{(a_1, b_2, t_1)\}, \{(a_2, b_2, t_2)\}
\end{align*}
\]

Note the preferences satisfy substitutes. When agents joining the same side must share the same terms, there are only four possible feasible and individual rational allocations, and one can easily check that none of these is stable:

\[
\begin{align*}
Y_1 = \{(a_2, b_1, t_2)\} & \rightarrow \text{blocked by } Y_2 \\
Y_2 = \{(a_2, b_2, t_2)\} & \rightarrow \text{blocked by } Y_3 \\
Y_3 = \{(a_1, b_1, t_2)\} & \rightarrow \text{blocked by } Y_4 \\
Y_4 = \{(a_1, b_1, t_1)\} & \rightarrow \text{blocked by } Y_1
\end{align*}
\]

Thus, there is no term-constrained stable allocation. It is worth noting that these preferences also (trivially) satisfy the law of aggregate demand. Intuitively, the latter assumption would not help because it makes no demands on the contract terms.

However, if we allow agents who join the same platform to have different terms, then there are two (unconstrained) stable allocations:

\[
\begin{align*}
Y_5 = \{(a_1, b_1, t_1), (a_2, b_2, t_2)\} \\
Y_6 = \{(a_1, b_2, t_1), (a_2, b_1, t_2)\}
\end{align*}
\]

As mentioned above, in the example of crowdfunding, it is possible that
d\(^{16}\)Since there is one platform and side \(B\) agents do not face contract terms, we suppress the notation of platform \(i\) and the term for side \(B\) in the listed contracts.
agents’ preferences express a lexicographic order over several components of the contract. Other examples include charities, religious organizations, and professional associations. Under the additional assumption of lexicographic preferences, we have the following existence result for term-constrained stability.

**Theorem 1.5.** When preferences are substitutable and lexicographic, a term-constrained stable allocation exists.

The proof, in Appendix A.1.3, is constructive and uses a modified version of Hatfield and Milgrom (2005)’s cumulative offer algorithm.

1.4 Concluding remarks

The main contribution of this paper is a new application of two-sided matching that provides a complementary perspective for analyzing certain platform markets. Our focus is on agents rather than platforms, and the "microstructure" of interactions at platforms. We allow agents to have preferences over the quality of interactions with agents on the other side, and not just the quantity. The solution concept of stability is a reasonable equilibrium criterion as it eliminates switching. We use the structure of the set of stable allocations to analyze equilibrium market structure.

As Weyl (2010) points out, there is a gulf between the standard platform ("two-sided") markets literature, and the two-sided matching (and more broadly, market design) literature. This paper is a first attempt at bridging that gap. Several interesting problems are left for future work. The most straightforward extension is to dispense with the single-homing assumption, which is not essential to our results. Hatfield and Kominers (2015) relax unitarity and show that many of the properties of stable allocations hold for the more general case in which each artist and backer pair may sign multiple contracts. In our model, this would allow agents to multi-home. Another extension is the introduction of platform objectives. The simplest way to do this is to assume platforms have a revenue target, but otherwise have no preferences over agents. In this setting, more care must be taken in defining an appropriate solution concept. This is left for future work.
Chapter 2

Sorting Under Public School Choice

2.1 Introduction

Public school options in the United States have evolved significantly in the last two decades. Households face a landscape that may include vouchers, charter schools, and open enrollment. In principle, such policies are meant to provide households with choices beyond the standard neighborhood school. However, it is not clear to what extent households value such options over the traditional form of choice: sorting to secure assignment in a neighborhood school.

In this paper, I estimate the value for school choice relative to neighborhood assignment. In particular, I exploit a natural experiment in the twin cities of Champaign-Urbana, Illinois in which households can substitute between two kinds of public school districts: Urbana offers traditional neighborhood assignment, while Champaign offers open enrollment.\(^1\) Data from before and after the Champaign school district adopted open enrollment provides variation in housing prices due to sorting. Given observed prices in the treatment (open enrollment) and control (neighborhood assignment) states, I estimate the expected marginal willingness to pay for school choice relative to residence-based assignment in a hedonic regression.

Of course, sorting means that treatment is endogenous: households choose their school district. I estimate a difference-in-differences model for the hedonic regression, controlling for variation in neighborhood amenities. In particular, I control for neighborhood school quality and distance to neighborhood school, as well as home-level observables such as number of bedrooms and bathrooms. The result of combining causal inference with hedonic regres-

\(^{1}\)In what follows, I use school choice to refer exclusively to open enrollment programs unless otherwise indicated.
sion is an estimate of the average treatment effect (ATE) which directly corresponds to the marginal willingness to pay for school choice relative to residence-based assignment.

I find that, on average, the value for school choice is 5-7% higher than for neighborhood assignment. That is, households are willing to pay more for school choice than for a guaranteed neighborhood school. Moreover, the results are robust to various forms of regularization, including lasso and ridge regression, as well as alternative mixed-effects specifications.

Single family home sale prices are modeled in a hedonic framework. Rosen (1974)’s hedonic model provides a theoretical framework for estimating implicit prices of attributes of differentiated goods such as houses. There are several challenges to a traditional hedonic approach, the most important being potential bias arising from omitted attributes (for example, curb appeal). Recent literature has shown that causal inference may yield valid estimates of implicit prices where a standard approach might fail; see Klaiber and Smith (2013); Parmeter and Pope (2013) for excellent reviews. The idea is to isolate the impact of the policy change on prices through the attribute of interest.

Causal inference, of course, is not without its pitfalls. By definition, the treatment effect is the difference between the price of a home in the treatment state and the price of a home in the control state. The fundamental problem of causal inference is that the counterfactual price is unobserved: if a home sells in the treatment state, then we never observe the price of the home in the control state. In other words, we only observe one of two potential outcomes. The unique nature of the quasi-experimental setting in Champaign-Urbana provides both a reasonable control group, in the sense that there are enough similarities between both districts as to make them substitutes for households that can afford to sort, and variation in housing prices from before and after the policy change in each district.

This paper contributes to an emerging literature on empirical matching and school choice. A large body of literature on public school choice builds on the pioneering work of Gale and Shapley (1962) on two-sided matching. This research program is largely theoretical and focused on how critical design elements—such as assignment mechanisms, priorities, and quotas—affect the match between students and schools with respect to efficiency, fairness, and
incentives.² More recently, this literature has inspired an empirical research program to evaluate the impact of school choice programs used in practice. Many of these papers leverage the school choice mechanisms themselves to provide random variation, as summarized in Abdulkadiroğlu et al. (2015). Abdulkadiroğlu et al. (2013) use this approach to estimate the effect of small high schools on student achievement. Abdulkadiroğlu et al. (2015) examine whether centralized design offers welfare improvements over decentralized assignment. In cases where school choice mechanisms are not strategy-proof, and thus revealed preference analysis is not reliable, Agarwal and Somaini (2014) use equilibrium analysis in the spirit of partial identification to recover preferences from reported rankings.

A separate body of literature, largely independent of the theoretical work on school choice mechanisms, has investigated the general equilibrium impact of alternative forms of school choice—particularly voucher programs, reviewed in Epple et al. (2015b), and charter schools, reviewed in Epple et al. (2015a). Related work by Epple and Romano (2003) compares residence-based assignment to open enrollment in a general equilibrium framework. In the conclusion, they directly address the consequences of one district adopting open enrollment while a neighboring district remains residence-based. In particular, they provide an example to illustrate how stratification can arise across school districts in equilibrium. One possibility is that high-income households relocate to the residence-based district, lowering the tax base (and hence the quality) of education in the choice district. More recently, Avery and Pathak (2015) formalize the ideas in this example by modeling the impact of open enrollment on sorting. They show that higher types indeed do sort out of school choice districts, which has a negative impact on school quality in the choice district.³ In a sense, my paper is an attempt to quantify the impact of such sorting.

A brief history of assignment policy changes in Champaign-Urbana is given in section 2.2. The empirical framework is presented in section 2.3, including a description of the data. The main results are presented in section 2.4. Section 2.5 concludes. All figures and tables are in Appendix B.1 and B.2, respectively.

²The foundational papers include Abdulkadiroğlu and Sönmez (2003); Abdulkadiroğlu et al. (2005c,a); Ergin and Sönmez (2006). See Abdulkadiroğlu (2013) for a review.
³In Epple and Romano (2003)’s model, ability is positively correlated with income.
2.2 Public school assignment policies in Champaign-Urbana, IL

Under residence-based assignment, a household’s choice set is typically restricted to its neighborhood school. As a result, residence-based assignment has historically been associated with increased segregation in schools and across neighborhoods. The major legal challenges to school districts accused of discrimination stem from post-

Brown v. Board of Ed desegregation efforts that were largely static and unresponsive to sorting and changing demographics. The idea behind open enrollment is that a household’s choice set includes all of the schools in the district. The principal goals of such programs are to increase public school options and to promote diversity.

Open enrollment programs assign students to schools through a mechanism that elicits each household’s ranking over schools in its choice set. In addition to preferences, the mechanism also considers each applicant’s priority at any given school. For example, an applicant with a sibling enrolled at one of its ranked schools has higher priority at that school. School choice mechanisms differ in how they incorporate preferences and priorities, and the substantial theoretical literature reviewed in Section 2.1 investigates properties of such mechanisms.

In practice, such priorities mean that households remain tied to certain schools. Thus, it is not obvious a priori if the adoption of school choice increases or decreases a home’s value. On the one hand, a particular location in the district no longer guarantees your child a spot at any single school. On the other hand, your child will have higher priority for any school within walking distance of your home. Since neighborhood school attendance areas are typically redrawn to balance the distribution of students across schools, such walk zone schools may in fact be closer (and generally more preferred) than a child’s previous neighborhood school. Moreover, it is possible that a household has multiple walk zone schools in its choice set.

My approach exploits public school assignment policy changes in the cities of Champaign and Urbana, located in Champaign County, Illinois. The two cities are adjacent and share the flagship campus of the University of Illinois.

---

4Exceptions are made when only certain schools offer a program, such as special needs, English as a second language (ESL), and bilingual programs. Nevertheless, the choice set for such households is restricted.
An attractive feature of this setting is that it offers a unique “natural experiment” in a small urban area, without other forms of school choice such as vouchers and charter schools confounding the effects of open enrollment. Prior to 1998, both districts offered residence based assignment. Beginning academic year 1998-99, Champaign’s school district adopted elementary school choice. Urbana’s school district continues offering residence-based assignment to before and after the policy change. Figure B.1 presents a map of the attendance areas in each school district prior to the policy change.

Complaints alleging discrimination against black students in Champaign’s school district arose in May of 1996. In September of 1997, the parties agreed to adopt a form of school choice beginning academic year 1998-99, even as the complaints gave way to a civil suit. In 1998, Champaign adopted a “controlled choice” system, which seeks to balance elementary school access and diversity through the use of race-based quotas. Since adopting school choice, the Champaign school district closed one elementary school, opened several others, and re-branded an existing school as a magnet school.

Urbana has maintained a system of neighborhood-based public school assignment, with attendance boundaries periodically re-drawn to balance enrollment. Beginning academic year 2002-03, Urbana modified its attendance boundaries for this purpose. Thus, I focus on the period 1996 (the earliest year for which I have housing data) through 2001. The key policy changes are summarized in Figure B.4. Following attendance area changes in Urbana, the choice set for individuals in the control regime changes significantly and thus we would expect the hedonic price function to adjust as well. Therefore, I restrict my sample to the period 1996-2001.

While the two cities are far from homogeneous, they share enough similarities to be close substitutes for a large portion of home buyers. In particular, performance on standardized tests, a popular measure of school quality, is comparable across the two districts. In Section 2.4.1, I present descriptive statistics to support this point. More importantly, close proximity means some families (especially middle to upper middle class) can consider housing and schooling options in both cities. That is, some families can choose between school choice and residence-based assignment. It is therefore important to evaluate treatment effects conditional on observables, such as the size of the house.
2.2.1 How school choice works in Champaign

School choice programs like Champaign’s typically work as follows:

- Each family ranks the schools in the district
- Each school in the district ranks applicants
- A mechanism collects both rankings and computes a match of students to schools

The details of how each step is implemented vary. In Champaign during the period 1996-2001, each household was allowed to rank up to three (of eleven) elementary schools.\(^5\)

Determining how schools rank applicants is important for public school choice. Since primary education is considered by many to be a basic right, public schools are not allowed to subjectively rank applicants. Public schools do, however, give certain applicants higher priority than others. The two priority classes used in Champaign are:

- Sibling priority: if a student has a sibling enrolled in a school, then it is desirable to keep both siblings together;
- Proximity priority: if a student lives within walking distance\(^6\) of a school, then it is desirable to let the student attend a school close to home.

Mathematically, priorities are like preferences. The key difference is that priorities are set by the district and thus cannot be manipulated.

The choice of mechanism corresponds to the district’s rules for determining assignments. These rules have important implications for how a family ranks schools. In Champaign, like many other choice districts, sibling priority at a school is generally satisfied. Proximity priority at a school, however, is only available if an applicant ranks that school first. In the literature, this mechanism is known as the “Boston mechanism,” since it corresponds to the school choice program originally used by the Boston Public Schools.\(^7\)

---

\(^5\) Today, households may rank all twelve existing elementary schools.

\(^6\) A proximity priority school is defined as any school within 1.5 miles of a student’s home, or else the closest school to a student’s home.

\(^7\) In fact, Champaign’s school choice mechanism was designed by the same consulting firm that designed BPS’s mechanism.
The Boston mechanism creates incentives for applicants to misreport their preferences over schools; e.g., if my true first choice school is overdemanded, I would rather not waste my proximity priority on it and may instead rank another school in my walk zone to which I am more likely to be admitted.

Champaign's (and Boston's) school choice program is technically one of "controlled choice:" assignments are determined subject to satisfying racial quotas. Operationally, this means the district sets both a floor and a ceiling for the number of seats it reserves for black applicants. The mechanism fills open seats first, and the remaining students compete for the reserved seats. Racial quotas thus represent a third priority class—another way for schools to objectively rank applicants.\footnote{Following the 2007 Supreme Court ruling in the Seattle “PICS” case, school districts were barred from using racial quotas. Thus, in 2009, Champaign moved toward using quotas based on socio-economic status (SES). Of course, SES is still correlated with race.}

2.3 Empirical framework

My goal is to estimate the value for public school choice relative to traditional residence-based assignment. Sorting by households in response to a policy change—the idea that households “vote with their feet”—reveals their preferences for the policy. However, directly observing such sorting is rare in practice. I therefore adopt Rosen (1974)'s hedonic model for housing prices, which assumes that prices reflect the marginal willingness to pay for bundled amenities, such as the local public school assignment mechanism. Unfortunately, estimating these values is difficult in practice, largely due to unobserved factors that confound observed prices. The key to my approach is the assumption that, following the policy change, some households can substitute between either regime. Using housing prices from before and after public school choice was instituted, in both the treatment and control regimes, I estimate the casual impact of the policy change in a standard difference-in-differences model. See Parmeter and Pope (2013) for a review of hedonic amenity valuation using causal inference.
2.3.1 The hedonic price model

The hedonic pricing model provides a general equilibrium framework for markets with differentiated products, such as houses. The key ingredient is the hedonic price function, $f(z)$, where $z \in \mathbb{R}^d$ is a vector of attributes such that the bundle $z$ completely characterizes the house for sale. Thus, $f(z)$ reflects the price of the bundle $z$.

Consumers have utility function

$$U(x, z; \zeta)$$

where $x$ is a consumption good and the consumer faces budget constraint $w = x + f(z)$. Note that $\zeta$ represents unobserved (to the econometrician) consumer taste shifters. On the supply side, producers face cost function

$$C(z, M, \eta)$$

where $M$ is the quantity of bundles $z$ produced. Here, $\eta$ has a similar interpretation as $\zeta$.

Let $\theta(z; u, w)$ denote the bid function, which represents a consumer’s willingness to pay for bundle $z$, holding income $w$ and utility fixed at $u$. Let $\psi(z; \pi)$ denote a seller’s offer curve, which provides the ask for bundle $z$, holding profits fixed at $\pi$. In equilibrium, one can show that

$$\frac{\partial f}{\partial z} = \frac{\partial \theta}{\partial z} = \frac{\partial \psi}{\partial z}$$

(2.1)

where the consumer’s first order condition implies

$$\theta_z \equiv \frac{\partial \theta}{\partial z} = \frac{U_z(w - \theta, z; \zeta)}{U_x(w - \theta, z; \zeta)}.$$

Thus, Rosen’s approach reduces inference on the consumer’s marginal willingness to pay for an amenity to inference on the hedonic gradient. Underlying the equilibrium analysis are several important assumptions:

**Assumption 2.1.** Assumptions for equilibrium:

1. Unit demand;

2. Numeraire $f(x) = 1;$
3. Interior solution.

The equilibrium condition (2.1) is key to Rosen’s approach, which may be summarized as follows.\(^9\) Suppose we have prices, \(y\), and data \(z^{\text{obs}} \in \mathbb{R}^k\), which represent the subset of attributes observable to the econometrician. Then we obtain an estimate of the hedonic price function from a regression of the form:

\[
y = f(z^{\text{obs}}, \theta) + \epsilon
\]

where the hedonic may be parameterized by some \(\theta\). We can then use this to estimate the gradient of \(f\) at \(z_j\), i.e., \(\hat{f}_j\), for an observed amenity of interest, \(z_j \in z^{\text{obs}}\). Under the assumptions below, in addition to the assumptions for equilibrium, this estimate represents the implicit price of \(z_j\).

**Assumption 2.2.** Assumptions for identification:

1. Single market;
2. Stability of \(f(z)\) over time;
3. \(\zeta\) are iid across consumers.

The implicit prices \(\{\hat{f}_j\}_{j=1:k}\) provide the value for the observed bundle \(z^{\text{obs}}\). In a potential outcome setting, in which \(z_j\) represents a policy of interest, \(\hat{f}_j\) represents the value for a marginal policy change, say from \(z_j = 0\) to \(z_j = 1\), under the above assumptions. See Parmeter and Pope (2013) for more details.

A fundamental problem with empirical implementation of this approach is that the hedonic bundle characterizing the house, \(z\), is not fully observed by the econometrician. The classic example is “curb appeal” while arguably an important component of a homebuyer’s decision, it is not tangible much less observable to an analyst. Unobserved attributes which affect decision making therefore create an endogeneity problem. One approach to this problem is to use causal inference, carefully conditioning on observed confounders in a difference-in-differences model.

\(^9\)In the literature, this is often referred to as the first stage of Rosen (1974)’s approach. The second stage involves recovering the marginal willingness to pay function, \(\theta_z(w, z, \zeta)\), a much more difficult problem as it requires specification of \(U(\cdot)\). See Bartik (1987); Epple (1987); Ekeland et al. (2002); Heckman et al. (2010); Bishop and Timmins (2015) for background and various approaches.
2.4 Main results

In Section 2.2, I claim that Champaign and Urbana share enough similarities that comparisons are meaningful. In the next subsection, I present descriptive statistics to illustrate similarities in sale prices, housing attributes, and school quality across the two districts. In subsection 2.4.2, I describe the model specification and discuss results from various estimates.

2.4.1 Describing the data

Basic property data for Champaign County is available to the public on the Champaign County Assessor’s website. While the County provides records of sale date and price for each property, it lacks even basic attribute information such as lot size. I construct a more complete data set by combining County data\(^{10}\) with single-family home sale and attribute data obtained from both DataQuick and the Multiple Listing Service (MLS).\(^{11}\) Besides sale date and price, the data includes basic information on the buyer (including loan amount and the buyer’s address) and attributes of the home (such as lot size and square feet).\(^ {12}\)

Figure B.5 presents the density of the log sale price for Champaign and Urbana, before and after the policy change. While the distributions are not the same, they do not appear to differ in a meaningful way. In Figure B.6, the median of the log sale price is plotted over the period 1996-2001. Note that, while levels differ, Champaign and Urbana appear to follow a similar trend in log sale price over the study period.

Housing level descriptive statistics, including the log sale price, are presented in Table B.1. The most important variable is, of course, the log sale price. Note that prices tend to be higher in Champaign than in Urbana, both before and after the policy change. In particular, the median (aver-

---

\(^{10}\) The data sets are combined using the unique Property Identification Number (PIN) assigned to each home by the County.

\(^{11}\) While the MLS data includes a lot of attribute information, one major drawback is that it is incomplete as it only includes listings for sales through a licensed broker. Moreover, the data is not consistent and is subject to selective reporting (i.e., with the purpose of selling the house).

\(^{12}\) Attributes from DataQuick are based on assessments from 2012, the most recent property assessments when the data was purchased. MLS provides attributes as listed on the MLS at the time the property was on the market.
sale price is roughly $82,000 ($81,200) in pre-choice Champaign, versus roughly $66,000 ($68,600) in pre-choice Urbana. After the policy change, the median sale price is about $94,000 in Champaign and about $78,000 in Urbana. However, there is also more variation in prices in Champaign than in Urbana, which reflects the larger housing market in Champaign.

Square feet (Sqft), lot size (Lotsize), and basement (Basement) are presented in 1000 square feet. Garage is an indicator that takes the value 1 if the property has a garage or carport, and is 0 otherwise. It is worth noting that square footage and lot size tends to be larger in Champaign than in Urbana, but once again there is also more variation in these attributes in Champaign. Another major difference is that neighborhood schools—defined as a home’s pre-choice residence-based assignment—are on average further from homes in Champaign than from homes in Urbana. Once again, there is more variation in these distances for Champaign, and this pattern simply reflects the fact that Champaign is a bit larger than Urbana (see, for instance, the map in Figure B.1). On the other hand, homes in Urbana are generally older. More importantly, the other key attributes—number of bedrooms, number of bathrooms, basement size, and presence of garage or carport—are generally similar in both cities.

The housing data provides addresses which I geocoded and matched to basic Census tract and level demographics. Table B.2 presents descriptive statistics on demographics at the Census tract level, by district and by treatment period. Once again, there appear to be some differences (e.g., with respect to number of households and median income) that are also accompanied by more variation in Champaign. Differences in the median value of owner-occupied units at the Census tract level appear to reflect the same trend found in the housing-level data. One key attribute that is similar for both cities, as reflected in the housing-level data, is median number of rooms.

Public school data is gathered from multiple sources, and Table B.3 presents descriptive statistics for schools, broken down by district and treatment period. The Illinois State Board of Education (ISBE) provides annual school and district level “report cards” for every public school district in Illinois.

---

13 I used several free services, including the Census geocoder, Nokia HERE’s geocoder, and Google’s geocoder.
14 Unfortunately, I could not obtain data a finer scale than Census tract for Champaign-Urbana.
Most importantly, the school level data includes figures on enrollment and performance on standardized tests, broken down by various demographics. In the table, I present summaries for measures used in the analysis: enrollment, average class size, percent black students, percent low income students, and the fraction of third grade standardized test takers who exceed state goals on reading and math.\textsuperscript{16} At the district level, the local school district property tax rate is available from Champaign County. In any given fiscal year, the school district tax rate is higher in Urbana than in Champaign. More importantly, note that the percent of students who exceed state goals on third grade mathematics and reading exams is similar across both districts, before and after the policy change. In fact, the standard deviations are similar as well. Note that enrollment is generally larger in Champaign, which is a bigger district, and this is one reason for the differences in average class sizes. A more important contrast is that Champaign has a higher percentage of black students, while Urbana has a higher percentage of low income students.\textsuperscript{17}

The National Center for Education Statistics (NCES) provides an annual series called the Common Core of Data (CCD), a comprehensive list of public schools in the United States. The CCD includes some basic demographics as well as the school’s physical address. This is geocoded and merged with the ISBE data.\textsuperscript{18}

School assignment policy information is provided directly by the school districts. This includes parent outreach materials, board minutes, and attendance area maps.\textsuperscript{19} Using scanned pdf images of attendance areas, I used GIS software to reconstruct attendance area maps for each district. These maps are used to compute, for instance, distances between homes and schools, walk-zones, and the number of schools within a home’s walk-zone.

\textsuperscript{16}ISBE provides the fraction of standardized test takers who are below, meet, or exceed state goals, for subjects including reading, math, science, and social studies, at various grade levels. Testing begins in the third grade.

\textsuperscript{17}Recall that Champaign has nine elementary schools before the policy change, closing one and opening two new schools in the first academic year of the policy change, while Urbana always has the same six elementary schools.

\textsuperscript{18}Note that CCD provides longitude and latitude information. However, the coordinates for the same physical address are not always consistent. Geocoding is done using the Census geocoder.

\textsuperscript{19}Current maps for Urbana School District 116 are on the district’s website, http://www.usd116.org/schools/elementary-attendance-areas/. Historical attendance area maps for both Champaign and Urbana were obtained by FOIA requests.
For the selection equation, I need a variable that affects selection but not the price of a home. The variable I use is the buyer’s origin: is the buyer coming from Champaign, Urbana, or elsewhere? The Dataquick data provides a buyer’s address for each transaction. I construct a three-level factor based on the buyer’s zip code.

As mentioned in section 2.2, I restrict my dataset to the period 1996-2001, just prior to Urbana changing its attendance area boundaries. Moreover, I drop sale prices in the top and bottom two percentiles, i.e., I drop the top 2% and bottom 2% of house prices. This seems to provide a reasonable sale price distribution; see Figure B.5. Table B.4 tabulates number of observations in each district by treatment and control. Note that the number of observations in the treatment state is roughly equal to the number of observations in the control state.\footnote{The table also demonstrates how much larger the Champaign housing market is relative to Urbana’s.}

### 2.4.2 Model specification

In this section, I present the hedonic specification underlying the difference-in-differences approach.

Let $Y_{ijt}$ be the log sale price for home $i$ in district $j$ at time $t$. The baseline linear hedonic diff-in-diff specification is:

$$
Y_{ijt} = \gamma_j + \gamma_t + D_{ijt}\delta + z_h^T \gamma_h + z_s^T \gamma_s + z_n^T \gamma_n + x^T \gamma_x + v_{ijt} \quad (2.2)
$$

$$
\begin{align*}
z_h &= \left( \text{Sqft, Lotsize, Age Home, Garage, Basement, Bedrooms, Baths} \right)^T \\
z_s &= \left( \text{School Distance, District Tax Rate, Average Class Size, Math, Reading} \right)^T \\
z_n &= \left( \text{Median Income, Median Rooms, Owner Occupied Units, Owner Occupied Median Value} \right)^T
\end{align*}
$$

where $\gamma_j$ is a school district fixed effect, $\gamma_t$ is a year fixed effect, and $z_h, z_s, z_n$ represent house-, school-, and neighborhood-level observables, respectively. Note that school-level observables—School Distance, Average Class Size,
Math, and Reading\textsuperscript{21}—correspond to the property’s \textit{pre-choice} neighborhood school. See Section 2.4.1 for more details. An important control in (2.2) is $x$, which represents Buyer Origin (Champaign, Urbana, or Other). This variable should affect treatment and not house price, hopefully providing control for unobservable confounders.

The key to causal inference in the current specification (2.2) is the $D_{ijt}$ term, where $D_{ijt} = 1$ if and only if home $i$ is in Champaign, post-choice. That is, $D_{ijt}$ designates treatment. In the next subsection, I present results from estimating (2.2).

2.4.3 Estimation results

The coefficient of interest in (2.2) is $\delta$, which is an estimate of the average treatment effect, $\Delta \equiv \mathbb{E}[Y_1 - Y_0]$.\textsuperscript{22} Results are shown in Table B.5.

Importantly, the DD model suggests that, on average, households prefer school choice over residence-based assignment. In particular, the estimate of $\delta = 0.0682$ implies households are willing to pay \textit{at least} 7\% more for school choice.\textsuperscript{23}

I also consider mixed-effects models, with random intercepts varying by school district and year. Results are shown in Table B.6. Model 1 is the random intercept specification, with observations grouped by school district and year. Model 2 is the random intercept and slope specification, with slopes varying by treatment assignment. Both specifications are fairly consistent with the DD model estimated in Table B.5, albeit a bit lower. The estimates of $\delta \approx 0.06$ imply a lower bound on the treatment effect of about 5.8-6\%.

One might naturally wonder if these results are sensitive to the particular specification in (2.2). Rather than estimating various permutations of the same basic model, I consider regularization methods for data-driven variable

\textsuperscript{21}Math and Reading represent the percentage of third grade students that exceed state goals on each standardized test.

\textsuperscript{22}More precisely, $\delta$ is a lower bound on the percentage difference between $Y_1$ and $Y_0$, since the model is estimated in log sale prices. This follows from Jensen’s inequality.

\textsuperscript{23}This follows from the fact that $\mathbb{E}[y_1/y_0] = \mathbb{E}[\exp(Y_1 - Y_0)] \geq \exp(\mathbb{E}[Y_1 - Y_0]) = \exp(0.0682) = 1.07$, where $y_1, y_0$ are prices in levels and $Y_1, Y_0$ are prices in logs.
selection. Lasso regression coefficient estimates are shown in Table B.7. Regularization by elastic net includes lasso and ridge regressions as special cases, and can mix continuously between the two; see Friedman et al. (2010). For brevity, I present only results for the lasso, but it is worth noting that ridge regression, as well as various mixtures of the two, yield similar results. Importantly, all of the regularized models consistently estimate $\delta \approx 0.05$, implying a lower bound of about 5% on the willingness to pay for school choice relative to residence-based assignment.

2.5 Conclusion

This paper is an attempt to estimate the relative values of two market designs. In particular, I leverage a natural experiment to compare the value of school choice relative to traditional residence-based assignment. Using a linear hedonic difference-in-differences model, I find that households are willing to pay (at least) between 5-7% more for school choice on average.

One limitation of the present study is the linear specification. It is very unlikely that the true hedonic price function is linear. Another limitation is that the results simply reflect a lower bound on the average willingness to pay. This measure lumps all households together, but it is unlikely that the effects are the same for higher-income households, who have the ability to pay for their desired school, and for lower-income households, who do not. It would be worthwhile to decompose the effect among different types of households, and moreover to estimate the full distribution rather than simply the average. One possibility that addresses both of these concerns is to estimate a semi- or non-parametric kernel regression. Another possibility is to model the full distribution of both counterfactual prices and a household’s selection into their district in a Bayesian semi-parametric mixture model. Such extensions are left for future work.

---

24 Regularization methods penalize regression models for increasing complexity that does not improve prediction, e.g., as measured by MSE. As such, they are analogous to placing a particular prior on the parameters (normal for ridge, Laplace for lasso). Ridge regression shrinks coefficients toward zero, while lasso performs variable selection by discarding coefficients.

25 The regularization parameter is chosen to minimize cross-validation error.
3.1 Introduction

Suppose you are tasked with pairing a group of men and a group of women together when each person has preferences over potential partners. Can you accomplish this task such that, once your matching is in place, no man or woman can obtain a better partner on their own? In a seminal paper, Gale and Shapley (1962) model this situation as a marriage market and present an elegant solution: the deferred acceptance algorithm. Gale and Shapley (1962) use their algorithm to prove that such stable matchings exist in two-sided markets. While Gale and Shapley (1962) did not aim to provide guidance on applied market design, their algorithm has come to play a key role in the design of centralized markets.\(^1\) The question we address in this paper is: how well can agents do without centralization?

Gale and Shapley (1962)’s algorithm has been independently discovered by practitioners many times. As Roth (1984a) showed, the National Resident Matching Program (NRMP) for new doctors had been using a version of deferred acceptance since 1951. Several other instances of the algorithm in the field were later documented, including the job market for clinical psychologists and dorm room assignment at MIT (Roth, 2008). Such observations renewed interest in Gale and Shapley (1962)’s stylized model of matching, with stability as the principal objective and deferred acceptance as the foundation for practical market design. Eventually, deferred acceptance became fundamental to the intentional reorganization of existing markets as centralized clearinghouses, most notably in the redesign of the assignment mechanisms for Boston Public Schools (Abdulkadiroğlu et al., 2005d) and New York City

\(^1\)Gale and Shapley (1962) do state their hope “that some of the ideas introduced here might usefully be applied to certain phases of the admissions problem.”
The deferred acceptance algorithm is practical in applications since it can find a stable matching in polynomial time. With \( n \) men and \( m \) women in a marriage market, deferred acceptance computes a stable matching in at most \( n \cdot m \) steps. Its speed and simplicity make it a natural candidate for centralized design, but there is nothing inherently centralized about deferred acceptance. The standard interpretation of the algorithm as one side of the market making successive proposals has a decentralized flavor to it since proposing agents don’t need information beyond their own preferences. However, agents have to be coordinated in two ways to properly execute deferred acceptance. First, only one side of the market can make proposals. Second, proposals must be grouped into rounds where agents propose at most once (though the order of proposals within rounds can be arbitrary). We argue that coordination is the distinguishing feature between centralized and decentralized markets.

In particular, we consider a class of decentralized matching markets introduced by Roth and Vande Vate (1990). Their idea, roughly, is to let agents take turns making proposals to a more preferred partner. Two agents that prefer each other to their current respective partners are said to form a blocking pair. Roth and Vande Vate (1990) show that, starting from an arbitrary matching, such random proposal processes eventually converge to a stable matching if each blocking pair has positive probability of being selected.

While these results suggest agents can attain stability without a centralized authority, Ackermann et al. (2011) show some markets almost certainly require exponentially many proposals to reach stability when blocking pairs are rematched uniformly at random. One is tempted to conclude that, in practice, decentralized markets cannot be ensured of reaching a stable matching. However, the process described above portrays market participants as naive. In particular, agents will accept any proposal when they are single. We introduce slightly more sophisticated behavior into the process and show that the process converges to stability in polynomial time for moderately sized markets. Moreover, since centralization by way of deferred acceptance requires at most a quadratic number rounds, we compare welfare under our process after a comparable number of rounds to the outcome of deferred acceptance.

In section 3.2, we introduce the matching environment and the model for proposal processes. Computational results are presented in section 3.3. In section 3.3.1, we explore convergence to stable matching in terms of number of
proposals, while in section 3.3.2, we compare welfare of our proposal processes before convergence to welfare under stable matching. Figures are presented in Appendix C.1 and tables are presented in Appendix C.2.

3.2 Model

3.2.1 Environment

We consider a standard one-to-one matching environment with strict preferences. In particular, a matching market of size $n$ is a triple $\theta = (M, W, \succ)$ consisting of a set of $n$ men $M$, a set of $n + m$ women $W$, and preferences $\succ = \{\succ_i\}_{i \in M \cup W}$ for each agent $i$ over potential partners on the other side of the market.

Preferences are strict linear orders over the set of potential partners for that agent and the option to remain unmatched, denoted by $\emptyset$. Agent $i$ finds agent $j$ acceptable if $j \succ_i \emptyset$. We will consider environments $\theta$ where every man finds every woman acceptable and vice-versa. Let $\Theta_n$ denote the set of all such markets of size $n$. An environment $\theta$ is balanced for some $n$ if $m = 0$, and otherwise it is unbalanced.

A matching is a function $\mu : M \cup W \rightarrow M \cup W \cup \{\emptyset\}$ describing how agents are paired. Let $\mathcal{X}(\theta)$ denote the set of all matchings for $\theta$. The partner of agent $i$ under matching $\mu$ is $\mu(i)$. The agent is single if $\mu(i) = \emptyset$. Valid matchings satisfy the following properties:

1. If $\mu(i) \neq \emptyset$, then $\mu(\mu(i)) = i$, i.e. if an agent $i$ has a partner, $\mu(i)$, then $\mu(i)$’s partner is $i$.

2. $\mu(M) \subseteq W \cup \{\emptyset\}$ and $\mu(W) \subseteq M \cup \{\emptyset\}$, i.e. all agents are either matched to an agent from the other side or single.

A matching $\mu$ is stable in $\theta$ if there is no pair of agents $m, w$ such that $m \succ_w \mu(w)$ and $w \succ_m \mu(w)$ and no agent $i$ such that $\emptyset \succ_i \mu(i)$. Since we consider only environments where all agents are mutually acceptable, the second requirement that all agents prefer their partners to being single is satisfied in every matching. Let $\mathcal{S}(\theta)$ denote the set of stable matchings for market $\theta$. 

38
3.2.2 Proposal processes as models of matching

Given a matching market $\theta$, agents begin at some initial matching $\mu_0$ with distribution $f_0(\mu) = Pr(\mu_0 = \mu)$. We will primarily consider $\mu_0$ as the empty matching where $\mu_0(i) = \emptyset$ for all $i \in M \cup W$. Matchings evolve according to a sequence of proposals between agents, with one proposal being made at discrete time steps, $t = 1, 2, \ldots$.

A proposal process, $P$, describes who makes a proposal at each step, who the proposer proposes to, and the conditions for whether a proposal is accepted. Given matching $\mu_t$ at step $t$, if agent $i$ proposes to agent $j$ and the proposal is accepted, then $\mu_t$ is updated as follows:

1. $i$ and $j$ are paired together: $\mu_{t+1}(i) = j$ and $\mu_{t+1}(j) = i$,
2. $i$ and $j$’s previous partners (if any) are now single: $\mu_{t+1}(\mu_t(i)) = \mu_{t+1}(\mu_t(j)) = \emptyset$,
3. and $\mu_{t+1}(k) = \mu_t(k)$ for all other agents $k$.

If the proposal is not accepted, then no change occurs, and $\mu_{t+1} = \mu_t$.

Deferred acceptance as a proposal process

In general, a proposal process may be random, inducing a random walk over the set of matchings, although deterministic proposal processes can also be considered. For instance, two versions of the deferred acceptance algorithm—with either men proposing or women proposing—are important examples of deterministic proposal processes.

Deferred acceptance can also be implemented semi-randomly by picking a fixed permutation of the proposing side and having the agents cycle through proposals according to this schedule. Supposing men are proposing, each man knows every other man has acted exactly once since his last proposal. This fixed schedule provides a monotonicity guarantee that underlie deferred acceptance. If an agent is unsure whether or not someone else has acted since their last proposal, the same monotonicity guarantee is no longer present and proposers are unable to work down their preference lists in the same way. Someone who once rejected me might now accept me, so the set of possible partners to consider making proposals doesn’t shrink. Of course, it
wouldn’t be worthwhile to only always propose to my first choice, so a lack of knowledge about when others have acted necessitates some randomness in how proposals are made. The continued need to explore all possible partners particularly holds if both men and women make proposals.

Uncoordinated proposal processes

To represent agents being unable to fully coordinate or observe the actions of others, we consider proposal processes where at each point in time, one agent is selected at random to act. The probability that a man makes a proposal might differ from the probability a woman makes a proposal, but within each side, all agents have an equal chance of acting.

With no knowledge about the actions of others, how should an agent choose who to propose to? A simple answer is to randomly propose to someone better than that agent’s current match. A woman who has repeatedly rejected a man might still potentially accept him now, so the man might plausibly think it’s worth another shot. If the man can keep track of all the partners of the women he is proposing to, the best he can do is never propose to a women who is with the same partner as a time when she rejected him.

**Definition 3.1** (Random better (best) reply). Begin at random matching \( \mu_0 \) with some probability \( p(\mu_0) \). Given \( \mu_{t-1} \) at time \( t \):

1. Pick a proposing agent \( i_t \in M \cup W \) at random

2. **Better (best) reply**: Agent \( i_t \) proposes to an agent (the best agent) \( j_t \) such that

\[
j_t \succ_{i_t} \mu_{t-1}(i_t)
\]

choosing uniformly at random if multiple such \( j_t \) exist.

3. Agent \( j_t \) accepts if \( i_t \succ j_t \mu_{t-1}(j_t) \)

   (a) If \( j_t \) accepts, update \( \mu_t \)

   (b) Else, set \( \mu_t = \mu_{t-1} \)

Set \( t = t + 1 \), and return to step 1.

Ackermann et al. (2011) show that these processes may take exponentially many rounds to reach stability. On some level, naive behavior by both
proposers and responders causes unnecessary delay. We now introduce two proposal processes that model plausible behavior for both proposers and responders without requiring a high level of sophistication.

To address naive behavior by responders, we introduce aspirations as a way for responders to keep from settling for inferior matches early on in the process. An agent's aspiration level is simply the minimum partner rank the agent is willing to accept. A responder will accept a proposal subject to their current aspiration level, which evolves over time.

Let $\rho(i,j)$ denote agent $i$’s ranking of agent $j$ wrt $\succ_i$:

$$\rho(i,j) \equiv \{|k : k \succ_i j|\}$$

We say that agent $i$ has aspiration level $\alpha_t(i)$ at $\mu_t$ if $i$ accepts a proposal from any agent $j \in \{j : \rho(i,j) \leq \alpha_t(i)\}$ and rejects a proposal otherwise. Aspiration levels are set and updated as follows:

**Definition 3.2 (Updating aspiration levels).** Given an environment $\theta$ and a proposal process $P$ with aspiration adjustment $a_n \in \mathbb{R}_+$:

- At $t = 0$, initial aspiration is $\alpha_0(j) = 1, \forall j$
- At $t > 0$, if $j$ receives a proposal from $i$, then

$$\alpha_t(j) = \begin{cases} 
\rho(j,i), & \mu_t(j) = i \\
\alpha_t(j) + a_n, & \mu_t(j) = \emptyset
\end{cases}$$

otherwise, $\alpha_t(j) = \alpha_{t-1}(j)$

In other words, a responder $j$ who accepts a proposal from a proposer $i$ sets their aspiration to their current partner’s rank. If instead $j$ rejects $i$’s proposal, then $j$ must adjust $\alpha_t(j)$ to become less picky if $j$ remains single, or else $j$’s aspiration remains set to their current partner, $\mu_t(j)$.

To address naive behavior by proposers, we allow agents to learn who is better than them. In particular, if a man $i$ is rejected by a woman $j$ when she is partnered with some other man $k$, then man $i$ should not waste another proposal on woman $j$ if she remains partnered with $k$. That is, man $i$ learns by revealed preference that $k \succ_j i$. Let $\lambda(i)$ denote $i$’s unattainable set. This is the set of all pairs of agents in which the woman’s partner is preferred
to man $i$,

$$\lambda(i) = \{(j, k) \in W \times M : k \succ_j i\},$$

with a symmetric definition for a woman $i$. Since agent preferences are private information, this set is empty at $t = 0$ and updated by agent $i$ whenever he is rejected in favor of another man. We thus define learning as follows:

**Definition 3.3** (Learning). Given an environment $\theta$, a proposal process $P$ admits learning if each agent $i$’s unattainable set $\lambda_t(i)$ is updated at $t > 0$ as

$$\lambda_t(i) = \begin{cases} 
\lambda_{t-1}(i) \cup (j, \mu_t(j)), & j \text{ rejects } i \text{ at } t \\
\lambda_{t-1}(i), & \text{else}
\end{cases}$$

with $\lambda_0(i) = \emptyset, \forall i \in M \cup W$.

An agent $j$ is **attainable at** $\mu_t$ for agent $i$ if

$$(j, \mu_t(j)) \notin \lambda_t(i),$$

that is, if $j$ is not partnered with an agent $\mu_t(j)$ for whom $i$ has been rejected before. It is natural, for example, for a man making a proposal to propose to an attainable woman, rather than proposing to a woman who will surely reject him.

We are now ready to define our first proposal process.

**Definition 3.4** (Random best attainable). Begin at random matching $\mu_0$ with some probability $p(\mu_0)$. Given $\mu_{t-1}$ at time $t$:

1. Pick a proposing agent $i_t \in M \cup W$ at random

2. **Best attainable proposal**: Agent $i_t$ proposes to the best agent $j_t$ such that

$$j_t \succ_{i_t} \mu_{t-1}(i_t) \text{ and } (j_t, \mu_{t-1}(j_t)) \notin \lambda_t(i_t)$$

(a) If no such $j_t$ exists, set $\mu_t = \mu_{t-1}$, set $t = t + 1$, and return to step 1.

(b) Else, go to step 3.
3. Agent \( j_t \) accepts if \( \rho(j_t, i_t) \succ j_t \alpha_{t-1}(j_t) \)

(a) If \( j_t \) accepts, update \( \alpha_t, \mu_t \)
(b) Else, update \( \alpha_t \) and set \( \mu_t = \mu_{t-1} \)

Set \( t = t + 1 \), and return to step 1.

Together, learning and aspirations guide pairs of men and women into “good” matches early on, so that only small adjustments are necessary in order to get to a stable match. In contrast, better and best reply dynamics are characterized by an inordinate amount of matching, breaking up, and re-matching, with adjustments toward stability fairly random.

Learning allows proposers to use their proposals more wisely. In principle, proposing agents can learn the preferences of agents on the other side, given sufficient proposals. Beyond learning all of the men preferred to himself, a man can learn which other men do not represent “competition.” Thus, learning can facilitate earlier pairwise matching.

On the other side, aspirations ensure responders do not settle into inferior matches too quickly. The danger from doing so, e.g., in the best reply dynamics, is that a lot of time is spent re-matching.

One potential problem with this process is that a man may cycle through a long list of attainable women before actually securing a match. Indeed, once a man learns which other men are preferred he has no proposal to make if the women remain matched to such men.

An alternative is for a proposing man to pursue single women first. If a man does not have a best attainable woman available, but there are unmatched women waiting around, then it seems reasonable to go after a single woman rather than losing the opportunity to propose. If it is desirable for agents to secure early matches, then it is natural that men propose to single women first. Of course, if no single woman is available, the best a man can do is to pursue his best attainable woman, if one is available.

We are now ready to define our second proposal process.

**Definition 3.5** (Random singles first). Modify the Random best attainable proposal process as follows:

2. **Best singles proposal**: Agent \( i_t \) proposes to the best agent \( j_t \) such that

\[
\mu_{t-1}(j_t) = \emptyset
\]
(a) **Best attainable proposal**: If no best single $j_t$ exists, $i_t$ proposes to the best attainable $j_t$

(b) If no such $j_t$ exists, set $\mu_t = \mu_{t-1}$, set $t = t + 1$, and return to step 1.

(c) Else, go to step 3.

The random singles first process is even more biased toward early matching than the random best attainable process. It represents a more risk-averse approach, in the sense that being matched is more important than finding the best match. Combined with learning and aspirations, the process should settle more quickly into “good” matches so that the path toward stability is less volatile in terms of re-matching.

### 3.3 Computational results

In this section, we present computational results on convergence to stability and welfare for our two proposal processes, in both balanced and unbalanced markets.

In what follows, the aspiration adjustment for an environment $\theta$ of size $n$ is set to decrease with the inverse of $n$ as:

$$ a_n \equiv \frac{10}{n}, $$

where $n = 10$ is the smallest market we consider.

#### 3.3.1 Convergence to stability in uncoordinated matching markets

The most natural way to simulate random sampling of a matching market $\theta \in \Theta_n$ is to sample a preference list for each agent. Since we only consider markets in which every agent prefers being matched to being unmatched, this amounts to sampling a permutation of $\{1, 2, \ldots, n\}$. Sampling is carried out uniformly at random, independently for each agent. Sufficient random sampling in this way should capture most typical instances, but convergence
of a proposal process is best characterized by convergence in the worst cases.\footnote{A natural question is what effect correlated preferences would have on finding a stable matching. It turns out that correlation makes finding a stable matching much easier, because in some sense the correlation coordinates agent behavior.}

How much random sampling is needed to credibly capture the worst cases is unclear in general. In the balanced case, however, we actually know what the worst case instances look like. Consider the class of markets \( \theta \), represented as a weighted graph, in Table C.1. A market can be represented as a weighted graph as follows: let \( \omega(m, w) \in \{1, \ldots, n\} \) denote the weight of edge \( (m, w) \). Then

\[
\begin{align*}
    m \succ_w m' & \iff \omega(m, w) < \omega(m', w) \\
    w \succ_m w' & \iff \omega(m, w) > \omega(m, w')
\end{align*}
\]

Thus, the instances represented in Table C.1 are such that a woman’s favorite man ranks that woman last in his preferences, while a woman’s least favorite man ranks her first; a woman’s second favorite man ranks that woman second to last, and so on.

In such markets, any matching such that every woman is matched to their \( k^{th} \) ranked man is stable. To see this, suppose \( k = 2 \). Then the weight for each pair \( (m, w) \) must be 2, in which case a man \( m \) can only improve his partner’s rank by matching with a woman for whom \( m \) is worse than her current partner. Thus, there are no blocking pairs. Ensuring that each woman is matched to a man just so is what prevents random better and best reply processes from finding stability.

The class of preferences shown in Table C.1 are presented by Ackermann et al. (2011) as instances in which random better and best reply proposal processes can take \( 2^{\Omega(n)} \) steps to converge on a stable matching. Ackermann et al. (2011) do not claim that this class uniquely represents the worst case scenario. However, Hoffman et al. (2013)’s characterization of the time to reach a given stable matching in terms of the size and depth of its jealousy graph shows that such instances indeed are the most problematic for random better (and, by extension, best) reply processes.

The number of proposals needed to reach a stable matching, as a multiple of \( n^3 \), is shown in Figures C.1-C.6. We consider the random best attainable process first.
In random balanced markets, the best attainable process appears to consistently find a stable matching at a small fraction of $n^3$ steps, especially as $n$ grows large. This is illustrated in Figure C.1 for 200 simulations from $n = 10, \ldots, 500$. The effect of adding a single agent on one side is shown in Figure C.2. The pattern is similar to the exactly balanced case, albeit with higher variance, based on 500 simulations over the same range of $n$. However, as shown in Figure C.3, the hard instances prove challenging. Over 150 simulations for $n = 10, \ldots, 300$, the number of proposals needed to find a stable matching is growing in $n$. Note that while the larger instances appear to converge at a reasonable multiple of $n^3$, we cap the simulations at $25n^3$. In other words, the random best attainable process appears to be exploding with $n$.

We now turn to the random singles first process. Figure C.4 shows results of 500 simulations of random balanced environments for $n = 10, \ldots, 500$. Once again, the rate of convergence is a fraction of $n^3$, which is unsurprising as this process is an improvement over the best attainable process. Moreover, the addition of another agent on one side of the market has a negligible impact on convergence, as seen in Figure C.5. With an additional woman in the market, the singles first proposal process still finds a stable matching at a rate less than $n^3$, based again on 500 simulations for $n = 10, \ldots, 500$.

In Figure C.6, we plot convergence for hard instances of a balanced market, for $n = 10, \ldots, 500$, based on 300 simulations. Note that the time to reach stability is still bounded by a small multiple of $n^3$ below approximately $n = 400$. Beyond this amount, potentially exponential growth starts to take over. Still, this is stark improvement over the naive random better and best reply processes where exponential growth is immediately apparent. This is encouraging if we hope moderately large uncoordinated markets reach stability.

### 3.3.2 Welfare of uncoordinated markets prior to reaching stability

While a polynomial number of proposals is much more feasible than an exponential number of proposals, it can still be unrealistic for growth at rates faster than $O(n^2)$. Since the deferred acceptance algorithm can require $O(n)$
proposals from each agent to find a stable match, we should expect a randomized proposal process to take at least as long. If we interpret a proposal as an indication of interest rather than a formal proposal, a person in a market with \( n = 500 \) people on each side could plausibly make 1,000 or 2,000 proposals, corresponding to \( O(n^2) \) proposals total. On the other hand, 250,000 or 500,000 proposals per agent—corresponding to \( O(n^3) \) proposals total—pushes the bounds of believability even with a loose interpretation what counts as a proposal. Is there a downside to agents stopping their search early before reaching stability? In this section, we investigate the welfare of matches found by random proposal processes after a small multiple of \( n^2 \) proposals.

The biggest potential cost of uncoordinated matching is that too many agents are single at a given point in time. The Random Singles First process attempts to address this problem in a greedy fashion by having agents break up an existing couple only when unavoidable. Figures C.8 and C.7 show the proportion of single agents at multiples of \( n^2 \) for balanced and almost balanced markets respectively at \( n = 50 \). Nearly every agent is matched with some partner within the first \( n^2 \) proposals. Figure C.9 of the proportion of single agents in an almost balanced market of \( n = 1000 \) shows this isn’t simply a feature of small markets. Once again, most agents are partnered within the first \( n^2 \) proposals and, after the initial matching, about 5% of agents are single at any one time. Since most agents are matched, it is now reasonable to focus on the welfare of matched agents.

A natural measure of welfare in environments characterized by rank-order preferences is the average rank of agents at a given match. Consider an environment \( \theta_n \) with \( n \) men and \( n + m \) women. Let \( \rho_t(i) \equiv \rho(i, \mu_t(i)) = |\{ j : j \succeq i, \mu_t(i) \}| \) denote the rank of \( i \)'s partner at matching \( \mu_t \). Then the average rank of men’s partners at \( \mu_t \) is \( R^M_t \equiv \frac{1}{n_t} \sum_{i \in M} \rho_t(i) \), where \( n_t = |\{ i \in M : \mu_t(i) \in W \}| \) is the number of men matched to women under \( \mu_t \). If \( \mu_{MOSM} \) is the man-optimal stable matching, the average rank of men at \( \mu_{MOSM} \) is \( R^M_{MOSM} \).

For a given environment \( \theta \), we can easily compute the optimal stable matches \( \mu_{MOSM} \) and \( \mu_{WOSM} \) in order to compare average rank at the current matching of a proposal process. In particular, we will focus on the average
improvement in rank at $\mu_t$ relative to woman-optimal stable matching:

$$Q_t^M \equiv \frac{1}{n_t} \sum_{i \in M} (\rho_{WOSM}(i) - \rho_t(i)) \quad (\text{Men})$$

$$Q_t^W \equiv \frac{1}{n_t} \sum_{i \in W} (\rho_{WOSM}(i) - \rho_t(i)) \quad (\text{Women})$$

$$Q_t^T \equiv \frac{(Q_t^M + Q_t^W)}{2} \quad (\text{Total})$$

The literature has developed a good idea of what the average ranks of optimal matches look like in terms of market size $n$ when preferences are drawn uniformly at random. For balanced $\theta$ with $n$ men and $n$ women, Pittel (1989) has shown that $R_{MOSM}^M \xrightarrow{p} \log n$ and $R_{WOSM}^M \xrightarrow{p} \frac{n}{\log n}$. Ashlagi et al. (forthcoming) explore unbalanced markets and find that the difference in welfare between the man-optimal and woman-optimal matches collapses rapidly with the addition of even a single person. In a market with $n$ men and $n+1$ women, the average rank of the men’s partners is $\log n$ and the average rank of the women’s partners is $\frac{m}{\log n}$ with high probability in every stable match. In other words, being on the short side of the market provides the same advantage as being on the proposing side in a balanced market. Because the addition of a single person drastically changes the set of stable matches, we should view balanced markets as a special case. Without loss of generality, we assume men are on the short side when considering unbalanced markets.

In 500 simulated almost balanced markets with $n$ men and $n+1$ women, all but one match found by the Random Singles First process after $5n^2$ proposals had a better average rank among all matched agents relative to the MOSM. In the sole simulation that didn’t have strictly better average welfare, the process found the unique stable match within the given number of proposals. Relative to the WOSM, 485 simulations had strictly better average ranks, 12 were equal to the WOSM, and 3 were worse. Every time the total average rank was better than the WOSM average rank, men did relatively worse and women did relatively better.

As shown in figure C.10, the short side of the market still has a better average rank than the long side in almost every instance despite doing worse than in any stable match. The short side is unable to fully use its advantage when matching is uncoordinated. Figure C.11 shows how the average rank
of men compares to the WOSM as $n$ changes. Figure C.12 shows the average rank across both men and women relative to the WOSM. Not only is the uncoordinated match is more egalitarian between the two sides relative to stability, it results in an overall better average ranks.

The situation is even more striking when the imbalance between sides grows. We now consider markets with $n$ men on the short side and $1.5n$ women on the long side of the market. Figure C.13 shows the short side of the market is roughly logarithmically worse in the uncoordinated match relative to the WOSM. With 500 men and 750 women, this means the short side is matched with their 2nd or 3rd rank partner rather than their 1st or 2nd rank partner on average. In contrast, a person on the the long side of the market is paired with their 140th best partner rather than their 220th best partner.

These results suggest that decentralized matching can be better for agents overall, particularly if an agent is unsure a priori whether they will be on the long or short side of the market. The evident hardness of finding stability in large coordinated markets ends up being an advantage rather than a failing. Unless a market is obviously failing due to instability or there is a reason all matches should happen quickly and simultaneously, this is a reason to avoid centralization.

### 3.4 Conclusion

We present two proposal processes that suggest uncoordinated two-sided matching markets perform well when agents aren’t completely naive. While our results don’t fully resolve the question of whether uncoordinated markets tend to reach stable matches, either answer turns out to be encouraging. Simulations show the random singles first proposal process reaches stability in small to medium sized markets within a small multiple of $n^3$ proposals. This holds even for the hardest-to-match set of preferences. On the other hand, if the process is cut-off before reaching stability, the resulting matches are more egalitarian and have better average welfare. Either way, our results suggest centralization has no advantage unless the market is unraveling or suffers another clear market failure.

Substantial work remains in the study of decentralized matching markets.
The asymptotic behavior of uncoordinated matching for random preferences or more realistic correlated preferences remains an open question. Another open question is whether a more sophisticated proposal process reaches stability with high probability in polynomial time for all possible preferences.
Appendix A

A.1 Proofs

A.1.1 Proof of Theorem 1.2

Let the set of terms be $T_A = T_B = \{t_L, t_H\}$, with $t_L < t_H$ and $t_L$ strictly preferred to $t_H$ by all agents, all else equal.

Let $Y$ denote the $A$-optimal stable allocation. Suppose $|\{a \in A : (Y^a)_{T_A} = t_L\}| < |\{a \in A : (Y^a)_{T_A} = t_L\}|$, for some stable allocation $Y$. Then there is some agent $\overline{a} \in A$ that has either: (i) no contracts at $\overline{Y}$, or (ii) only $t_H$-contracts at $\overline{Y}$.

Consider case (i). If some $b \in (Y^\overline{a})_B$ has no contracts at $\overline{Y}$ then $(a,b)$ form a blocking pair for $\overline{Y}$. So all $b \in (Y^\overline{a})_B$ have contracts with some side $A$ agents in $\overline{Y}$. Moreover, these contracts must be preferred to those in $Y$ or they can block $\overline{Y}$ with $\overline{a}$. But this contradicts $\overline{Y}$ as the $B$-pessimal stable allocation.

Consider case (ii). Let $y \in Y^{\overline{a}}$ such that $y_{T_A} = t_L$. Suppose $y_B \in (Y^{\overline{a}})_B$, and let $\overline{y} \in Y^{\overline{a}}$ be the associated contract. Suppose $y_I = \overline{y}_I$. By the assumption that all agents prefer $t_L$-contracts to $t_H$-contracts, $y \in C_\pi(Y \cup \overline{Y}) \neq Y^{\overline{a}}$, regardless of $y_{T_B}, \overline{y}_{T_B}$. This contradicts $A$-optimality of $\overline{Y}$. So $y_I \neq \overline{y}_I$, and there exists $z \in X^{\overline{a}}$ that coincides with $\overline{y}$ except $z_{T_A} = t_L$. Then $\overline{a}$ and $y_B$ mutually prefer $z$ to $\overline{y}$, contradicting stability of $\overline{Y}$.

Therefore $y_B \in Y^{\overline{a}} \Rightarrow y_B \not\in (Y^{\overline{a}})_B$. Moreover, such agents $y_B$ must have contracts in $\overline{Y}$ or they can form a blocking pair for $\overline{Y}$ with $\overline{a}$. Thus such agents $y_B$ have contracts with different side-$A$ agents at $\overline{Y}$. These contracts are preferred by $y_B$ to $y$, otherwise they can block $\overline{Y}$ with $\overline{a}$. But this contradicts $\overline{Y}$ as $B$-pessimal stable.

An analogous argument shows that the number of side-$B$ agents that sign
$t_H$-contracts at $Y$ is weakly higher than at any other stable allocation. \qed

A.1.2 Proof of Theorem 1.3

Consider an allocation $X'$. Without loss of generality, assume agents are associated with two platforms $i$ and $i'$ such that $|\{a \in A : (X'^a)_I = i \text{ and } (X'^a)_T = t\}| > |\{a \in A : (X'^a)_I = i' \text{ and } (X'^a)_T = t\}| > 0$. Consider an agent $b \in B$ such that $(X'^b)_T = t$. Suppose $(X'^b)_I \neq i$. Let the set of side $A$ agents whose term is $t$ and who are associated with platform $i$ be $A'$. Let $Y = \{x \in X : x_B = b, x_A \in A', \text{ and } x_I = i\}$. Then $Y' = X'_A' \cup Y$ forms a blocking set of contracts.

Therefore, $X'$ is not a stable allocation. That is, there cannot be a stable allocation in which one platform $i$ has more side-$A$ agents signing $t$ contracts than another platform $i'$. Similarly, there cannot be a stable allocation in which platform $i$ has more side-$B$ agents signing $t$ contracts than platform $i'$. Then either $i$ has all agents who sign $t$ contracts, or $i$ and $i'$ have the same number of agents.

Obviously, if all agents signing $t$ contracts are at platform $i$, this is stable. Suppose $i$ and $i'$ have the same number of agents signing $t$ contracts. By the law of aggregate demand for similar terms, the set of agents who join $i'$ prefer the allocation in which all agents join $i$. So this cannot be stable. \qed

A.1.3 Proof of Theorem 1.5

The proof below incorporates the following algorithm, an extension of Hatfield and Milgrom (2005)'s cumulative offer algorithm to the many-to-many setting introduced in Fung and Hsu (2014).

The many-to-many cumulative offer algorithm

The (side $A$-proposing) many-to-many cumulative offer algorithm works as follows.

- Step 1: Let $O_a(1) = X_a$ for all $a \in A$. An arbitrary side $A$ agent $a_1$ proposes $Y(1) = C_{a_1}(O_{a_1}(1))$. Side $B$ agents hold $C_b(Y_b(1))$ for all $b \in B$. 

Let $O_b(1) = Y_b(1)$ for all $b \in B$.

Let $O_B(1) = \cup_{b \in B} O_b(1) = Y(1)$.

• Step $t$: Let $O_a(t) = X_a - R_B(O_B(t - 1))$ for all $a \in A$. An arbitrary side $A$ agent $a_t$ such that $C_{a_t}(O_{a_t}(t)) \not\subset C_B(O_B(t - 1))$ proposes $Y(t) = C_{a_t}(O_{a_t}(t))$. Hospitals hold $C_b(O_b(t - 1) \cup Y_b(t))$ for all $b \in B$.

- Let $O_b(t) = O_b(t - 1) \cup Y_b(t)$ for all $b \in B$.
- Let $O_B(t) = \cup_{b \in B} O_b(t) = O_B(t - 1) \cup Y(t)$.

The algorithm terminates in some step $T$ such that $C_a(O_a(T)) \subset C_B(O_B(T - 1))$ for all $a \in A$. The final allocation is $X' \equiv C_A(O_A(T))$.

In Fung and Hsu (2014), we show the algorithm produces a stable allocation in unitary many-to-many matching with contracts markets, under substitutes as well as under weaker preference assumptions. We now prove Theorem 1.5.

Proof. Suppose the preferences are lexicographic. Consider the following construction.

For any $X$, construct a reduced problem $\bar{X}$ such that for any $\bar{x} \in \bar{X}$ we have

$$\bar{x} = (a, b, i),$$

where $a \in A$, $b \in B$ and $i \in I$. For any agent $a \in A$, we can construct a reduced preference $\bar{P}_a$ such that if $x \bar{P}_a \bar{x}'$ for any $\bar{x}, \bar{x}' \in \bar{X}$, then $x P_a x'$ for some $x, x' \in X$, where $\bar{x}_A = \bar{x}'_A = x_A = x'_A = a$, $\bar{x}_B = x_B$, $\bar{x}'_B = x'_B$, $\bar{x}_I = x_I$, and $\bar{x}'_I = x'_I$.

We use the following algorithm to produce a term-constrained stable allocation. There are three steps.

- Step 1: For the reduced problem $\bar{X}$ and reduced preferences $\bar{P}_{s \in A \cup B}$, run the many-to-many cumulative offere algorithm. The outcome is $\bar{X}'$.

- Step 2: For each platform $i$ that has a positive number of agents associated with it, let the side $B$ agent with the smallest index, say $b$, chooses his favorite term among $X$. In other words, let the term associated with a platform $i$ be $(C_b(X))_{T_A, T_B}$. 

53
For each platform $i$ that has a positive number of agents associated with it, let $i(t_A, t_B)$ be the term that is produced by the above method.

- Step 3: Let $X' \subseteq X$ be such that

  1. $(X'^{a})_B = (\bar{X}^{a})_B$, $(X'^{a})_I = (\bar{X}^{a})_I := i$, and $(X'^{a})_{T_A, T_B} = i(t_A, t_B), \forall a \in X'_A$;

  2. $(X'^{b})_A = (\bar{X}^{b})_A$, $(X'^{b})_I = (\bar{X}^{b})_I := j$, and $(X'^{b})_{T_A, T_B} = j(t_A, t_B), \forall b \in X'_B$.

We claim that $X'$ is a term-constrained stable allocation. Note that if the preference is not substitutable nor lexicographic, the algorithm may not produce a feasible allocation.

In the following, we show that $X'$ is term-constrained stable.

1. The allocation $X'$ is feasible: By lexicographic preference, for any allocation $\bar{X}' \subseteq \bar{X}$, if we create an allocation $Y \subseteq X$ by assigning some terms $(t_A, t_B)$ to each platform $i$, so that for agents associated with platform $i$, they also have the terms in their contracts, then the allocation $Y$ is feasible. This is precisely what the three-step algorithm does.

2. The allocation $X'$ is IR: Implied by lexicographic preference, for any allocation $\bar{X}' \subseteq \bar{X}$, if we create an allocation $Y \subseteq X$ by assigning some terms $(t_A, t_B)$ to each platform $i$, so that for agents associated with platform $i$, they also have the terms in their contracts, then the allocation $Y$ is IR. Therefore, $X'$ is IR.

3. The allocation does not have feasible blocking set: Since preferences are substitutable, $\bar{X}'$ is a stable allocation in the reduced problem. In other words, no agents would jointly change the contracts so that they can switch platforms or be matched with different agents in the reduced problem.

Since preferences are lexicographic, for each agent, the set of agents on the other side and the platform in the contracts are more important than the terms. No agents would jointly change the contracts so that they can switch platforms or be matched with different agents. Within a platform $i$, the term is chosen by the side $B$ agent with the smallest index. In other words, agents associated with $i$ cannot jointly change the contracts so that they can have different terms in their contracts, since there is at least one agent who would disagree.
Based on the above argument, the allocation $X'$ is term-constrained stable allocation.
Appendix B

B.1 Figures

B.1.1 Elementary school attendance areas in Champaign-Urbana

Figure B.1: Champaign and Urbana attendance areas, circa 1998. Points represent schools.
Figure B.2: Champaign attendance areas, circa 1989-1998. Points represent schools corresponding to attendance area.
Figure B.3: Urbana attendance areas, circa 1989-2002. Points represent schools corresponding to attendance area.

B.1.2 Timeline of policy changes

Figure B.4: Timeline of assignment policy changes
B.1.3 Descriptives

Figure B.5: Distribution of log sale price, before (left) and after (right) the policy change.
Figure B.6: Median log sale price over time.
B.2 Tables

B.2.1 Descriptives
<table>
<thead>
<tr>
<th>School District</th>
<th>Pre Choice</th>
<th>Post Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td><strong>Champaign</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Sale Price</td>
<td>4.397</td>
<td>4.407</td>
</tr>
<tr>
<td>Sq.Ft</td>
<td>1.783</td>
<td>1.652</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>3.324</td>
<td>3.609</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>2.054</td>
<td>2.000</td>
</tr>
<tr>
<td>Basement</td>
<td>1.060</td>
<td>1.044</td>
</tr>
<tr>
<td>Garage</td>
<td>0.974</td>
<td>1.000</td>
</tr>
<tr>
<td>Age.Home</td>
<td>29.612</td>
<td>27.921</td>
</tr>
<tr>
<td>Neighborhood.School.Distance</td>
<td>1.155</td>
<td>0.661</td>
</tr>
<tr>
<td><strong>Urbana</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log Sale Price</td>
<td>4.229</td>
<td>4.190</td>
</tr>
<tr>
<td>Lotsize</td>
<td>17.485</td>
<td>8.820</td>
</tr>
<tr>
<td>Sq.Ft</td>
<td>1.493</td>
<td>1.344</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>3.055</td>
<td>3.000</td>
</tr>
<tr>
<td>Bathrooms</td>
<td>1.680</td>
<td>1.914</td>
</tr>
<tr>
<td>Basement</td>
<td>1.097</td>
<td>1.039</td>
</tr>
<tr>
<td>Garage</td>
<td>0.940</td>
<td>1.000</td>
</tr>
<tr>
<td>Age.Home</td>
<td>43.385</td>
<td>38.093</td>
</tr>
<tr>
<td>Neighborhood.School.Distance</td>
<td>0.815</td>
<td>0.481</td>
</tr>
</tbody>
</table>

Table B.1: Descriptive statistics by school district: housing
<table>
<thead>
<tr>
<th>School District</th>
<th>Pre</th>
<th>Post</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>Median</td>
</tr>
<tr>
<td>Champaign</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Households</td>
<td>1789.828</td>
<td>1875.000</td>
</tr>
<tr>
<td>Median.Household.Income</td>
<td>50912.986</td>
<td>51896.000</td>
</tr>
<tr>
<td>Median.Rooms</td>
<td>5.738</td>
<td>5.600</td>
</tr>
<tr>
<td>Owner.Occupied.Units</td>
<td>1026.257</td>
<td>966.000</td>
</tr>
<tr>
<td>Owner.Occupied.Median.Value</td>
<td>114163.811</td>
<td>105100.000</td>
</tr>
<tr>
<td>Urbana</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Households</td>
<td>2577.355</td>
<td>2583.000</td>
</tr>
<tr>
<td>Median.Household.Income</td>
<td>35273.957</td>
<td>35714.000</td>
</tr>
<tr>
<td>Median.Rooms</td>
<td>5.041</td>
<td>5.100</td>
</tr>
<tr>
<td>Owner.Occupied.Units</td>
<td>1103.004</td>
<td>1167.000</td>
</tr>
<tr>
<td>Owner.Occupied.Median.Value</td>
<td>90737.748</td>
<td>87600.000</td>
</tr>
</tbody>
</table>

Table B.2: Descriptive statistics by school district: Census tract level
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Champaign</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mean</td>
<td>4.3875</td>
<td>4.3875</td>
<td>0.0228</td>
<td>4.1975</td>
<td>4.1883</td>
<td>0.0804</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School Tax Rate</td>
<td></td>
<td>4.3875</td>
<td>4.3875</td>
<td>0.0228</td>
<td>4.1975</td>
<td>4.1883</td>
<td>0.0804</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class.Size</td>
<td>20.4050</td>
<td>20.6000</td>
<td>2.4631</td>
<td>21.5114</td>
<td>22.3000</td>
<td>2.7105</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low.Income</td>
<td>39.2150</td>
<td>33.0500</td>
<td>16.1234</td>
<td>43.8886</td>
<td>41.1500</td>
<td>17.6696</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrollment</td>
<td>433.1500</td>
<td>450.5000</td>
<td>86.5012</td>
<td>392.7500</td>
<td>399.0000</td>
<td>78.7513</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Urbana</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School Tax Rate</td>
<td>4.9502</td>
<td>4.9502</td>
<td>0.0547</td>
<td>4.7877</td>
<td>4.7644</td>
<td>0.0684</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class.Size</td>
<td>17.6500</td>
<td>18.3000</td>
<td>2.6976</td>
<td>18.0917</td>
<td>17.7000</td>
<td>3.3221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low.Income</td>
<td>46.2000</td>
<td>45.9500</td>
<td>13.4631</td>
<td>49.5708</td>
<td>47.0500</td>
<td>15.2922</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enrollment</td>
<td>394.5000</td>
<td>392.0000</td>
<td>42.5836</td>
<td>352.9583</td>
<td>350.0000</td>
<td>36.5531</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Excel.Reading</td>
<td>25.0833</td>
<td>27.0000</td>
<td>11.5165</td>
<td>20.1583</td>
<td>20.1500</td>
<td>7.4026</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table B.3: Descriptive statistics by school district: schools
<table>
<thead>
<tr>
<th>School District</th>
<th>Pre-</th>
<th>Post-</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Champaign</td>
<td>2766</td>
<td>6371</td>
<td>9137</td>
</tr>
<tr>
<td>Urbana</td>
<td>959</td>
<td>2308</td>
<td>3267</td>
</tr>
<tr>
<td>All</td>
<td>3725</td>
<td>8679</td>
<td>12404</td>
</tr>
</tbody>
</table>

Table B.4: Number of observations by “treatment” and “control”
### Results

|                         | Estimate | Std. Error | t value | Pr(>|t|) |
|-------------------------|----------|------------|---------|----------|
| (Intercept)             | -0.7444  | 1.2279     | -0.61   | 0.5444   |
| Sqft                    | 0.0005   | 0.0000     | 32.19   | 0.0000   |
| Lotsize                 | 0.0001   | 0.0001     | 1.16    | 0.2449   |
| Age Home                | -0.0018  | 0.0003     | -5.49   | 0.0000   |
| Garage 1                | 0.2341   | 0.0253     | 9.24    | 0.0000   |
| Basement                | -0.0001  | 0.0000     | -5.01   | 0.0000   |
| Bedrooms                | -0.0440  | 0.0107     | -4.12   | 0.0000   |
| Baths Total             | 0.1077   | 0.0107     | 10.09   | 0.0000   |
| Agency Name URBANA SD 116 | -0.4804   | 0.1593     | -3.02   | 0.0026   |
| Neighborhood School Distance | -0.0803   | 0.0065     | -12.27  | 0.0000   |
| Median Household Income | 0.0000   | 0.0000     | 10.91   | 0.0000   |
| Median Rooms            | -0.0720  | 0.0112     | -6.41   | 0.0000   |
| Owner Occupied Units    | 0.0001   | 0.0000     | 5.69    | 0.0000   |
| Owner Occupied Median Value | -0.0000   | 0.0000     | -4.03   | 0.0001   |
| Year 1997               | 0.0889   | 0.0266     | 3.35    | 0.0008   |
| Year 1998               | 0.1116   | 0.0346     | 3.23    | 0.0013   |
| Year 1999               | 0.2015   | 0.0518     | 3.89    | 0.0001   |
| Year 2000               | 0.3738   | 0.0767     | 4.87    | 0.0000   |
| Year 2001               | 0.4094   | 0.0797     | 5.14    | 0.0000   |
| PIN                     | 0.0000   | 0.0000     | 0.42    | 0.6773   |
| School Tax Rate         | 0.9815   | 0.2779     | 3.53    | 0.0004   |
| Class Size              | -0.0130  | 0.0023     | -5.62   | 0.0000   |
| Math                    | 0.0052   | 0.0006     | 8.49    | 0.0000   |
| Reading                 | 0.0011   | 0.0006     | 1.76    | 0.0783   |
| Buyer Origin Other      | -0.2554  | 0.0141     | -18.16  | 0.0000   |
| Buyer Origin URBANA SD 116 | -0.0673   | 0.0186     | -3.62   | 0.0003   |
| d.obs                   | 0.0682   | 0.0178     | 3.83    | 0.0001   |

Table B.5: Estimates for DD hedonic model
<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.0077</td>
<td>3.1098</td>
</tr>
<tr>
<td></td>
<td>(0.2253)</td>
<td>(0.2215)</td>
</tr>
<tr>
<td>Sqft</td>
<td>0.0005</td>
<td>0.0005</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Lotsize</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0001)</td>
<td>(0.0001)</td>
</tr>
<tr>
<td>Age.Home</td>
<td>-0.0017</td>
<td>-0.0017</td>
</tr>
<tr>
<td></td>
<td>(0.0003)</td>
<td>(0.0003)</td>
</tr>
<tr>
<td>Garage1</td>
<td>0.2335</td>
<td>0.2324</td>
</tr>
<tr>
<td></td>
<td>(0.0254)</td>
<td>(0.0253)</td>
</tr>
<tr>
<td>Basement</td>
<td>-0.0001</td>
<td>-0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>-0.0442</td>
<td>-0.0440</td>
</tr>
<tr>
<td></td>
<td>(0.0107)</td>
<td>(0.0107)</td>
</tr>
<tr>
<td>Baths.Total</td>
<td>0.1083</td>
<td>0.1080</td>
</tr>
<tr>
<td></td>
<td>(0.0107)</td>
<td>(0.0107)</td>
</tr>
<tr>
<td>Neighborhood.School.Distance</td>
<td>-0.0794</td>
<td>-0.0800</td>
</tr>
<tr>
<td></td>
<td>(0.0065)</td>
<td>(0.0065)</td>
</tr>
<tr>
<td>Median.Household.Income</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Median.Rooms</td>
<td>-0.0718</td>
<td>-0.0721</td>
</tr>
<tr>
<td></td>
<td>(0.0112)</td>
<td>(0.0112)</td>
</tr>
<tr>
<td>Owner.Occupied.Units</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>Owner.Occupied.Median.Value</td>
<td>-0.0000</td>
<td>-0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>PIN</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>(0.0000)</td>
<td>(0.0000)</td>
</tr>
<tr>
<td>School.Tax.Rate</td>
<td>0.1462</td>
<td>0.1247</td>
</tr>
<tr>
<td></td>
<td>(0.0438)</td>
<td>(0.0428)</td>
</tr>
<tr>
<td>Class.Size</td>
<td>-0.0125</td>
<td>-0.0126</td>
</tr>
<tr>
<td></td>
<td>(0.0023)</td>
<td>(0.0023)</td>
</tr>
</tbody>
</table>

Table B.6: Estimates for mixed-effects DD hedonic model
<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Math</td>
<td>0.0053</td>
<td>0.0053</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>Reading</td>
<td>0.0010</td>
<td>0.0011</td>
</tr>
<tr>
<td></td>
<td>(0.0006)</td>
<td>(0.0006)</td>
</tr>
<tr>
<td>Buyer.OriginOther</td>
<td>−0.2547</td>
<td>−0.2552</td>
</tr>
<tr>
<td></td>
<td>(0.0141)</td>
<td>(0.0141)</td>
</tr>
<tr>
<td>Buyer.OriginURBANA SD 116</td>
<td>−0.0660</td>
<td>−0.0652</td>
</tr>
<tr>
<td></td>
<td>(0.0185)</td>
<td>(0.0185)</td>
</tr>
<tr>
<td>d.obs</td>
<td>0.0580</td>
<td>0.0562</td>
</tr>
<tr>
<td></td>
<td>(0.0172)</td>
<td>(0.6206)</td>
</tr>
<tr>
<td>AIC</td>
<td>17320.7928</td>
<td>17324.0587</td>
</tr>
<tr>
<td>BIC</td>
<td>17499.0114</td>
<td>17531.9804</td>
</tr>
<tr>
<td>Log Likelihood</td>
<td>-8636.3964</td>
<td>-8634.0294</td>
</tr>
<tr>
<td>Num. obs.</td>
<td>12404</td>
<td>12404</td>
</tr>
<tr>
<td>Num. groups: Year</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Num. groups: Agency.Name</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Var: Year (Intercept)</td>
<td>0.0049</td>
<td>0.0068</td>
</tr>
<tr>
<td>Var: Agency.Name (Intercept)</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Var: Residual</td>
<td>0.2297</td>
<td>0.2296</td>
</tr>
<tr>
<td>Var: Year d.obs</td>
<td></td>
<td>0.0007</td>
</tr>
<tr>
<td>Cov: Year (Intercept) d.obs</td>
<td>-0.0022</td>
<td></td>
</tr>
<tr>
<td>Var: Agency.Name d.obs</td>
<td></td>
<td>0.3847</td>
</tr>
<tr>
<td>Cov: Agency.Name (Intercept) d.obs</td>
<td>0.0000</td>
<td></td>
</tr>
</tbody>
</table>

Table B.6 (cont.)
<table>
<thead>
<tr>
<th>Feature</th>
<th>Lasso estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>3.5313</td>
</tr>
<tr>
<td>Sqft</td>
<td>0.0005</td>
</tr>
<tr>
<td>Lotsize</td>
<td>0.0001</td>
</tr>
<tr>
<td>Age.Home</td>
<td>-0.0016</td>
</tr>
<tr>
<td>Garage1</td>
<td>0.2328</td>
</tr>
<tr>
<td>Basement</td>
<td>-0.0001</td>
</tr>
<tr>
<td>Bedrooms</td>
<td>-0.0410</td>
</tr>
<tr>
<td>Baths.Total</td>
<td>0.1073</td>
</tr>
<tr>
<td>Agency.NameURBANA SD 116</td>
<td>0.0660</td>
</tr>
<tr>
<td>Neighborhood.School.Distance</td>
<td>-0.0780</td>
</tr>
<tr>
<td>Median.Household.Income</td>
<td>0.00001</td>
</tr>
<tr>
<td>Median.Rooms</td>
<td>-0.0643</td>
</tr>
<tr>
<td>Owner.Occupied.Units</td>
<td>0.0001</td>
</tr>
<tr>
<td>Owner.Occupied.Median.Value</td>
<td>-0.000001</td>
</tr>
<tr>
<td>Year1997</td>
<td>0.0122</td>
</tr>
<tr>
<td>Year1998</td>
<td>0.0217</td>
</tr>
<tr>
<td>Year2000</td>
<td>0.0995</td>
</tr>
<tr>
<td>Year2001</td>
<td>0.1241</td>
</tr>
<tr>
<td>PIN</td>
<td>0.0041</td>
</tr>
<tr>
<td>School.Tax.Rate</td>
<td>-0.0120</td>
</tr>
<tr>
<td>Class.Size</td>
<td>0.0054</td>
</tr>
<tr>
<td>Math</td>
<td>0.0008</td>
</tr>
<tr>
<td>Reading</td>
<td>-0.2514</td>
</tr>
<tr>
<td>Buyer.OriginOther</td>
<td>-0.0570</td>
</tr>
<tr>
<td>Buyer.OriginUrbana SD 116</td>
<td>0.0495</td>
</tr>
<tr>
<td>d.obs</td>
<td></td>
</tr>
</tbody>
</table>

Table B.7: Estimates for regularized DD hedonic model
Appendix C

C.1 Figures

Figure C.1: Number of proposals to reach stability in Random Best Attainable process in balanced random environment
Figure C.2: Number of proposals to reach stability in Random Best Attainable process in almost balanced random environment

Figure C.3: Number of proposals to reach stability in Random Best Attainable process in hard environment
Figure C.4: Number of proposals to reach stability in Random Singles First process in balanced random environment

Figure C.5: Number of proposals to reach stability in Random Singles First process in almost balanced random environment
Figure C.6: Number of proposals to reach stability in Random Singles First process in hard environment

Figure C.7: Proportion single after multiple of proposal in almost balanced environment for $n = 50$. 
Figure C.8: Proportion of single agents after multiple of proposal in balanced environment for $n = 50$.

Figure C.9: Proportion of single agents after multiple of proposal in almost balanced environment for $n = 1000$. 
Figure C.10: Average ranks for $n$ men and $n + 1$ women after $5n^2$ proposals.

Figure C.11: Men’s average rank relative to the WOSM with $n$ men and $n + 1$ women after $5n^2$ proposals.
Figure C.12: Total average rank relative to the WOSM with $n$ men and $n + 1$ women after $5n^2$ proposals.

Figure C.13: Men’s average rank relative to the WOSM with $n$ men and $1.5n$ women after $5n^2$ proposals.
Figure C.14: Total average rank relative to the WOSM with $n$ men and $1.5n$ women after $5n^2$ proposals.
C.2 Tables

<table>
<thead>
<tr>
<th></th>
<th>$m_1$</th>
<th>$m_2$</th>
<th>$m_3$</th>
<th>...</th>
<th>$m_{n-2}$</th>
<th>$m_{n-1}$</th>
<th>$m_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w_1$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>$n-2$</td>
<td>$n-1$</td>
<td>$n$</td>
</tr>
<tr>
<td>$w_2$</td>
<td>$n$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>$n-2$</td>
<td>$n-1$</td>
</tr>
<tr>
<td>$w_3$</td>
<td>$n-1$</td>
<td>$n$</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>...</td>
<td>$n-2$</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>$w_{n-1}$</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>...</td>
<td>$n$</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>$w_n$</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>...</td>
<td>$n-1$</td>
<td>$n$</td>
<td>1</td>
</tr>
</tbody>
</table>

Table C.1: An instance of hard preferences in balanced markets, reproduced from Ackermann et al. (2011).


