

THE UNIVERSITY OF CHICAGO

ESSAYS ON MATCHING, EFFICIENCY, AND OPTIMAL SOCIAL POLICY

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE IRVING B. HARRIS
GRADUATE SCHOOL OF PUBLIC POLICY STUDIES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

BY
JONATHAN MARTIN VILLARS DAVIS

CHICAGO, ILLINOIS

JUNE 2016

Copyright © 2016 by Jonathan Martin Villars Davis
All Rights Reserved

To my family

TABLE OF CONTENTS

LIST OF FIGURES	vi
LIST OF TABLES	vii
ACKNOWLEDGMENTS	viii
ABSTRACT	ix
1 DEFERRED ACCEPTANCE MECHANISMS CAN IMPROVE MATCH QUALITY: QUASI-EXPERIMENTAL EVIDENCE FROM A TEACH FOR AMERICA PILOT	1
1.1 Introduction	1
1.2 The Setting	6
1.2.1 Teach for America	6
1.2.2 The Chicago Mechanism	7
1.2.3 The Deferred Acceptance Algorithm	13
1.3 Measuring the Impact of Adopting the Deferred Acceptance Algorithm	17
1.3.1 Quasi-Experimental Design	17
1.3.2 Data	18
1.3.3 Estimation and Inference	19
1.4 The Effect of Adopting the DAA on Teacher Retention	22
1.4.1 Retention through the Start of the First School Year	23
1.4.2 Retention through the End of the First School Year	27
1.4.3 How valuable are these impacts?	29
1.5 The Effect of Adopting the DAA on Unraveling and Interview Day Hiring	30
1.6 Conclusion	42
2 RETHINKING THE BENEFITS OF EMPLOYMENT PROGRAMS: THE HET- EROGENEOUS EFFECTS OF SUMMER JOBS (W/ SARA B. HELLER)	44
2.1 Introduction	44
2.2 Program Description	47
2.2.1 Summer 2012	48
2.2.2 Summer 2013	49
2.3 Experimental Design	50
2.3.1 Summer 2012	50
2.3.2 Summer 2013	51
2.4 Data	52
2.5 Analytical Methods	54
2.6 Descriptive Statistics	58
2.7 Participation	60
2.8 Main Results	64
2.9 Who Benefits From Summer Jobs?	74
2.10 Conclusion	85

3	DESIGNING ORGANIZATIONAL VERSUS PUBLIC MARKETS (W/ B. PABLO MONTAGNES)	88
3.1	Introduction	88
3.2	Model	90
3.3	Results	93
3.3.1	Aligned Preferences	96
3.3.2	Imperfect Information	98
3.3.3	Outside Options	100
3.3.4	More General Solutions	102
3.4	Conclusion	104
	REFERENCES	105

LIST OF FIGURES

1.1	First Round Thresholds	12
1.2	School Schedule and Ranking Form	15
1.3	Trends in Retention Through Start of First School Year	24
1.4	Trends in Retention Through End of First School Year	28
1.5	Actual and Counterfactual Distribution of Hires By Round	34
1.6	Results of Pairwise Teacher Interview Competitions	36
1.7	Teacher and School Ranks Over Matches	37
1.8	Change in Teacher Preferences Over Matches	40
1.9	Change in Teacher Preferences Over Matches	41
2.1	Density of Causal Forest Year One Violent Crime Estimates	80
2.2	Causal Forest Point Estimates with 95% Confidence Intervals	84

LIST OF TABLES

1.1	Characteristics of Two Potential Schools	5
1.2	Market Outcomes in Example 1 under the Chicago and DAA Mechanisms, R=2	13
1.3	National Sample Summary Statistics	20
1.4	The Impact of DAA on Retention	26
1.5	Correlates of Pre-DAA Attrition	31
1.6	Impact of DAA on Hires by Round	33
1.7	Interview Day Attendance and Hiring	38
1.8	Impact of DAA on Interview Day Outcomes	39
2.1	Baseline Balance	59
2.2	Participation	60
2.3	Program Participation and Other Employment	63
2.4	Impact of One Summer Plus 2012 and 2013 on Arrests	65
2.5	Impact of One Summer Plus 2012 and 2013 on Formal Employment	68
2.5	Impact of One Summer Plus 2012 and 2013 on Formal Employment - Continued	69
2.6	Impact of One Summer Plus 2012 and 2013 on Schooling	71
2.6	Impact of One Summer Plus 2012 and 2013 on Schooling - Continued	72
2.7	Estimated Social Savings per Participant from Crime Reduction	75
2.8	Heterogeneity in Impacts of One Summer Plus 2012 and 2013 on Year One Crime	78
2.9	Selected Partial Correlations with Conditional Average Treatment Effects	82
2.9	Selected Partial Correlations - Continued	83
2.10	Average Effects Under Different Eligibility Criteria	86

ACKNOWLEDGMENTS

First and foremost, I am indebted to my wife, Kimberlee Pelster, and my parents, Becke and Marty Davis. This dissertation would not have been possible without their incredible support.

I thank my committee, Jens Ludwig, Kerwin Charles, and Bob LaLonde, and Scott Kominers, for their exceptional mentorship. I am especially grateful to my chair, Jens Ludwig, for generously providing many rounds of constructive feedback. I have also benefited tremendously from working with Dan Aaronson, Dan Alexander, Dan Black, Stephane Bonhomme, Alec Brandon, Sara Heller, Ariel Kalil, Bhash Mazumder, Pablo Montagnes, and the rest of the faculty of the Harris School.

I gratefully acknowledge the generous support of graduate research fellowship 2014-IJ-CX-0011 from the National Institute of Justice.

Chapter 1 would not have been possible without the support of the Teach for America Chicago-Northwest Indiana Region, especially the Districts and School Partnerships team, and the Teach for America national research team. Karen Ott, Soukrpida Phetmisy, and Yoon Ha Choi in particular provided exceptional support. This research benefited from useful feedback from Ioana Marinescu and seminar participants at the Federal Reserve Bank of Chicago.

The research in Chapter 2 was generously supported by award B139634411 from the U.S. Department of Labor and grant 2012-MIJ-FX-0002 from the Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. We thank Chicago Public Schools, the Chicago Police Department, and the Department of Family and Support Services for providing the data used for this analysis.

Chapter 3 benefited from very useful feedback from John Hatfield, Fuhito Kojima, and Alex Teytelboym.

Any errors are my own.

ABSTRACT

The first chapter of this dissertation provides the most rigorous quasi-experimental evidence to date of the impact of deferred acceptance algorithms on match quality. I worked with Teach For America (TFA) to match high school teachers to schools in Chicago using the DAA at a series of "interview days", while keeping the mechanism for matching elementary school teachers unchanged. I show that TFA's original interview day mechanism promotes strategic early hiring in theory and in practice. I estimate the effect of adopting the DAA using a difference-in-difference strategy - comparing changes in teacher retention rates over time for TFA high school teachers to the change among TFA elementary school teachers in Chicago to the same changes in four other TFA regions. Adopting a variant of the DAA reduces attrition through the start and end of teachers' first school year by 7 and 8 percentage points, respectively. These effects are arguably a lower bound for other markets because substantial heterogeneity in schools' preferences over teachers reduced inefficiency prior to the intervention. Achieving similar retention increases via higher salaries would cost nearly \$3.2 million.

The second chapter, which is co-authored with Dr. Sara Heller, uses medium-term results from two randomized controlled trials in Chicago to better understand why summer jobs reduce violence and for whom such programs work best. We see almost identical decreases in arrests for violent crime- equal to 5 arrests per 100 youth- in both the initial and follow up study, but we find no effect on other types of crime. We find limited effects on schooling, including a marginally significant reduction in graduation two years after the program. We find an increase in post-program formal employment with the program providers, but not among other employers. One hypothesis for the program's success is that it targets many youth prior to school exit, acting as unemployment prevention rather than remediation. We use recent developments in supervised machine learning to explore which subgroups' violent crime rates are most responsive to the intervention. We find that criminally involved males who are still enrolled in school but have poor attendance and minimally criminally involved

females early in their high school career benefit most. We find that more disconnected males who are older, more criminally involved and do not have recent formal work experience are most adversely affected by the program. As this group is often the target of similar social programs, this has potentially important implications about optimally targeting similar “light touch” social programs.

The third chapter, which is co-authored with Dr. B. Pablo Montagnes, develops theory to demonstrate matching problems within an organization are distinct from traditional applications in public markets. Organizational assignment problems are constrained by individual rationality so an organization may select any assignment that is acceptable to its members. We show that there are no guarantees that assignment mechanisms that respect preferences will perform well from an organization’s perspective. In some cases, an organization can better achieve its objectives by ignoring preferences and randomly choosing assignments, even when market participants have preferences aligned with the organization’s, have outside options, and have private information about match qualities.

CHAPTER 1

**DEFERRED ACCEPTANCE MECHANISMS CAN IMPROVE
MATCH QUALITY: QUASI-EXPERIMENTAL EVIDENCE
FROM A TEACH FOR AMERICA PILOT**

1.1 Introduction

Centralized clearinghouses using variants of the Deferred Acceptance Algorithm (DAA) [30, 62] have been adopted in a number of markets in order to improve market outcomes. Most famously, in the 1940s the market for medical interns unraveled to the point that hospitals were hiring medical students two years before their graduation from medical school [63]. This early matching was bad for hospitals who would have preferred to hire medical students with more complete academic records and for medical students who often had to commit to a position before knowing all of their options. This unraveling ended in the 1950s after the National Residency Matching Program (NRMP), which matches new doctors to hospitals using a variant of the DAA, was adopted.

The DAA is often able to prevent market unraveling because it yields a stable match in the sense that no unmatched doctor and hospital would prefer to be matched together over their assigned match. [61] shows stability is an important feature of successful centralized mechanisms. He compares the longevity of centralized clearinghouses using stable and unstable mechanisms across seven regional markets for doctors in the United Kingdom and finds that the stable clearinghouses almost always outlast the unstable clearinghouses.

Even when stability is not a concern, centralized clearinghouses often adopt mechanisms related to DAA because it is strategy-proof for the side of the market making proposals when each proposing agent has unit demand. Boston Public Schools began matching its 60,000 students to schools using a variant of the DAA in the mid-2000s despite its task force originally recommending the distinct Top Trading Cycles mechanism [2]. A deciding factor

was DAA's ability to "level the playing field" by removing opportunities to improve one's match by strategically misrepresenting their preferences over schools [58].

Adopting a centralized clearinghouse is difficult because its success hinges on the voluntary participation of the vast majority of the market. In general, a centralized clearinghouse will create winners and losers. Incumbents who were benefiting from the original market mechanism may put forth substantial resources to prevent the clearinghouses' adoption. Federal appellate judges have tried to reform the market for their clerks on several occasions to no avail [10, 11, 6]. Even the successful re-design of the NRMP in the 1990s was threatened by litigation claiming the mechanism suppressed medical residents' wages [55]. Moreover, decentralized markets may be preferable for normative reasons [29]. Consequently, credible empirical evidence about the actual efficiency gains created by stable clearinghouses is critical for understanding whether their benefits justify the costs.

Unfortunately, the empirical evidence of the effects of adopting stable mechanisms on match quality is limited. Measuring the impact of market level interventions is challenging because treatment is administered at the market level. A rigorous evaluation requires entire comparison markets that do not undergo the intervention. The best studies to date use cross-sectional comparisons across markets or time series comparisons within a single market. As mentioned above, [61] shows that clearinghouses using stable mechanisms outlast their unstable counterparts using a cross-sectional comparison of regional labor markets in the United Kingdom. In a short note, [55] show that clearinghouses do not affect average wages by comparing average wages across a cross-section of subspecialties of internal medicine that match with and without a centralized clearinghouse. The validity of cross-sectional comparisons relies crucially on the assumption that the markets receiving the intervention are similar in all other aspects to the comparison markets. [56] show that the entry-level gastroenterology labor market had greater geographic mobility in years when a centralized clearinghouse was used than in the years before the clearinghouse was adopted or after it was eliminated. In the only study I am aware of to directly examine the impact of a market

level intervention on match quality, [28] show that television ratings of college football "bowl" games increased after reforms to end unraveling in the market for bowl game contracts were implemented. This type of pre/post research design cannot disentangle the impact of the intervention from other changes that occur concurrently with the reforms.

The present paper is the first I am aware of to use a credible quasi-experimental source of identifying variation to estimate the gains from adopting a variant of the DAA. My application is Teach for America (TFA), which places high-achieving college graduates and professionals in teaching positions in disadvantaged urban and rural areas, in my case Chicago and Northwest Indiana¹ and four comparison regions. A substantial portion of TFA Chicago's teachers are hired at a series of "interview days", which facilitate a large number of interviews between teachers and schools in a short period of time. Traditionally, teachers were matched to schools at interview days using what I refer to as the "Chicago Mechanism". I use a simple model to demonstrate that this mechanism promotes strategic early hiring in the unique symmetric equilibrium because schools are able to hire teachers at any time during the day and teachers are required to accept their first offer. Similar to [34], I find that the extent of this unraveling is driven by correlation in schools' preferences over teachers. These interview days are similar to many other unraveled labor markets in that they are an entry level labor market where schools have incentive to hire early and workers have incentive to accept these early offers.

I worked with TFA to adopt the School Position Proposing Deferred Acceptance Algorithm (DAA) at its high school interview days starting in 2014. Under the new mechanism, schools and teachers submit rank ordered lists of their preferences at the end of the interview day and matches are determined using the DAA. Because all matches are determined simultaneously, the market cannot unravel. Elementary school interview days were unaffected by this change so TFA's elementary school teachers serve as my comparison group. Simulations of my model suggest that the Chicago Mechanism may actually outperform the

1. I refer to this region as "Chicago" for ease of exposition.

DAA in terms of expected match quality in some cases. This ambiguity is best resolved with empirical evidence about the impact of matching using the DAA and related mechanisms.

My main outcome is teachers' retention rates through the start and end of their first school year. Retention is both a measure of match quality and an important outcome in itself because teacher attrition is disruptive to the teacher, the school, and the students and reflects poorly on TFA. A study of the costs of teacher turnover in Texas estimated replacing one teacher cost 25% as much as a teacher's salary and benefits [1]. TFA teachers have been shown to be more effective at teaching mathematics and equally effective at teaching reading than non-TFA teachers teaching at the same school in a randomized controlled trial [32]. Moreover, teachers hired mid-year to fill an unexpected vacancy are likely to be of below average quality. I show that trends in high school and elementary school teachers' retention rates are quite similar, supporting the common-trends assumption that is key to my identification.

As a thought experiment in the potential importance of teacher to school match quality, Table 1.1 shows characteristics of two real traditional high schools in Chicago. School A is smaller, lower performing, and in a more violent neighborhood than School B. Nearly one in five students at School B have limited English, whereas virtually no students at School A do. Teachers may prefer working at either of these schools for a number of valid reasons. For example, a teacher who joined TFA to make a difference for the most disadvantaged students may prefer School A, whereas a teacher who is fluent in Spanish may prefer School B where she can get more use of this skill.

I find replacing TFA's Chicago Mechanism with the DAA increases retention through the start and end of teachers' first school year by 7 and 9 percentage points, respectively. This is a 31% and 24% reduction in attrition through these points in the first school year, respectively. Using schools' submitted preference reports, I show that schools' preferences over teachers are quite heterogeneous. This suggests these estimates are plausibly a lower bound on the gains from adopting the DAA in other markets. In addition to improving retention rates, adopting

Table 1.1: Characteristics of Two Potential Schools

	School A	School B
Type	Traditional Public HS	Traditional Public HS
Students	250	1000
Low Income	99%	96%
Limited English	0%	20%
5-Year Graduation Rate	50%	75%
Neighborhood	72 per 100k	33 per 100k
Homicide Rate	(\approx Honduras)	(\approx South Africa)

the DAA increased interview day hiring by 50%. While my research design cannot isolate whether the gains are driven by improving match quality directly or increasing the share of interview day hires, a comparison of actual interview day matches and simulated matches under the Chicago Mechanism suggests the increased retention is driven by higher teacher satisfaction with their interview day matches.

This increase in retention rates is comparable in magnitude to the change from raising new teachers' salaries by \$8,000 [39]. This is suggestive of an opportunity for large social welfare gains from better matching teachers to schools. More generally, the magnitude of this benefit suggests costly interventions to correct unraveling in other settings may be justified.

The remainder of the paper is organized as follows. Section 1.2 provides an overview of TFA and its interview days and explores the theoretical properties of Chicago Mechanism and the DAA. Section 1.3 discusses the quasi-experimental design as well as my data, estimation and inference. Section 1.4 examines the impact of adopting the DAA on retention through the start and end of the first year. Section 1.5 shows the impact of the intervention on unraveling and hiring. Section 1.6 concludes.

1.2 The Setting

1.2.1 *Teach for America*

TFA is a non-profit organization that recruits thousands of high achieving college graduates and professionals to teach disadvantaged youth in urban and rural areas for at least two years. Over 2000 TFA teachers have been assigned to the Chicago region. The 2014 "corps" consisted of 5,300 teachers nationally [82] and over 400 matriculants² were assigned to the Chicago region.³ Being accepted to TFA's corps is competitive. In 2014, the acceptance rate was 15 percent [82].

As part of their application, applicants rank regions and select which subject areas they would be willing to teach. If accepted, applicants are given an offer specifying a grade level and subject area in a particular region. For example, someone may be given an offer to teach high school science in Chicago. If this offer is unacceptable to the applicant, she can reject it and not join Teach for America. [27] discusses how this initial match is implemented, focusing only on the regional component of the offer. Grade level and subject area assignments are determined using a similar algorithm.

After accepting this offer, teachers must match with a teaching position in their assigned grade level and subject area. In Chicago, most teachers attend an "interview day". Interview days are designed to facilitate a large number of interviews between teachers and schools in a short period of time. Typically, each teacher is scheduled to have 25 minute interviews with 6 or 7 different schools.⁴ Teachers only interview for positions in their assigned subject

2. Note that TFA does not include matriculants who exit the organization very early or defer their acceptance to a later year in its official numbers. According to the official numbers, the Chicago region had 291 TFA teachers in 2014. Because not all matriculants are included in TFA's official numbers, I refer to them as matriculants here for clarity. Henceforth, I will refer to the set of matriculants as teachers for ease of exposition.

3. For comparison, the largest traditional certification program according to Institutional Postsecondary Education System (IPEDS) data, Ashford University, had 1,745 education graduates in 2013 (the last year data is available).

4. TFA sends resumes to participating schools before the interview day so administrators can request interviews with their most preferred candidates.

area, but schools often interview teachers for several subject areas over the 7 interview slots. Importantly for my research design, separate interview days are held for high school and elementary school teachers. Typically, three interview days are held for each grade level per year. Teachers who are not matched at interview days are matched at in-person or phone interviews setup by TFA staff members. A small number of teachers may exit the organization before being matched.

This interview day structure shares several features with other unraveled markets. Most importantly, the market can unravel because schools may optimally make strategic and potentially inefficient hiring decisions before completing all interviews. Moreover, TFA's interview days are an entry level labor market for teachers. Entry level labor markets may be more likely to unravel because new workers often have relatively fixed start dates and because their future productivity is more uncertain than that of more experienced workers. Finally, the requirement that teachers accept their first offer is similar to the norm that federal appellate clerks cannot refuse an offer from a federal judge [10].

1.2.2 *The Chicago Mechanism*

Before I began working with TFA, all interview day matches were determined by the *Chicago Mechanism*. This mechanism is described below.

Step 1. Complete round 1 interviews. After completing the round 1 interview, each school decides whether to make an offer to the teacher they interviewed. If given an offer, the teacher must accept it or exit Teach for America.

Step $2 \leq r \leq R$. All round r interviews involving unhired teachers are completed. When possible, unhired teachers waiting on standby fill interview vacancies created by hired teachers. Each school decides whether to make an offer to the teacher they interviewed. If given an offer, the teacher must accept it or exit Teach for America.

In order to analyze schools' and teachers' strategic behavior under the Chicago Mechanism, I consider a finite horizon non-stationary two-sided matching market where agents exit the market once matched. This setup is similar to the model of [24] with a few key differences which are highlighted as they arise.

There are two disjoint sets of teachers, $T = \{t_1, \dots, t_N\}$, and schools, $S = \{s_1, \dots, s_N\}$. An outcome of the game is a matching between teachers and schools, $\mu : T \cup S \rightarrow T \cup S$, where $\mu(t_i) = s_j$ if and only if $\mu(s_j) = t_i$. $\mu(t_i) \in S$ unless $\mu(t_i) = t_i$ which denotes teacher t_i is matched with her outside outcome. Similarly, $\mu(s_j) \in T$ unless $\mu(s_j) = s_j$ which denotes school s_j matching with its outside option.

Each potential match is associated with a match quality. Match quality between teacher t_i and school s_j is given by β_{ij} where $\beta_i = (\beta_{i1}, \dots, \beta_{iN})'$ is a vector of exchangeable random variables that are independent and identically distributed across teachers according to distribution $F(\cdot)$. Denote the j^{th} marginal distribution of $F(\cdot)$ by $F_j(\cdot)$ which are all assumed to have finite first moments. This specification allows schools' preferences over teachers to be correlated. In contrast, [24] assume there is a continuum of teacher and school types, and preferences are vertical over these types. Given the setup of [24], deferred acceptance yields the optimal assortative match. This is not the case given my specification, which can be thought of as more of a "worst case" scenario for deferred acceptance. Both teachers and schools are risk neutral, derive utility from their match quality, and do not discount. All schools have the same outside option with match quality normalized to 0. On the other hand, and in contrast to [24], teachers have stochastic outside options, \underline{u}_i , where $\underline{u}_i \sim Uniform(\underline{U}, \bar{U})$ for all i .

Schools sequentially interview teachers over R rounds, where $R \ll N$.⁵ Each teacher meets a single school in each round and schedules are randomly determined. During an interview between teacher t_i and school s_j , both teacher t_i and school s_j observe β_{ij} . After

5. This assumption ensures that schools do not meet all unhired teachers before completing all interview rounds.

each interview, each school chooses whether or not to hire the teacher being interviewed. When a teacher is hired, both the teacher and the hiring school exit the market. After all interviews are completed, hired teachers decide whether to accept the position or exit TFA. When a teacher exits, both the teacher and the school match with their outside options.

For simplicity, consider the case where $R = 2$. Teacher t_i accepts an offer if the expected quality of the position is weakly greater than the quality of her outside option. Having observed β_{ij} , school s_j believes teacher t_i will be retained with probability $\frac{\beta_{ij}-\underline{U}}{\bar{U}-\underline{U}}$. Therefore school s_j 's expected utility conditional on hiring teacher t_i is:

$$E[u_{ij}|\beta_{ij}] = \begin{cases} \beta_{ij} & \text{if } \beta_{ij} \geq \bar{U} \\ \frac{\beta_{ij}-\underline{U}}{\bar{U}-\underline{U}}\beta_{ij} & \text{if } \underline{U} < \beta_{ij} < \bar{U} \\ 0 & \text{if } \beta_{ij} \leq \underline{U} \end{cases}.$$

Let v^r denote the expected value of not hiring in round r .⁶ School j should hire in round r whenever $E[u_{ij}|\beta_{ij}] \geq v^r$. If school j does not hire in the final round, she will be left with her outside option with quality 0, so $v^2 = 0$. Therefore, school s_j should hire teacher t_i if $E[u_{ij}|\beta_{ij}] \geq 0$.

In the first round, school s_j 's expected utility from not hiring is equal to the expected utility of a second round hire. Because teachers hired in the first round exit the market, the second round pool of teachers is generally of lower quality. The expected match quality of a second round hire is:

$$v^1 = P(\beta_{ij} > 0 | \beta_{ij'} < v^1) E[u_{ij} | \beta_{ij} > 0, \beta_{ij'} < v^1].$$

The following proposition proves that there is a unique symmetric threshold v^{1*} . This establishes the existence and uniqueness of the symmetric equilibrium.

Proposition 1. *A symmetric sequential equilibrium exists and is unique. In this equilibrium,*

6. Exchangeability implies that $v_j^r = v^r \forall j$.

schools hire any teacher in the first or second round with match quality exceeding v^{1} or 0, respectively. Teachers accept any offer they prefer to their outside option.*

Proof. I have already shown teachers' best response and schools' second round threshold. I finish the proof of the proposition by verifying that the first round threshold, v^1 , is determined by a unique fixed point.

First, let $A = \{v : 0 \leq v \leq E[\beta_{ij} | \beta_{ij} > 0]\}$. A is a nonempty, convex set. Note that $v^r \in A$ for all r because at worst a school will match with its outside option and the average teacher in the second round market is weakly worse than the average teacher in the full market.

Consider $g(\cdot)$ defined as:

$$g(v^1) = P(\beta_{ij} > 0 | \beta_{ij'} < v^1) E[u_{ij} | \beta_{ij} > 0, \beta_{ij'} < v^1].$$

Notice that both $P(\beta_{ij} > 0 | \beta_{ij'} < v^1)$ and $E[u_{ij} | \beta_{ij} > 0, \beta_{ij'} < v^1]$ are continuous and monotonically increasing in v^1 , so the whole expression is continuous and monotonically increasing.⁷ Moreover, $0 \leq P(\beta_{ij} > 0 | \beta_{ij'} < v^1) \leq 1$ and $0 < E[u_{ij} | \beta_{ij} > 0, \beta_{ij'} < v^1] \leq E[\beta_{ij} | \beta_{ij} > 0]$ so $f(v) \in A$ for all v . This implies g is a monotonically increasing continuous function from A into itself. Brouwer's Fixed Point Theorem implies f has a fixed point and monotonicity implies uniqueness.

Denote this fixed point by v^{1*} . Then schools hire any teacher whose quality exceeds v^{1*} in the first round. □

The following example provides more insight in to the properties of this equilibrium.

Example 1. *Suppose $\beta_{ij} \sim \mathcal{N}(0, 1)$ with $Cov(\beta_{ij}, \beta_{ij'}) = \rho$ for $j \neq j'$ and $Cov(\beta_{ij}, \beta_{i'j'}) = 0$ for all $i \neq i'$ and \underline{u}_i are independent and identically distributed $Uniform(0, 0.25)$.*

Notice that the vector β_i is multivariate normally distributed. If school s_j believes other schools hire any teacher whose quality exceeds v^1 in the first round, the distribution of match

7. Monotonicity follows because A is a subset of \mathbb{R}_+ .

quality in the second round is given by:

$$\beta_{ij}|\beta_{ij'} < v^1 \sim \mathcal{N}\left(-\rho\frac{\phi(v^1)}{\Phi(v^1)}, 1 - \rho^2\left[1 - v^1\frac{\phi(v^1)}{\Phi(v^1)} - \left(\frac{\phi(v^1)}{\Phi(v^1)}\right)^2\right]\right).$$

Then, the distribution of match quality conditional on exceeding school j 's outside option is given by:

$$\beta_{ij}|\beta_{ij} > 0, \beta_{ij'} < v^1 \sim \mathcal{N}\left(\mu + \sigma\lambda\left(\frac{-\mu}{\sigma}\right), \sigma^2\left[1 - \lambda\left(\frac{-\mu}{\sigma}\right)\left(\lambda\left(\frac{-\mu}{\sigma}\right) + \frac{\mu}{\sigma}\right)\right]\right),$$

where $\lambda(x) = \frac{\phi(x)}{1-\Phi(x)}$ is the inverse mills ratio. Denote this distribution by $G(\cdot)$ and its density by $g(\cdot)$.

Then:

$$P(\beta_{ij} > 0|\beta_{ij'} < v^1) = 1 - \Phi\left(\frac{\rho\phi(v^1)}{\sigma\Phi(v^1)}\right).$$

and:

$$E[u_{ij}|\beta_{ij} > 0, \beta_{ij'} < v^1] = \frac{1}{1 - G(0)} \int_0^{0.25} 4b^2g(b)db + \frac{1}{1 - G(0)} \int_{0.25}^{\infty} bg(b)db.$$

The expected quality of a round two match is given by:

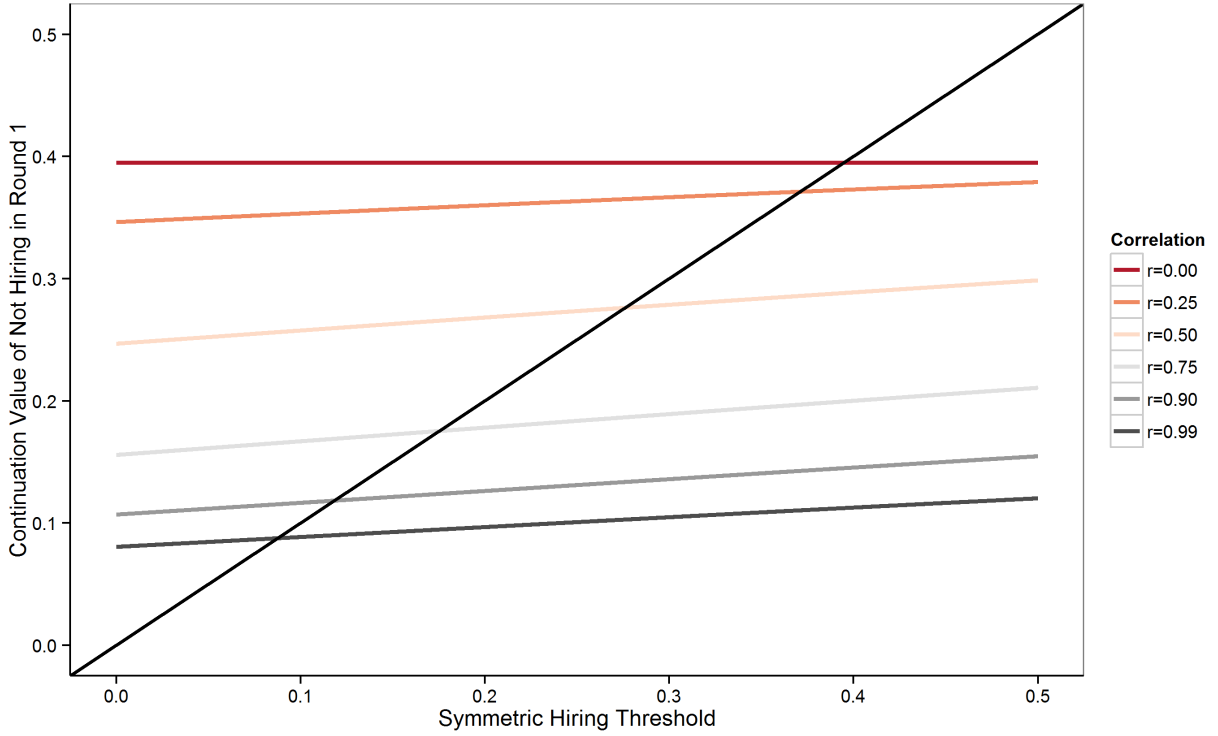
$$P(\beta_{ij} > 0|\beta_{ij'} < v^1)E[u_{ij}|\beta_{ij} > 0, \beta_{ij'} < v^1].$$

The equilibrium first round threshold is the value of v^{1*} satisfying:

$$v^{1*} = P(\beta_{ij} > 0|\beta_{ij'} < v^{1*})E[u_{ij}|\beta_{ij} > 0, \beta_{ij'} < v^{1*}].$$

Figure 1.1 plots the expected quality of a round two match against the first round threshold for $\rho \in \{0.00, 0.25, 0.50, 0.75, 0.90, 0.99\}$. The equilibrium threshold is determined by the

Figure 1.1: First Round Thresholds



Notes. Equilibrium round 1 strategies in Example 1 are determined by the intersection of the blue shaded lines with the 45 degree line. The equilibrium hiring threshold v^{1*} is decreasing in the correlation in schools' preferences over teachers.

intersection of the expected quality with the 45 degree line. The equilibrium threshold falls as preferences become more correlated. This is consistent with [34]'s finding that correlation between ordinal rankings drives unraveling.

When preferences are uncorrelated, v^{1} is 0.39. This implies 51.5% of hires are made in the first round interview. When preferences are highly correlated, $\rho = 0.99$, v^{1*} falls to 0.087 which causes the first round share of hires to increase to 83%.*

Expected match quality given the equilibrium strategies, σ^ , is given by:*

$$E[\beta_{ij}|\sigma^*] = P(\beta_{ij} > v^{1*})E[u_{ij}|\beta_{ij} > v^{1*}] + (1 - P(\beta_{ij} > v^{1*}))P(\beta_{ij} \geq 0|\beta_{ij'} < v^{1*})E[u_{ij}|\beta_{ij} > 0, \beta_{ij'} < v^{1*}].$$

When preferences are uncorrelated, the ex-ante expected match quality is 0.63 and the expected teacher exit rate is 4.8%. When preferences are highly correlated, the expected match quality falls to 0.12 and the expected teacher exit rate is 5.4%. These market outcomes are shown in Table 1.2 for comparison.

Table 1.2: Market Outcomes in Example 1 under the Chicago and DAA Mechanisms, R=2

Mechanism	Hire Rate	Attrition	Expected Match Quality
Optimal Assignment, $\rho = 0.00$	99.6%	0.0%	1.77
Optimal Assignment, $\rho = 0.50$	90.6%	1.0%	1.29
Optimal Assignment, $\rho = 0.99$	56.5%	7.6%	0.49
Chicago, $\rho = 0.00$	67.3%	4.8%	0.63
Chicago, $\rho = 0.50$	62.0%	4.8%	0.53
Chicago, $\rho = 0.99$	55.9%	7.3%	0.43
DAA, $\rho = 0.00$	63.2%	6.1%	0.61
DAA, $\rho = 0.50$	58.2%	6.9%	0.55
DAA, $\rho = 0.99$	49.7%	9.1%	0.41

Notes: See examples 1 and 2 for details. The hire rate is the share of schools that choose to make an offer to a teacher. Attrition is defined as a teacher quitting after being hired. Expected match quality is calculated for all schools accounting for the probability of hiring the outside option due to not making an offer or teacher attrition. Outcomes for optimal assignments and DAA values are averaged over 10,000 simulations assuming $N = 25$.

1.2.3 The Deferred Acceptance Algorithm

The general structure of interview days remained the same after the adoption of the Deferred Acceptance algorithm. As before, each teacher is scheduled to have 25 minute interviews with 6 or 7 different schools. However, rather than having schools hire teachers throughout the day, schools complete all scheduled interviews. After all interviews are completed, schools submit a rank ordered list (ROL) of their preferences over the teachers they are willing to hire for each of their open positions. An example of a blank school preference form is shown in Figure 1.2. Each teacher submits a ROL of her preferences over *all* of the schools she interviewed with. Teachers are prohibited from refusing an offer without exiting TFA in

order to keep with the spirit of the Chicago Mechanism that "TFA teachers go wherever they are needed" and schools are only permitted to rank teachers for positions in their assigned certification areas.

Teachers submit a rank ordered list over schools. Because schools rank teachers for each position, teacher preferences are expanded to be over positions at each school. Suppose a teacher ranked school s_1 first, which ranked teachers for two positions, and school s_2 second, which ranked for three positions. Her preferences would be modified to rank position 1 at s_1 first, position 2 at s_1 second, position 1 at s_2 third, position 2 at s_2 fourth, and position 3 at s_2 fifth.

Once schools and teachers submit their preferences, matches are determined using the *School-Position Proposing Deferred Acceptance Algorithm* (DAA). This algorithm operates as follows:

Step 1. Each school position makes an offer to its most preferred teacher. If a teacher receives multiple offers, she tentatively accepts her most preferred offer. If a teacher receives a single offer, she tentatively accepts it.

Step 2. All unmatched school positions make an offer to their most preferred teacher who has not already rejected them. Teachers accept their most preferred offer, including new offers and tentatively accepted offers. If a previously unmatched teacher receives a single offer, she tentatively accepts it.

Step k. Repeat Step 2 until all school positions are filled or all unfilled positions are unwilling to hire their next most preferred teacher.

As before, teachers can accept their match or exit TFA. Consider how this mechanism changes the outcome in Example 1.

Figure 1.2: School Schedule and Ranking Form

SESSION 1 SCHEDULE FOR: <<School Name>> _____

Note: You will interview all corps members below. **At the end, you will turn this form in at check-out.**

Round	Corps Member Name	Certification Area
Round 1: 8:00 – 8:25am	<<CM Name>>	<<CM Cert Area>>
Round 2: 8:30 – 8:55am		
Round 3: 9:00 – 9:25am		
Round 4: 9:30 – 9:55am		
Round 5: 10:05 – 10:30am		
Round 6: 10:35 – 11:00am		
Round 7: 11:05 – 11:30am		

Ranking to Hire: Once interviews have concluded, rank all corps members you are willing to hire for each position(s) below. These rankings are utilized by the program to make an 'offer' to your highest ranking first; if another school also ranks the corps member, the corps member's preference will break the tie and the program will then offer to your next candidate and so on until you match with (hire) a candidate or your list concludes. As a result, it is most advantageous for you to rank more corps members than less in each area understanding that any candidate below is a potential hire.

	CERT AREA: _____	CERT AREA: _____	CERT AREA: _____	CERT AREA: _____
	# of hires I'd like from the list below: _____	# of hires I'd like from the list below: _____	# of hires I'd like from the list below: _____	# of hires I'd like from the list below: _____
1				
2				
3				
4				
5				
6				

Scheduling Follow-up Interviews: Fill out the below if you would like follow-ups to be scheduled on Friday (4/11).

<input type="checkbox"/> I'm choosing not to submit rankings today and would like to follow-up with the candidates indicated below. <input type="checkbox"/> Regardless of how hiring goes today, I would like to schedule a follow-up with the candidates indicated below.
--

Example 2. *As in Example 1, suppose $\beta_{ij} \sim \mathcal{N}(0, 1)$ with $\text{Cov}(\beta_{ij}, \beta_{ij'}) = \rho$ for $j \neq j'$ and $\text{Cov}(\beta_{ij}, \beta_{i'j'}) = 0$ for all $i \neq i'$ and \underline{u}_i are independent and identically distributed $\text{Uniform}(0, 0.25)$. In order to keep the example as consistent with Example 1 as possible, assume there are still 2 interview rounds.*

The assumption that schools and teachers both derive their utility from expected match quality implies that schools and teachers have aligned preferences. This, in turn, implies a unique stable match [57]. This match can alternatively be recovered by the following (infeasible) algorithm that requires knowledge of match qualities:

Step 1. *Match the school-teacher pair with the maximum match quality.*

Step 2. *Match the school-teacher pair with the maximum match quality among unmatched schools and teachers.*

Step k. *Repeat Step 2 until all school positions are filled or all unfilled positions are unwilling to hire their next most preferred teacher.*

This alternative algorithm allows matches to be more easily characterized. For example, when $N = 2$, the expected match quality is given by:

$$E[u|DAA] = \frac{1}{2}(E[u|\beta \text{ is max of 4 draws}] + E[u|\beta \text{ not max}]),$$

which equals 0.61. For $N = 3$, the expected match quality calculation is similarly given by:

$$E[u|DAA] = \frac{1}{3}(E[u|\beta \text{ is max of 6 draws}] + E[u|\beta \text{ is max of remaining 3}] \\ + E[u|\beta \text{ not max of 6 or 3}]),$$

which equals 0.62. Notice when $N = 3$ the maximum match is always the maximum realization of 6 match qualities. The second match is always the maximum of the remaining 3 match qualities. The final match is the only remaining match.

When $N = 4$, the calculation becomes somewhat more complicated. The first match is always the maximum realization of 8 match qualities and the second match is always the maximum of the remaining 5 match qualities. However, the third match can be the maximum of 2, 3, or 4 remaining match qualities. The first two scenarios each occur with probability $\frac{4}{9}$ and the final scenario occurs with probability $\frac{1}{9}$. Again, the final match is the only remaining match. The expected match quality across these three scenarios is 0.62.

Table 1.2 compares expected market outcomes using the optimal certainty equivalent match, the Chicago Mechanism, and the DAA. The Optimal Assignment and DAA calculations assume $N = 25$ and are averaged over 10,000 simulations of this example. Perhaps surprisingly, the DAA only outperforms the Chicago Mechanism in terms of average match quality when preferences are moderately correlated. This is in part because of the “worst case” specification of match quality as DAA necessarily yields the optimal match in [24]’s specification.

1.3 Measuring the Impact of Adopting the Deferred Acceptance Algorithm

1.3.1 Quasi-Experimental Design

The previous section showed that there is no guarantee that the DAA will yield better matches than the Chicago Mechanism. Despite this ambiguity, there is little empirical evidence about the impact of determining matches with the DAA on match quality and efficiency.

Measuring the impact of market level interventions is challenging because treatment is administered at the market level. A rigorous evaluation requires comparison markets that do not undergo the intervention. Because I worked with TFA to implement the DAA only at high school interview days and not at elementary interview days, I am able to provide the most rigorous quasi-experimental evidence to date of the impact of an intervention to

correct unraveling on match quality and market outcomes. Specifically, I identify the impact of adopting the DAA by comparing the change in outcomes of high school teachers to the change in outcomes of elementary school teachers in Chicago to the same difference in four other TFA regions because I have panel data on both the market for high school teachers and for elementary school teachers in each of these five regions. This difference-in-difference design identifies the causal effect of using the DAA instead of the Chicago Mechanism so long as the difference between high school and elementary outcomes in the other four regions is an appropriate counterfactual for the difference in Chicago if the new mechanism had not been adopted. This is often referred to as a *parallel trends* assumption. I am able to show suggestive evidence in support of the parallel trends assumption by showing that the trends were approximately parallel in the six years before the intervention.

1.3.2 Data

The next challenge to evaluating the impact of adopting the DAA at TFA's interview days is getting high quality data covering both the pre- and post-intervention period for all of TFA's teachers in these five regions. I accomplish this using restricted access administrative data provided by TFA's National Research team ("national" data) and the TFA Chicago region ("Chicago" data).

The national data includes details about all 5,956 teachers who accepted a position in or transferred to one of these regions between 2008⁸ and 2014. Importantly, I observe teachers' grade level and subject area placements which determine whether they participate in high school or elementary school interview days. I exclude 1,268 special education teachers, 21 percent of my sample, from my main analysis because they are not given an initial high school or elementary assignment. I also observe teachers' point of departure if they exited TFA before completing their two-year commitment. Finally, I observe each teacher's gender, race and ethnicity, college, college graduation year, college major, and college GPA. I control

8. One region's data begins in 2010.

for college quality using the 75th percentile SAT math score at the college according to 2013 Integrated Postsecondary Education Data System (IPEDS) records.

Summary statistics for my analysis sample are shown in Table 1.3. Table 1.3 demonstrates that the size of TFA's corps grew in recent years. The size of the corps increased 50% between 2008 and 2013, but shrank somewhat between 2013 and 2014.

Few TFA teachers switch regions after receiving an initial offer from TFA. Just under half of teachers are assigned to teach high school, but nearly 10 percent of high school teachers switch to elementary school.

About 70% of teachers are female. 44% of teachers are non-White, including 11% who are Black, 10% who are Hispanic, and 22% who are another race.

On average, teachers had a 3.54 grade point average in college and attended schools with a 75th percentile SAT math score of 662. Only 5% of teachers were education majors in college, whereas 46% were social science majors and another 46% were either math, science, or humanities majors. 72% of the sample joined TFA right out of college and only 4% had been out of college for 6 or more years.

Finally, Chicago is the largest region in the data, consisting of 36% of the sample.

The Chicago data includes information about Chicago's 2013 and 2014 interview days, including teacher and school interview schedules and interview day hires. For 2014 high school interview day sessions, this data also includes the rank ordered lists submitted by teachers and schools.

1.3.3 Estimation and Inference

My data is a repeated cross-section of high school and elementary teachers in five TFA regions. I estimate the impact of adopting DAA at interview days using a triple difference

Table 1.3: National Sample Summary Statistics

	mean	sd
Corps Year: 2008	0.12	0.33
Corps Year: 2009	0.12	0.33
Corps Year: 2010	0.12	0.33
Corps Year: 2011	0.13	0.33
Corps Year: 2012	0.18	0.38
Corps Year: 2013	0.18	0.38
Corps Year: 2014	0.15	0.36
Initial Assignment: Chicago	0.36	0.48
Actual Assignment: Chicago	0.36	0.48
Grade Level Assignment: High School	0.45	0.50
Actual Grade Level: High School	0.34	0.47
Content Area Assignment: SPED	0.00	0.00
Content Area Assignment: Science	0.22	0.41
Content Area Assignment: Math	0.17	0.37
Content Area Assignment: Social Science	0.07	0.25
Content Area Assignment: Language	0.24	0.43
Content Area Assignment: General	0.30	0.46
Female	0.68	0.47
Black	0.11	0.32
Hispanic	0.10	0.30
Other Race	0.22	0.41
College GPA	3.54	0.38
College: 75th Percentile SAT Math Score	661.61	175.91
College Major: Education	0.05	0.21
College Major: Social Science	0.46	0.50
College Major: Math or Science	0.21	0.41
College Major: Humanities	0.25	0.43
College Major: Other	0.03	0.18
Double Major	0.34	0.47
Years Between College Degree and TFA: 0	0.72	0.45
Years Between College Degree and TFA: 1-2	0.18	0.38
Years Between College Degree and TFA: 3-5	0.07	0.25
Years Between College Degree and TFA: 6+	0.04	0.19
Chicago	0.36	0.48
Region 2	0.18	0.38
Region 3	0.29	0.45
Region 4	0.12	0.32
Region 5	0.06	0.23
Observations	4687	

model estimated with the following individual level regression:

$$y_{ijrt} = \beta \text{DAA}_{jrt} + \eta \text{High School} \times 2014_{jrt} + \alpha_j + \gamma_{rt} + X_i' \delta + u_{ijrt}, \quad (1.1)$$

$$i = 1, \dots, N, \quad j = \text{Elementary, High School}, \quad r = 1, \dots, 5, \quad t = 2008, \dots, 2014,$$

where i indexes individual teachers, j indexes whether teacher i was assigned to teach at a high school or elementary level, r is teacher i 's region, and t denotes teacher i 's TFA cohort. y_{ijrt} denotes teacher i 's outcome. DAA_{jrt} is an indicator for being a high school teacher in Chicago in 2014. This is the group which could have participated in a high school interview day using the DAA. $\text{High School} \times 2014_{jrt}$ is an indicator for being a high school teacher in 2014. α_j is a grade level assignment fixed effect which captures time-invariant level differences in outcomes between high school and elementary school teachers. γ_{rt} is a region-cohort fixed effect. X_i is a vector of individual level control variables, including indicators for gender, race, college major, being a double major, and for time out of college as well as college GPA, the 75th percentile math SAT score of the college, and an indicator for missing this SAT score measure.

I also present estimates from a difference-in-difference model using only teachers from Chicago which identifies the impact using the change in high school teachers' outcome relative to the change in elementary school teachers' outcome. I estimate this impact with the following individual level regression:

$$y_{ij1t} = \beta^{DD} \text{DAA}_{j1t} + \alpha_j^{DD} + \gamma_{1t}^{DD} + X_i' \delta^{DD} + u_{ij1t}^{DD}, \quad (1.2)$$

where all variables and indices are defined as above. The DD superscript indicates the estimates are from the difference-in-difference specification instead of the main triple differences specification.

Because DAA_{jrt} only varies at the grade level-cohort level, the error terms u_{ijrt} are effectively averaged over individuals within a region-grade level-cohort, rather than over

all individuals. As a result, standard errors should be clustered at the region-grade level assignment-cohort level. Unfortunately, there is a well-known downward bias in cluster-robust standard errors [13, 18] when there are a small number of clusters. Here, there are 66 clusters across all regions and 14 clusters in Chicago.

When measuring the impact of adopting the DAA on match quality I adjust for this bias using the Wild cluster bootstrap with the null hypothesis imposed [18]. Under the assumption that errors are symmetric around zero within each cluster, the Wild Bootstrap estimates the finite sample distribution of the cluster T-statistic using realizations across pseudo-samples generated by multiplying all of the errors from each cluster by 1 or -1 . This manipulation works well for clustered designs because it maintains the correlation structure within clusters. With 14 clusters, there are 16,384 possible pseudo-samples, including the true sample. I draw 2,000 pseudo-samples to calculate the finite sample distribution of the T-statistic. P-values are calculated by ranking the true estimates in the distribution of the absolute value of the estimates across pseudo-samples.

When estimating the impact of adopting the DAA on interview day participation and hiring, I assume that there were no cluster level random effects and use heteroskedasticity robust standard errors. This is by necessity, because I only have interview data from 2013 and 2014. With only 4 clusters, there are only 16 possible pseudo-samples using the Wild Bootstrap. This leaves a null rejection region at the 5 percent level.

1.4 The Effect of Adopting the DAA on Teacher Retention

TFA teachers commit to teaching at their placement school for two years, but in practice some TFA teachers exit the organization before this two year commitment is over. In this section, I examine the impact of the DAA on retention through the start and end of the first year of teaching.⁹

9. TFA's official retention rates are conditional on starting the first school year. Consequently, the retention rates reported here are lower than TFA's officially reported rates.

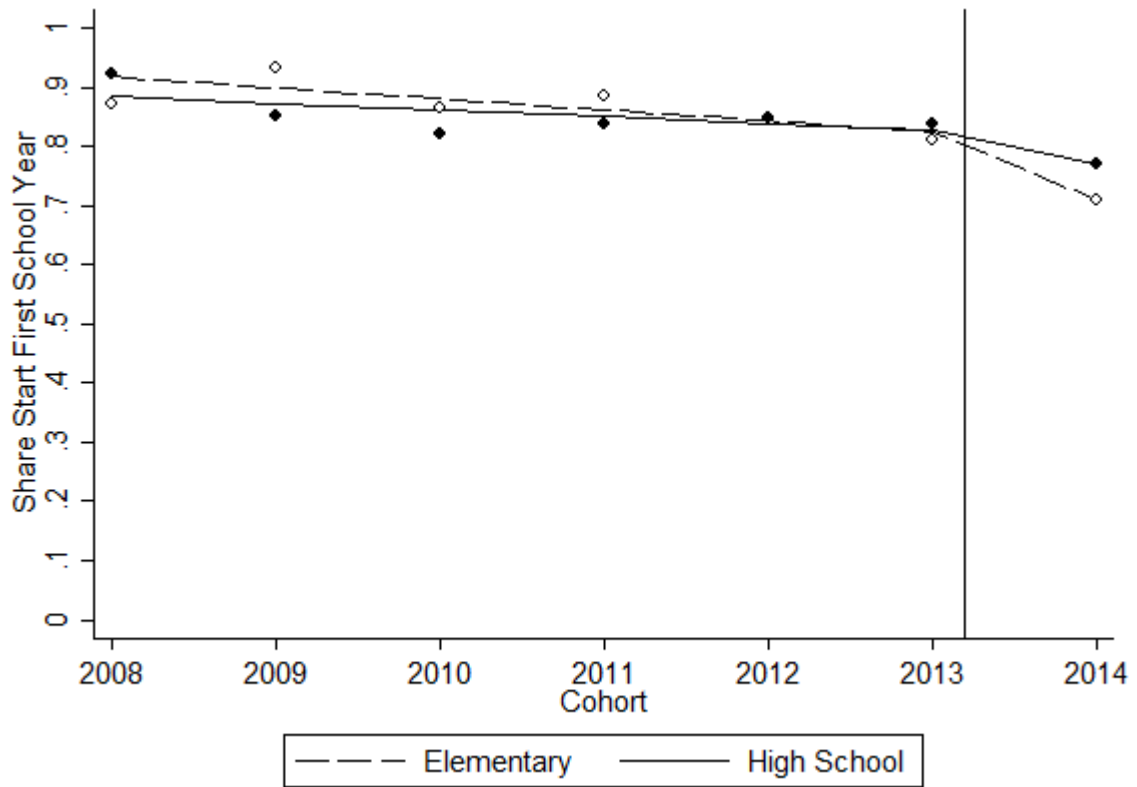
1.4.1 Retention through the Start of the First School Year

TFA teachers have between 3 and 6 months between their interview day and the start of the school year. This includes an intense six week summer institute, which is essentially a teaching boot camp. When a teacher exits TFA before the start of the school year, schools must scramble to quickly fill this teaching position. This separation is disruptive for the teachers too, who likely had planned to be affiliated with TFA for two years when they accepted the offer and are left looking for a new job. Some teachers on the margin of leaving or staying may decide to leave because they are unhappy with their school placement. The effect of the DAA on teacher retention through the start of teachers' first school year is an important measure of match quality as well as an important outcome for TFA, schools, and students.

Figure 1.3 plots the retention rates through the start of teachers' first school year by TFA cohort separately for high school and elementary school teachers in Chicago. Special education teachers are excluded from this analysis because they are not assigned a grade level when accepted to TFA. The DAA was adopted for high school teachers in the 2014 cohort. The black dots are the actual retention rates of high school teachers in each TFA cohort and the solid black line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA. Similarly, the white dots are the actual retention rates of elementary school teachers in each TFA cohort and the dashed line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA.

Figure 1.3 demonstrates most TFA teachers in Chicago, high school or elementary, start their first year of school. In 2013, the year prior to the intervention, 82% of all teachers started their first year. Increasing this retention rate is therefore an ambitious goal. Importantly for the difference-in-difference design, changes in elementary school retention rates closely track changes in high school retention rates. Statistically, the probability of seeing the observed differences between high school and elementary school start of school year retention rates from an F-test under the null hypothesis of equal cohort effects across the two groups is

Figure 1.3: Trends in Retention Through Start of First School Year



Notes. The black dots are the actual retention rates of high school teachers in each TFA cohort and the solid black line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA. Similarly, the white dots are the actual retention rates of elementary school teachers in each TFA cohort and the dashed line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA.

0.43.¹⁰

Table 1.4 shows estimates of the impact of adopting the DAA instead of the Chicago Mechanism, $\hat{\beta}$, on retention through the start and end of the first year of teaching. Column 1, which is the estimate of $\hat{\beta}$ from the triple-difference specification in Equation 1.1 without individual covariates, shows adopting the DAA increased retention through the start of the first year by 5.8 percentage points. Comparing this impact to the predicted retention rate in the absence of the intervention suggests that adopting the DAA reduced the estimated pre-school year quit rate of high school teachers by 28%, from 21% to 15%. These retention rates are estimated by subtracting the estimated coefficient from the observed retention rate for the treated cohort. Using HAC standard errors, the T-statistic for this estimate is 2.01. This is slightly above 1.986, the critical value for a 5 percent level test with a T-distribution with 64 degrees of freedom. However, this critical value may be downward biased given that the data includes only 66 clusters [13, 18]. To correct for this bias, I also present the p-value generated by the Wild Bootstrap [18]. The adjusted p-value is 0.047 so the estimate is still statistically significant at the 5% level.

10. The probability of seeing these differing patterns under the null of equal linear trends is 0.92. The F-statistic was calculated using an individual level regression with the standard set of controls.

Table 1.4: The Impact of DAA on Retention

Outcomes	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Start First Year			Finished First Year				
DAA	0.058	0.072	0.073	0.068	0.066	0.086	0.07	0.063
T-Statistic	2.078	2.56	5.294	3.738	2.29	2.979	4.095	3.779
P-Value Wild Bootstrap	0.047	0.023	0.007	0.019	0.145	0.025	0.031	0.004
Estimated Mean without DAA	0.791	0.776	0.766	0.772	0.669	0.648	0.714	0.721
Regions	All		Chicago Only		All		Chicago Only	
Cohort by Region Effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Observations	4687	4687	1698	1698	4687	4687	1698	1698

Column 2 shows the estimate of $\hat{\beta}$ from Equation 1.1 with individual covariates included. Including individual covariates increases the point estimate to 0.072 which is a 32% reduction in attrition relative to the estimated retention rate. This effect remains statistically significant at the 5% level using either a T-distribution with 64 degrees of freedom or the wild bootstrap.

Columns 3 and 4 present estimates of $\hat{\beta}^{DD}$ from the difference-in-difference specification in Equation 1.2. In Column 3, we see a very similar 7.3 percentage point increase in retention due to the adoption of DAA while the T-statistic nearly doubles. The 0.05-level critical value from a T-test with 12 degrees of freedom is 2.18, but here the downward bias in clustered standard errors with few clusters becomes more of a concern because there are only 14 clusters in the Chicago only sample. However, the effect remains significant at the 1% level after adjusting for this bias using the wild bootstrap.

1.4.2 Retention through the End of the First School Year

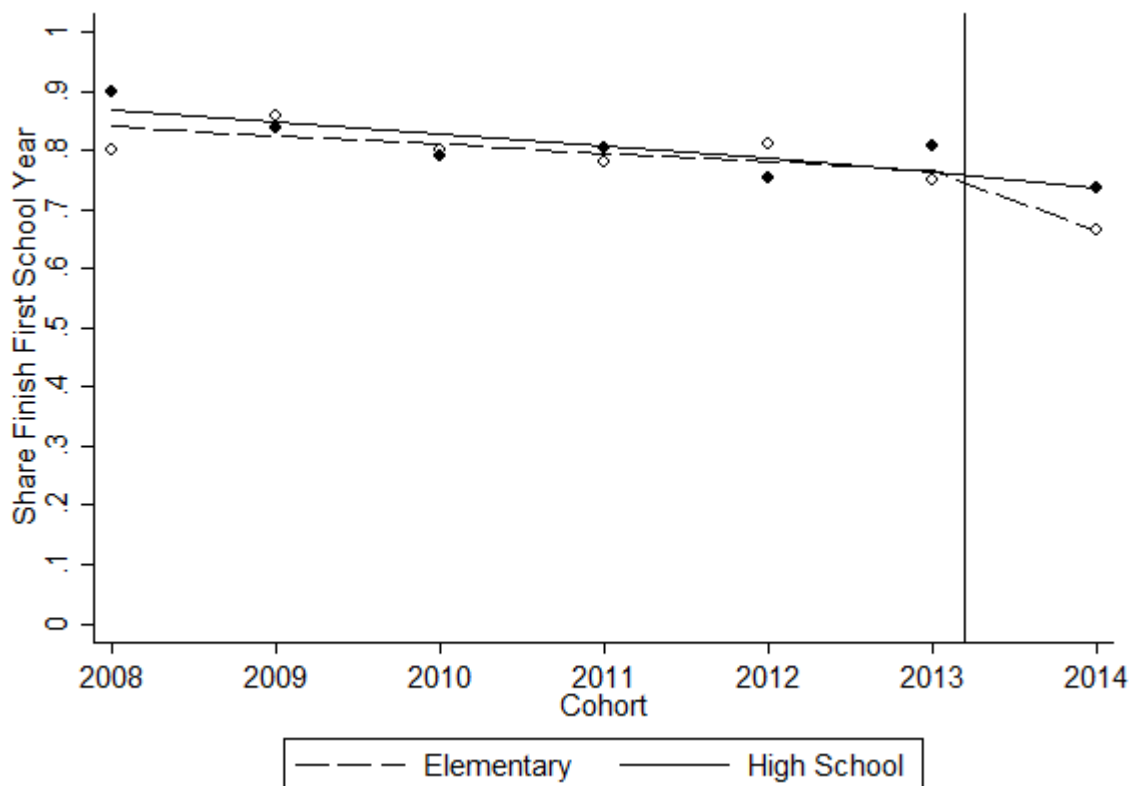
Figure 1.4 shows retention rates through the end of teachers' first school year by assigned grade level in Chicago. The black dots show the actual retention rates of high school teachers in each TFA cohort and the solid black line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA. The white dots show the actual retention rates of elementary school teachers in each TFA cohort and the dashed line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA.

Retention is quite high through the end of the first year. Across all pre-intervention years, 81% of TFA teachers complete their first year of teaching. As with retention through the start of the first year, the elementary school end of first year retention rate closely tracks the high school retention rate. The probability of seeing these differing annual retention rates from an F-test assuming equal cohort effects across groups is 0.42.¹¹

Table 1.4, columns 5 and 6 are analogous to Columns 1 and 2 above, but using retention

11. The probability from an F-test assuming equal linear cohort trends is 0.28.

Figure 1.4: Trends in Retention Through End of First School Year



Notes. The black dots are the actual retention rates of high school teachers in each TFA cohort and the solid black line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA. Similarly, the white dots are the actual retention rates of elementary school teachers in each TFA cohort and the dashed line is the estimated linear trend in high school teachers' retention prior to the adoption of DAA.

through the end of the first year as the outcome. The increase in retention grows over the school year. Without controls (Column 5), the estimates suggest that adopting the DAA increased retention through the end of the first year by 6.6 percentage points. Across all specifications, this is the only effect that is not statistically significant at conventional levels after adjusting for the downward bias in clustered standard errors. With individual controls, this impact grows to a statistically significant increase in retention of 8.6 percentage points. Comparing this effect to the predicted first school year quit rate without the DAA suggests a 24% reduction in quits through the first school year, from 33% to 24%.

Columns 7 and 8 are analogous to columns 3 and 4 but again using retention through the end of the first year. Using the difference-in-difference strategy, the increase in retention appears to be driven entirely by retention through the start of the school year. With individual controls, the effect falls slightly from 0.068 to 0.063.

1.4.3 How valuable are these impacts?

Given that TFA already had established the centralized structure of interview days, adopting the DAA was essentially a zero cost or even a cost saving intervention. This intervention caused a 6 to 9 percentage point increase in retention through the end of the first year of teaching, equal to a 24 percent reduction in attrition. How valuable is this higher retention?

Often, policies attempt to increase teacher retention by increasing teacher salaries. [39] estimates an \$8,000 increase (measured in 2015 dollars) in new teachers salaries in Milwaukee would lead to a 6 percentage point reduction in teachers exit rate from the profession, which is equal to the lower bound of the estimates. If applied to all 400 teachers in Chicago's 2014 corps, this alternative policy would cost nearly \$3.2 million.

There may be additional benefits from increased retention of TFA teachers in particular. [32] find that students randomly assigned to a TFA teacher instead of a non-TFA teacher teaching the same grade and subject at the same school performed 0.15 standard deviations better on a math exam and equally well on a reading exam, on average. These estimates

may understate the true efficacy of the TFA teachers who remain in teaching because of the DAA compared to the next best option because they are conditional on already working at a school. When a school leader is forced to hire quickly to fill an unexpected vacancy, she may need to hire a lower quality teacher. Estimates in [21] suggest test score gains of this order of magnitude translate in to substantial increases in income later in life.

Of course, teachers who quit TFA may not be comparable to the average TFA teacher. Table 1.5 shows the association between select teacher characteristics and quitting by the start or end of the first school year prior to the intervention. The results suggest teachers with better outside options, not necessarily the worst teachers, are more likely to exit. In particular, teachers with who have higher GPAs or who have been out of college longer are more likely to exit at either point in the school year. Men and non-white teachers are also somewhat more likely to quit. Match quality may be especially important for non-white teachers if they strongly prefer teaching students from the same demographic group.

This is suggestive of an opportunity for large social welfare gains from better matching teachers to schools. More generally, the magnitude of this benefit suggests costly interventions to correct unraveling in other settings may be justified.

1.5 The Effect of Adopting the DAA on Unraveling and Interview Day Hiring

One of the important potential benefits of the DAA is that it does not promote unraveling, or strategic early hiring, whereas the Chicago Mechanism does. This suggests that, on average, hires under the DAA should occur in later rounds than under the Chicago Mechanism. Table 1.6, Column 1 shows that the DAA increased the average round of hires, conditional on being hired, by 0.57 rounds. This effect is not significant at conventional levels. The average round of hire at DAA interview days was 3.5, so between the third and fourth interview rounds. This is exactly the average round of hire that would be seen if hires were

Table 1.5: Correlates of Pre-DAA Attrition

Quit By:	Start First Year	End of First Year
College GPA	0.048*** (0.016)	0.056*** (0.019)
Years Between College and TFA: 1-2	0.035** (0.015)	0.050*** (0.017)
Years Between College and TFA: 3-5	0.057** (0.023)	0.084*** (0.026)
Years Between College and TFA: 6+	0.097*** (0.033)	0.184*** (0.037)
Female	-0.020* (0.012)	-0.021 (0.013)
Black	0.105*** (0.019)	0.117*** (0.022)
Hispanic	0.048** (0.019)	0.062*** (0.022)
Other Race	0.057*** (0.014)	0.076*** (0.016)
N	3979	3979

Notes. Sample restricted to pre-DAA period (2008-2013).
Regression controls for all individual controls.

uniformly distributed across six rounds, the modal number of interviews. This is slightly lower than would be expected if hiring was randomly distributed across seven rounds, but given that schools are permitted to request interviews with candidates, they may be more likely to interview their most preferred candidates in earlier rounds. Perhaps more surprising is that this effect suggests that the average hire round would have occurred almost exactly in the third round if the Chicago Mechanism had been used instead of the DAA. This raises the question: Why did the market unravel so little with the Chicago Mechanism?

Table 1.6: Impact of DAA on Hires by Round

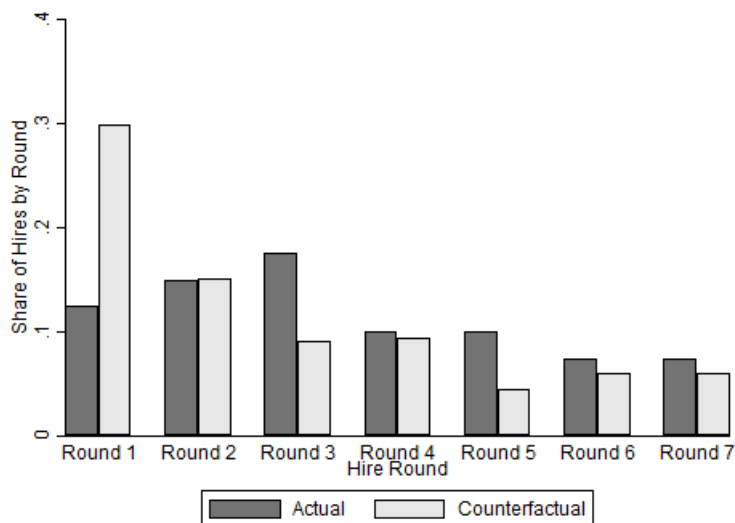
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Average	Round	Round	Round	Round	Round	Round	Round	Round
	1	2	3	4	5	6	7	8
DAA	0.57	-0.17	-0.00	0.08	0.01	0.05	0.01	0.01
	[0.68]	[0.14]	[0.14]	[0.11]	[0.13]	[0.12]	[0.10]	[0.09]
Observations	166	166	166	166	166	166	166	166

Standard errors in brackets

* $p < .1$, ** $p < .05$, *** $p < .01$

Table 1.6 shows the impact of the DAA on the distribution of hires across rounds. The outcome of each column is an indicator for being hired in the specified round. Because these estimates are conditional on being hired at an interview day, the estimates sum to 1 by construction. Consequently, coefficients are interpretable as changes in the distribution of hires across rounds. Figure 1.5 plots the realized distribution of hiring across rounds (in dark gray) and the counterfactual distribution implied by the coefficients in Table 1.6. Round 1's share of all hires falls from 30% to 13%, while the share of hires in rounds 3 through 7 increases from 35% to 53%. This suggests that at least one in six hires under the Chicago Mechanism would have been a strategic, potentially inefficient early hire.

Figure 1.5: Actual and Counterfactual Distribution of Hires By Round



Notes. The dark shaded bars are the actual share of hires in each round. The light shaded bars show the counterfactual distribution of hires implied by the estimates in Table 1.6.

The model of schools' behavior under the Chicago Mechanism demonstrated that similarity of preferences among schools is an important driver of unraveling in equilibrium. Intuitively, a school must make hiring decisions with more urgency when it knows that other schools are actively hiring from its pool of preferred teachers. Therefore, the market may not have unraveled further because schools' preferences are sufficiently dissimilar. I am not

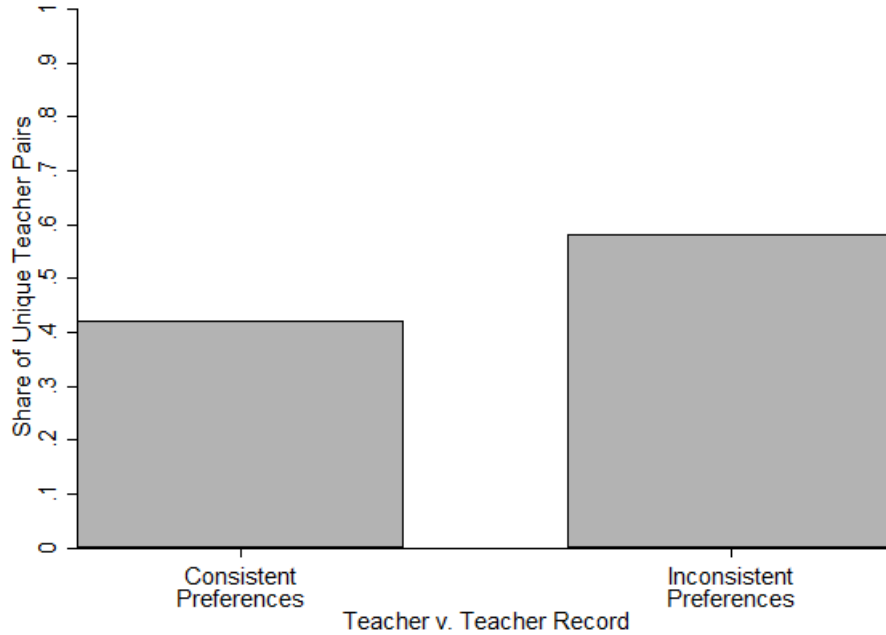
directly able to measure the similarity of schools' preferences because schools interview different subsets of teachers. However, I can learn about preference similarity whenever two teachers interview with the same school. Suppose all schools have identical preferences. Then whenever the same two teachers are both ranked by a school, the relative ordering of these teachers should be the same. On the other hand, if schools have completely idiosyncratic preferences, the relative ranking should switch in about half of cases. Changes in the relative ordering of two teachers across schools is therefore evidence of the similarity or dissimilarity in school preferences.

Across 587 interviews between 66 schools or school networks and 90 teachers¹² over three interview days, there were a total of 1,522 instances where two teachers both interviewed with the same school at a single interview day. To be informative about preference similarity, teachers need to interview with and be ranked by at least two of the same schools. In total, 116 pairs of teachers meet these criteria. Figure 1.6 shows the distribution of results over these 116 teacher-to-teacher competitions in my preference data. The left bar shows the share of cases where the relative ranking of teachers remains constant across interviews with 2 or 3 different schools and the right bar corresponds to cases where the relative rankings of two teachers changes across schools. Across all cases where teachers are interviewed by multiple schools, the relative ranking of the teachers switches 58% of the time. Just limiting attention to cases where two teachers were ranked by two schools, the relative ordering switches 54% of the time. This is quite close to what we would expect if schools have completely idiosyncratic preferences. This suggests that the dissimilarity of schools' preferences over teachers likely limited the amount of unraveling in the market.

The similarity of schools' preferences over teachers is of interest for several reasons beyond understanding unraveling in the present application. Econometrically, [17] assume schools preferences over teachers are determined using a single criterion function. [5] and [4] use a vertical preferences assumption, i.e. that every agent on one side of the market has identical

12. This includes 39 special education teachers.

Figure 1.6: Results of Pairwise Teacher Interview Competitions



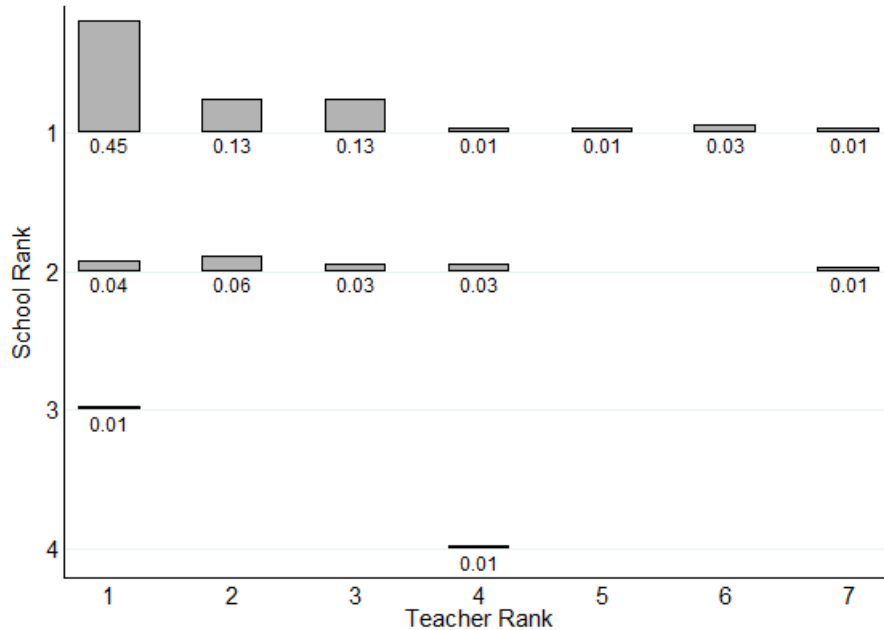
Notes. This figure shows the results of competitions between pairs of teachers who interviewed with at least two of the same schools where at least one teacher was ranked by each school. In total, 116 pairs of teachers meet these criteria. The left bar corresponds to cases where the relative ranking of teachers remains constant across interviews and the right bar corresponds to cases where the relative rankings of two teachers changes across schools.

preferences, to nonparametrically identify preferences in a two-sided matching market. The previous results suggest that the applicability of these types of assumptions to teacher labor markets may be limited. In the education literature, [43] argue that principals are unable to predict the quality of new teachers. These results are consistent with their claim in that there is substantial disagreement among school leaders about who will become the most effective teachers.

A benefit of dissimilar preferences is that more schools are able to match with their top choice teacher. Figure 1.7 shows the distribution of teacher and schools' preferences over matches. In 44% of matches, both the teacher and the school matched with their top choice. Just focusing on schools' preferences over matches, 78% of schools matched with their top choice and 18% matched with their second choice. 50% of teachers matched with their most

preferred school and 35% matched with either their second or third most preferred school. Only 3% of teachers matched with their least preferred school.

Figure 1.7: Teacher and School Ranks Over Matches



Notes. This figure shows the joint distribution of schools' and teachers' ordinal preferences over interview day matches when matches are determined by the DAA. The marginal distribution of schools' ordinal preferences can be recovered by adding left to right across columns. The marginal distribution of teachers' preferences can be recovered by adding top to bottom across rows. Teachers are required to rank all schools. Schools only rank teachers they are willing to hire.

Importantly, the DAA could affect match quality by changing the interview day hiring rate in addition to reducing unraveling. Table 1.7 summarizes participation and hiring at the 2013 and 2014 high school and elementary interview days, the two years for which I have interview day data. In 2013, 78% of elementary school teachers and 63% of high school teachers attended an interview day. Across all teachers, including those who did not attend an interview day, 29% and 39% of elementary and high school teachers were hired at interview days, respectively. In 2014, elementary teacher interview day attendance fell by a tenth to 71% while high school teacher attendance, at 62%, essentially did not change. Despite the higher attendance rate of elementary school teachers, nearly twice as many high

school teachers, 48%, were hired at an interview day than elementary school teachers, 25%.

Table 1.7: Interview Day Attendance and Hiring

Day	N	Attendees	Attendance Rate	Hires	Hire Rate
A. High School					
2014	84	52	62%	40	48%
2013	93	59	63%	36	39%
Change			-2%		9%
B. Elementary					
2014	221	157	71%	55	25%
2013	223	174	78%	64	29%
Change			-7%		-4%
Difference-in-Difference			5%		13%

The 13 percentage point differential between the change in the high school and elementary interview hire rates is suggestive of an increase in hiring due to the DAA. Table 1.8 shows estimates of the impact of the DAA on interview day outcomes using the specification shown in equation 1.1. Column 1 shows that the DAA actually increased participation by 6 percentage points despite the fact that the high school participation rate only increased by a single percentage point. This is because the difference-in-difference specification in equation 1.1 identifies the impact of the DAA as the difference between the change in outcomes of high school teachers and elementary school teachers. Because the elementary participation rate fell by 7 percentage points in 2014, the small increase in the high school participation rate between 2013 and 2014 hides the fact that the actual counterfactual change may have been a moderate decline in participation. Without a comparison group, the estimated impact would be close to zero. With that said, this 6 percentage point increase is imprecisely estimated and is not significantly different than zero at conventional levels.

Column 2 shows that the DAA increased interview day hiring by 16 percentage points. This effect is economically large and statistically significant at the 10 percent level. It suggests high school interview day hiring was 50% higher than it would have been without

Table 1.8: Impact of DAA on Interview Day Outcomes

	(1)	(2)
	Attended Interview Day	Hired at Interview Day
DAA	0.06 [0.09]	0.16* [0.09]
Mean without DAA	0.56	0.35
Observations	621	621

Notes. Robust standard errors in brackets.

Stars indicate * $p < .1$, ** $p < .05$, *** $p < .01$.

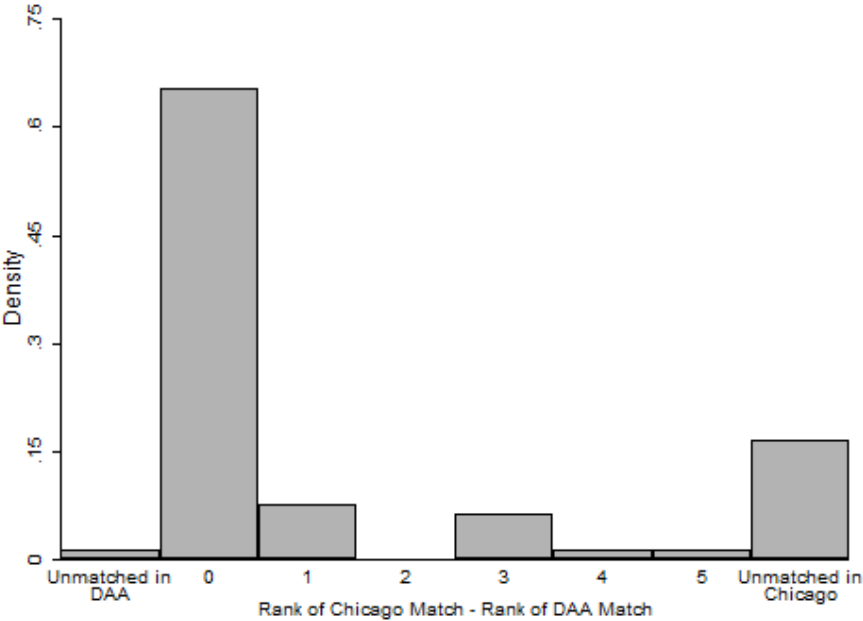
the DAA. This is unconditional on actually attending an interview day. Because just under two-thirds of high school teachers attended an interview day, the impact on interview day hiring conditional on attending is 50% larger, suggesting a 22 percentage point increase in interview day hiring among attendees. This effect is statistically significant at the 5 percent level.

While my research design is unable to identify whether the welfare gains are due to the decline in unraveling or the increase in interview day hiring, I can examine the impact of the DAA on teachers' and schools' preferences over their interview day matches by simulating matches under the Chicago Mechanism. Suppose two schools would like to hire a teacher. When matches are determined by the DAA the teacher is assigned to the school she prefers. When matches are determined by the Chicago Mechanism, the teacher is assigned to the school that interviewed her first because this school could hire her before she met the second school. If I make the simplifying assumption that the set of teachers schools are willing to hire is unchanged by the additional information they have when the DAA is used instead of the Chicago Mechanism, I can simulate the Chicago Mechanism by breaking ties with teachers' interview schedules instead of their preferences.

Figure 1.8 shows how teachers' preferences over their assigned matches differ from their simulated Chicago Mechanism matches. Positive values indicate that the DAA gives the preferable match and negative values indicate that the Chicago Mechanism gives the preferable match. No teachers prefer their match under the Chicago Mechanism to their match

under the DAA. However, 66% are indifferent and 1 percent are unmatched by the DAA but not by the Chicago Mechanism. The remaining third of teachers prefer their match under the DAA to their match under the Chicago Mechanism. 10 percent of teachers prefer their match under the DAA by 3 to 5 ranks on their rank ordered list. Because teachers ranked six to seven schools, this indicates that 10 percent of teachers went from one of their bottom choices to one of their top choices. If teachers who match with their top choices are less likely to quit or be fired this change can plausibly explain all of the change in retention. Moreover, 15 percent of teachers would not have been matched at all under the Chicago Mechanism. This is quite similar to the actual 16 percentage point increase in hiring.

Figure 1.8: Change in Teacher Preferences Over Matches

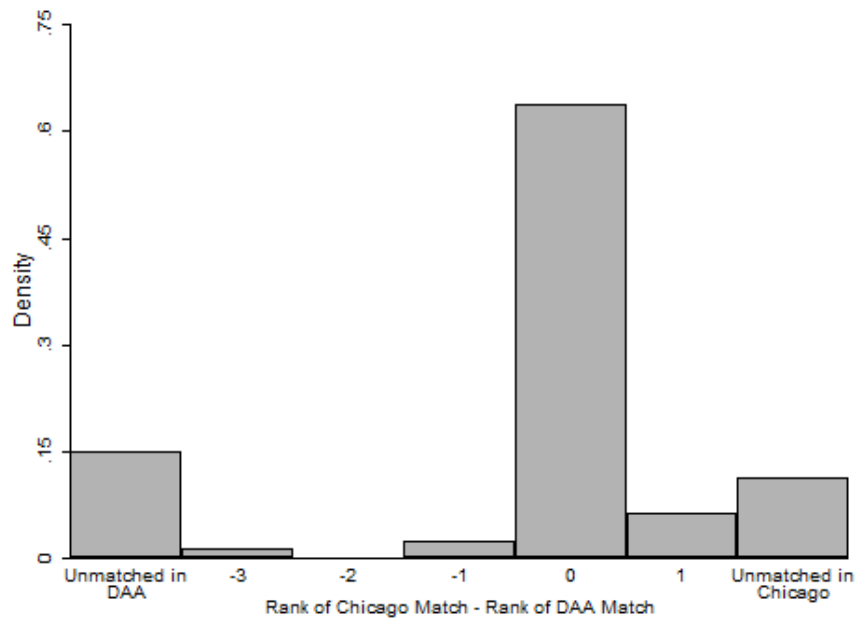


Notes. This figure shows the distribution of differences between teachers’ ordinal preferences over their actual match from the DAA and their counterfactual match under the Chicago Mechanism. A positive difference of x implies a teacher prefers her match under DAA to the match under the Chicago Mechanism by x ranks. Counterfactual matches were determined by running the school proposing DAA replacing teachers’ actual preferences with their interview schedule.

Figure 1.9 shows how schools’ preferences over their assigned matches differ from their simulated Chicago Mechanism matches. Again, positive values indicate that the DAA gives the preferable match and negative values indicate that the Chicago Mechanism gives the

preferable match. As with the teachers, nearly two-thirds of matched schools end up with the same match under the DAA as they would have under the Chicago Mechanism. However, in contrast to teachers, who almost universally did weakly better under the DAA, some schools prefer their match under the Chicago Mechanism. About 4% of schools prefer their match under the Chicago Mechanism and 14% are not matched under the DAA but are matched by the Chicago Mechanism. Still, 6% of schools prefer their match from the DAA to the match they would have gotten with the Chicago Mechanism and another 10% were matched by the DAA but not by the Chicago Mechanism. In total, 18% of schools would have preferred their match had the Chicago Mechanism been used, but 16% prefer their assigned match to their counterfactual match. Importantly, some of the teachers whose matches changed may have actually quit before the end of their first school year had their match not changed.

Figure 1.9: Change in Teacher Preferences Over Matches



Notes. This figure shows the distribution of differences between schools' ordinal preferences over their actual match from the DAA and their counterfactual match under the Chicago Mechanism. A positive difference of x implies a school prefers its match under DAA to the match under the Chicago Mechanism by x ranks. Counterfactual matches were determined by running the school proposing DAA teachers' actual preferences with their interview schedule.

1.6 Conclusion

This study presents the most rigorous quasi-experimental evidence of the impact of matching using the Deferred Acceptance Algorithm on match quality to date. Replacing Teach for America’s Chicago Mechanism, which promoted strategic early hiring by allowing schools to hire teachers at any time during interview days, with the School Position Proposing Deferred Acceptance Algorithm increased retention through the start and end of teachers’ first school year with TFA by 7 and 9 percentage points, respectively. These increases in retention are equal to a 31% and 24% reduction in quits through these points in the first school year, respectively. This increased retention benefits Teach For America, teachers, schools, and, perhaps most importantly, students. Previous estimates of the impact of teacher salaries on attrition suggest these impacts are similar in magnitude to the additional retention associated with an \$8,000 increase in new teacher salaries [39], a policy which would cost nearly \$3.2 million if applied to all TFA Chicago teachers in the 2014 cohort. These estimates are arguably a lower bound for the benefits in other markets because substantial heterogeneity in schools’ preferences over teachers likely limited the inefficiencies under the Chicago Mechanism. I estimate that about 15% of hires would have been strategic hires due to unraveling had the DAA not been adopted. Adopting the DAA also increased interview day hiring by 50 percent. Comparing actual matches to simulated matches under the Chicago Mechanism suggests that the welfare gains are driven by better, not necessarily more, interview day matches.

The present study is the first to apply principles of economic design to teacher labor markets and consequently has several implications about how to improve teacher quality. While Teach for America teachers are admittedly a somewhat unique subset of the teaching profession, matching mechanisms may be appropriate for teacher labor markets more generally because wages are often fixed by rigid union contracts, making them similar to other markets without prices. Moreover, while most teacher labor markets in the United States are currently decentralized, there are a number of countries where teachers are matched to

schools using a centralized assignment mechanism, including France [23], Mexico [59], Turkey [26], and Uruguay [77]. Because the quality improvements from a single year of teaching experience correspond to significant gains in student achievement [74], increasing the retention of new teachers could lead to substantial improvements in teacher quality. This is especially true in disadvantaged urban schools, where the least experienced teachers are most likely to work [49].

More generally, the large welfare gains from adopting the DAA at TFA's interview days suggest that the benefits from adopting stable centralized clearinghouses in other markets may also be large. These gains likely justify devoting substantial effort and resources to correcting unraveling in other markets, like the market for federal law clerks.

CHAPTER 2

RETHINKING THE BENEFITS OF EMPLOYMENT PROGRAMS: THE HETEROGENEOUS EFFECTS OF SUMMER JOBS (W/ SARA B. HELLER)

2.1 Introduction

American youth face major obstacles in the wake of the Great Recession. Youth employment over the summer, when teenagers are most likely to work, is near a 60-year low [75]. Minority and low-income youth face even more serious challenges: The 2010 employment rate for low-income black teens in Illinois - the location of this study - was less than one-fourth the rate for higher-income white teens (9 vs. 39 percent), and homicide kills more young black males nationwide than the other 9 leading causes of death combined [7, 20].

Decades of workforce development programs have targeted these problems, motivated by the idea that improved human capital and additional work experience should improve employment outcomes and increase the opportunity cost of crime. Yet despite over \$5 billion in annual spending on these programs in the U.S. [46], they have shown limited empirical success among young people. Only very intensive and expensive interventions appear to improve employment among disadvantaged youth [45, 53, 68], and even fewer reduce crime.¹ Although benefit-cost analyses of short-term results, which have to make assumptions about how long results will persist, sometimes suggest these investments are worthwhile [52, 60], conclusions based on longer-term measures tend to be more pessimistic [69]. The apparent difficulty of improving employment via jobs programs has often led to the conclusion that such programs require too much investment to be worthwhile (e.g., [36]).

1. Of the large, well-evaluated youth employment programs, only Job Corps and JobSTART reduce crime [19, 69] (whereas the Job Training Partnership Act may actually increase crime among male youth) [16]. The crime reductions, however, fade out quickly after the end of the programs, raising the possibility that only intensive programs reduce crime because their effects are limited to incapacitation during the program itself. The National Supported Work Demonstration also appears to have reduced crime among older participants, but not among youth [76].

It is therefore particularly surprising that two recent studies, one in Chicago and one in New York City, raise the possibility that a short, low-intensity employment intervention - supported summer jobs - actually can have major and lasting impacts on youth, even in the absence of changes in employment [31, 37]. Both studies find substantively large declines in violence,² a problem which generates a cost to victims of over \$600 billion per year (inflated to 2015\$ from [54]).

This paper uses medium-term results from the initial randomized controlled trial in Chicago, as well as a second randomized cohort of more criminally-involved youth the following year, to better understand why summer jobs reduce violence and for whom such programs work best. In the summers of 2012 and 2013, we randomly assigned a total of 6,855 youth to the One Summer Chicago Plus (OSC+)³ program or to a control group. Although some of the program elements varied across cohorts (see Section II), in both years treatment groups were offered a 6-8 week part-time summer job at minimum wage (\$8.25/hour) along with a job mentor - a constantly-available adult to assist youth in learning to be successful employees and to help them deal with barriers to employment. Most youth also participated in a curriculum built on cognitive behavioral therapy principles aimed at helping youth manage their cognitive and emotional responses to conflict, as well as encouraging them to set and achieve personal goals. We track youth in administrative data from the Chicago Public Schools (school records through the 2014-15 school year), the Chicago Police Department (arrest records through March 2015), and the Illinois Department of Employment Security (unemployment insurance records through the first quarter of 2015).

We find that the program caused very similar reductions in violent crime across the

2. [37] finds that the Chicago program we study here reduces violent-crime arrests by 43 percent over 16 months, a drop which accrues largely after the end of the program (i.e., is not due solely to incapacitation). [31] find that New York City's summer jobs program reduces mortality by 20 percent, likely driven by reductions in homicide.

3. The acronym in prior publications was OSP, but the City of Chicago has since updated the way it refers to the program. We update our abbreviation accordingly, noting that although the program continues to experiment with different program elements and study populations over time, it is basically the same intervention.

2012 and 2013 cohorts of about 5 arrests per 100 youth - a 29 percent reduction. This effect is statistically significant at the 5 percent level in the pooled sample and for the 2012 cohort alone and continues to be significant after making adjustments for multiple hypothesis testing. Excluding program months, we find a proportionally large (22%), but statistically insignificant reduction of 2.7 arrests in violent crime per 100 youth in the remainder of the first post-program year. The results stop accruing after the first year. We do not find significant effects on other crime types, but we see proportionally large increases in property and drug crime, which causes the program to have no impact on the total number of arrests. Our estimates suggest the program had basically no impact on youth's schooling outcomes, including attendance, GPA, and graduation. Similarly, we find no effect on youth's non-program provider formal employment outcomes. However, we see an increase in employment at the program providers after the program. This is likely driven in part by extended programming built in to the 2013 program, but the 2012 cohort was also more likely to be employed at program providers in the second year after the program.

After accounting for the social savings from the reduction in violent crime, we estimate that the program was approximately cost neutral overall. However, our results suggest there may be some important heterogeneity across outcomes and by type of youth. Our key pre-specified hypothesis was about school - serving youth before they face full time labor market on their own seems to be a key difference between summer jobs programs and other programs that have all been for disconnected youth. We find that the reduction in arrests for violent crime is driven by in school youth, but this effect is not statistically distinguishable from the small positive effect seen among out of school youth. Moreover, the reduction in arrests for violent crime comes at the cost of 5.9 additional arrests for drug crime per 100 in-school youth-an 85% increase compared to comparable members of the control group. We also find suggestive evidence that youth with a baseline arrest or without recent experience in the formal sector benefited more from the program.

As with any subgroup analysis, data mining becomes a concern. In particular, spurious

large effects may be mistaken for important heterogeneity. To guard against these concerns, we turn to recent methods from the supervised machine learning literature designed to address treatment heterogeneity in causal inference problems [9, 80]. These methods provide a more principled, data-driven method for detecting truly important measures of treatment heterogeneity.

Implementing [80]’s causal forest algorithm, we find evidence that criminally-involved in-school male youth with relatively poor attendance may be especially responsive to the program. Less criminally-involved females who are early in their high school careers also seem to benefit more than other groups. Our causal forest estimates suggest the most disconnected youth in our sample, older, highly criminally-involved male youth who are out of school and do not have any recent formal employment, may be most adversely affected by the program. This suggests similar “light touch” social programs may be best suited for less disconnected youth, whereas the most highly disconnected youth, who are often the target population of such programs, likely require more intense programming.

2.2 Program Description

The program was designed by the government agency that administers the program, Chicago’s Department of Family and Support Services (DFSS), primarily as a violence-reduction intervention. The details of OSC+ varied across the two summers (discussed separately below), although the basic structure remained the same: Youth were offered a 5 hour per day, 5 day per week summer job at minimum wage (\$8.25 per hour). All youth were assigned a job mentor - an adult to assist them in learning to be successful employees and to help them deal with barriers to employment - at a ratio of about 10 to 1. Characteristics of mentors varied: Some were staff at the program providers, some were college students home for the summer, and some were individuals who applied for the mentor jobs directly. Mentors participated in a one-day training (which has been revised and extended in later years of the program) and were paid a salary. DFSS administered the program through contracts with local non-profit

agencies. These agencies recruit applicants, offer participating youth a brief training, hire the mentors, recruit employers to host youth, place youth in summer jobs, provide daily lunch and bus passes when appropriate, monitor youths' progress over the course of the summer, and if youth are fired, work with them to find an alternative placement.⁴

One hypothesis for why prior youth employment programs require lengthy intervention to improve outcomes is that disadvantaged adolescents may lack the “soft skills” to benefit from lower-intensity programming. To test whether targeting some of these skills could improve the impact of the program, some youth also spent 2 of the 5 daily hours in a social-emotional learning (SEL) curriculum based on cognitive behavioral therapy principles.⁵ The curriculum varied somewhat by provider⁶, but the lessons focus on emotional and conflict management, social information processing, and goal setting; they aim to help youth learn to understand and manage the aspects of their emotions and behavior that might interfere with successful participation and employment (e.g., the inclination, not uncommon among adolescents, to snap defensively at a someone offering constructive criticism). Prior research has shown that similar programming can reduce violent crime and create lasting improvements in school engagement on its own [38].

2.2.1 Summer 2012

In the first year of the program, youth ages 14 - 21 were recruited from 13 Chicago public high schools selected for the high number of students at risk of violence. Any youth attending or planning to attend an eligible school could apply; about 13 percent of the prior year's student

4. In 2012, three agencies served as program providers: Sinai Community Institute, St. Sabina Employment Resource Center, and Phalanx Family Services. In 2013, the program expanded; providers were Black Star Project, Blue Sky Inn, Kleo Center Community Life Center, Phalanx Family Services, St. Sabina, Westside Health Authority, and Youth Outreach Services.

5. Over the course of the two program years, it became clear that 2 hours per day was too much time to devote to this curriculum; it has since been changed to once a week.

6. In both years, the SEL curriculum was provided by two agencies: Youth Guidance and SGA Youth and Family Services.

population did. Youth could work for a total of 8 weeks⁷ at range of employers in the non-profit and government sectors. The jobs involved tasks such as supervising younger youth at summer camps, clearing lots to plant a community garden, infrastructure improvement at local schools, and providing administrative support at aldermen's offices. Because of restrictions assigned by a funder, there were no private sector jobs in this program year. Half of the treatment group was randomly assigned to work fewer hours (3 per day) and attend the SEL curriculum for the same wage during the other 2 hours; the other half worked all 5 hours per day.

2.2.2 Summer 2013

In part because OSC+ is intended as an experimental program with which to test why the program works and for whom, and in part because of logistical constraints, the 2013 design differed in a few ways from the 2012 program. Because the school district lengthened the 2013-14 school year, the shorter summer necessitated a 6-week instead of 8-week program, during which all youth received the SEL programming. Funding restrictions were lifted, so private sector jobs were included. Eligibility was limited to youth ages 16-22 in order to reduce the burden of obtaining work permits among 14- and 15-year olds. DFSS also encouraged treatment youth to keep participating in programming offered by the community service agencies after the summer ended, which included a mix of additional SEL activities, job mentoring, and social outings such as sporting events and DJ classes. These activities were much lower intensity than the summer programming, and students received a small stipend (about \$200) rather than an hourly wage for participation.

Importantly, the eligibility requirements and recruiting process also changed in order to

7. OSC+ was originally designed to run over 7 summer weeks, but additional funding allowed for an optional week-long extension of the jobs component. Eight weeks of programming were offered but not required, and in the 8th week there was no SEL programming. One service provider also offered access to additional, optional programming outside of OSC+ (like drama, graphic design, and fitness activities), but these activities were not funded by the program. Program impacts were not limited to this provider, so these activities seem unlikely to be the key driver of the results.

test treatment heterogeneity. Because of their focus on violence reduction, DFSS decided to limit the program to male youth, who are disproportionately involved in violence. They also recruited applicants differently. About half of the applicants were referred directly from the criminal justice system (from probation officers, juvenile detention or prison, or a center to serve justice-involved youth). No one was required to apply, but adults in the justice agencies invited youth who they judged to be work-ready to fill out applications. The other half of the applicants had applied to Chicago’s broader summer programming; those who were ages 16-20, lived in one of the 30 highest-violence community areas, and included a social security number on their application⁸ were entered into the lottery.

2.3 Experimental Design

2.3.1 *Summer 2012*

Prior to the program, DFSS selected 13 Chicago public high schools to participate. Because OSC+ was designed to prevent violence, the schools chosen had the highest number of youth at risk of violence involvement in the city, as identified by a separate research partner. Program providers encouraged youth at these schools to apply to the program, marketing it as a summer jobs program with more work hours (and so more opportunity for income) than Chicago’s standard summer programming.

A total of 1,634 youth in the study schools chose to apply for the 700 available program slots. The research team blocked youth on school and gender (the former to match youth to the closest program provider and the latter to over-select males, who are disproportionately involved in violence). We then randomly selected 350 youth for the jobs-only treatment arm and 350 for the jobs + social-emotional learning treatment arm. The remaining applicants were randomly ordered within blocks and treatment groups to form a waitlist. When 30

8. The intention was to facilitate matching to employment records, but these hand-entered social security numbers turned out to be too error-prone for such matching. We explain our alternative source of SSNs below.

treatment youth declined to participate, the first 30 control youth (in the same block and treatment group as the decliners) were offered the program, for a total treatment group of 730. For further details on the 2012 experimental design, see the web appendix to [37].

2.3.2 *Summer 2013*

Program providers recruited youth from two applicant pools. The first pool ($n = 2,127$) were male⁹ youth ages 16-22 at the start of the program who were recruited from a number of criminal justice agencies.¹⁰ The second pool ($n = 3,094$)¹¹ were youth ages 16-20 who lived in one of the 30 Chicago neighborhoods with the highest rates of violent crime and provided a social security number on their application. Each of the applicant pools was entered into a lottery using a block-randomized design. Youth were blocked on applicant pool and on age (under or over 18), largely because the city had a legal obligation to keep probationers who were under age 18 separate from those who were over 18. In order to ensure that each of the service providers was assigned the number of youth they were able to serve, and to minimize the distance youth had to travel to the provider offices, we also blocked on the geographic location of applicants' home address.

Youth were randomly assigned to treatment or control groups within these applicant pool-age-geography blocks, and each block was assigned to a specific service agency. Our main

9. 21 female youth applied in this applicant pool.

10. These agencies include Juvenile and Adult Probation, the Cook County Sheriff, the Juvenile Temporary Detention Center, the Department of Juvenile Justice, and the Juvenile Intervention Support Center.

11. The web application through which these youth applied was not designed solely for OSC+, and it did not ask applicants to report their gender. Since the 2013 program was for boys only, we randomly assigned all applicants in this second pool ($n = 7,588$) and relied on program providers to discern gender (those assigned to treatment were contacted by the service providers and not offered the program if female, with the exception of a very small number of transgender youth who were female but identified as male and so were allowed to participate). After the lottery, we matched youth to other administrative data sources (school and arrest records) that include gender. The analyses reported here drop female applicants. This does not undermine the integrity of random assignment, since gender is a baseline characteristic. We do not observe gender for the observations that were not located in schooling records and had no pre-program arrests ($n=353$). We include these observations in the analysis, with an indicator variable for the fact that they are missing gender.

analysis consists of 5,221 youth (2,636 treatment and 2,585 control).¹² Note that because of the time-constrained recruiting process, the number of youth assigned to the treatment group far exceeds the number of available slots (1,000).¹³ One important implication is that the maximum take-up rate possible - even if the first thousand youth were immediately located and agreed to participate - is 38 percent (1,000 out of 2,636). Note that this is by design and should not be interpreted as indicating low demand for the program among the treatment group.

2.4 Data

We match our study youth to existing administrative datasets from a variety of government sources. Although this has the drawback of generating limited evidence on mechanisms, it makes the study feasible by keeping data collection costs minimal. Program application and participation records come from DFSS. We measure crime with Chicago Police Department (CPD) administrative arrest records. Youth were matched to these records probabilistically using their name and birth date. The arrest data include the date and a description of each offense, which we use to categorize offenses as violent, property, drug, or other (vandalism, trespassing, etc.). The data cover both juvenile and adult arrests from 2001 through March 2015. Youth who have never been arrested will not be in the CPD records, so we assign zero

12. Despite a de-duplication process at the time of application, 52 youth submitted duplicate applications that were not identified until after random assignment. We consider a youth to be in the treatment group if any of his applications were assigned to it, effectively using the maximum of a youth's random assignment indicators as their final assignment. This ensures that treatment assignment is still random, since the maximum of any series of random variables is also random. For the analysis, we only retain one observation per youth. Because those who entered the lottery more than once had a higher probability of being selected, treatment assignment is only random conditional on the number of applications submitted. To account for this, we include indicator variables for submitting one or two duplicate applications in all of our analyses.

13. In planning to serve a very mobile and arrest-prone population, it was clear that filling all the available slots would take considerable time. Rather than add to the recruiting time by giving providers the same number of names as available slots and asking them to wait for additional lists when not all youth could be located, we gave providers lists of hundreds more youth than available program slots upfront. As a result, providers were not expected to contact everyone on their lists of treatment youth. Instead, they stopped recruiting once their slots were filled. We count everyone on the "treatment" list as part of the treatment group, since we did not enforce the rule that providers work down the list in order.

arrests for individuals not matched to the CPD data.

We use student-level administrative records from Chicago Public Schools (CPS) to capture schooling outcomes, matching youth using their unique CPS identification numbers if provided on their application, or probabilistically using their name and birth date if their numbers were unavailable. These data include details about the youths' enrollment status, grade level, course grades, attendance, and disciplinary actions¹⁴ from the beginning of their enrollment in CPS through the 2014-15 academic year. Because the 2012 cohort was recruited through schools, they all have school records. In the 2013 cohort, 91.63% of youth were successfully linked to a CPS record, with no difference in the matching rate between treatment and control groups ($p=0.556$). We matched youth using historical CPS records beginning in 1988, before the oldest youth in the study was born, so that any youth who had ever enrolled in CPS would be matched to their schooling records.¹⁵ Since many of these youth were no longer enrolled in school during the pre-program year, many have missing CPS data in the pre- and/or post-program years (see descriptive statistics and results below for details).

To measure employment, we use quarterly Unemployment Insurance (UI) records. These data include the total earnings, employer name, and industry for each employer in the formal sector a youth worked for in each quarter. In order to match youth to UI data, the Illinois Department of Employment Security (IDES) requires youths' social security numbers (SSNs). We took advantage of the fact that the school district has historically asked for SSNs during the enrollment process (they no longer do so, but most of the study youth enrolled at a time

14. CPS underwent a major reform of how they recorded disciplinary incidents during this time, so it is not clear how comparable recording is across or even within schools. As such, we do not use the disciplinary data as outcome measures.

15. There are two reasons that study youth might not be found in the CPS data. First, they may not have ever attended school in Chicago. About 1% of the sample lived outside the school district limits (but still in Cook County) at the time of their application. These youth would not be expected in the CPS data unless they had previously lived in the city and enrolled in CPS during prior years. Since youth are legally allowed to drop out of school at age 17, some youth may also have been above the age of legally-required school enrollment when they moved into the district. Second, data errors may have generated false negatives in the matching process.

when they did).¹⁶

These data provide an incomplete measure of employment for a number of reasons. First, as with all UI data, the records only include employment eligible for UI withholding, which excludes many agricultural and domestic positions, family employment, and any employment in the informal sector. Field work by sociologists and ethnographers suggests that the informal economy may be a non-trivial source of income for youth living in low-income neighborhoods (e.g., [33] and [78]). Second, not all youth had SSNs available for matching, either because they were not in the CPS data at all (437 youth, or 6.37% of our sample), or because CPS did not have a SSN on record (1,340 of the 6,418 CPS records were missing SSNs). CPS also removed 238 of the matches because of “significant” conflicts between the names in the two files (i.e., they removed apparent false positive matches). In all these cases, youth might have had employment records in the UI data, but they would be missing in our data. Our main analysis treats anyone without an SSN or who was removed as a bad match as missing. For the remaining youth with “valid” SSNs,¹⁷ we assign zeros for employment and earnings, assuming anyone not found in the matching process never worked in the formal sector. This approach assumes that cases are missing completely at random; the appendix shows the results are robust to different approaches to missing data.

2.5 Analytical Methods

The analysis plan is as follows: Let Y_{ibt} denote some post-program outcome for individual i in block b during post-randomization period t . This outcome, Y_{ibt} , will be a function of treatment group assignment, denoted by Z_{ib} , and observed variables from administrative

16. Prior to May 2011, CPS asked parents and guardians to include SSNs in students’ enrollment information. So any One Summer Plus applicant who was enrolled before that date had the chance to provide SSNs, although the school district did not validate them, nor require their submission. CPS provided the numbers directly to IDES without researcher involvement, and removed them before we received the data. Since the decision to provide an SSN (or a valid SSN) is a pre-program characteristic, we expect the missing data to be balanced across treatment and control groups.

17. We call an SSN “valid” if it was both a) submitted to IDES and b) not removed by CPS because of a name mismatch

records measured at or before baseline, $X_{ib,t-1}$, as in equation 2.1 below:¹⁸

$$Y_{ibt} = Z_{ib}\delta_1 + X_{ib,t-1}\delta_2 + \xi_b + u_{ibt} \quad (2.1)$$

We control for the blocking variable with block fixed effects, ξ_b . The “Intent-To-Treat effect” (ITT) captures the effect of being assigned to the treatment group, and is given by the estimate of coefficient δ_1 in equation 2.1. Although baseline characteristics are not necessary for identification, we include them in the regression to improve the precision of estimates by accounting for residual variation in the outcomes.

The ITT framework fully exploits the strength of the randomized experimental design. Moreover, the coefficient δ_1 in equation 2.1 may be especially useful for policy, as it directly addresses the impact of offering services on the outcome Y . But because not all youth offered the treatment participate, the ITT estimates will understate the effects of actually participating in the program on those youth who participate. Under the typical relevance and exogeneity assumptions for instrumental variables,¹⁹ this latter set of effects can be recovered from the experimental data. We perform this estimation through a two-stage least squares strategy, in which random assignment, Z_{ib} , is an instrument for program participation, P_{ibt} , which is an indicator variable for starting the program, and \hat{P}_{ibt} is the predicted probability

18. Baseline covariates include controls for demographic characteristics and neighborhood characteristics, as well as for pre-program criminal involvement, academics, and formal employment. Demographic controls include indicators for age at the start of the program and for being male, Black, or Hispanic. Neighborhood controls include the census tract’s median income, proportion of those over 25 with a high school diploma or equivalent, and home ownership rate. Crime controls include separate indicators for having been arrested for 1 or 2 or more violent, property, drug, or other crimes. Academic controls include days absent and indicators for the student’s free lunch status, special education status, enrollment status in the year prior to the program (determined by June 2012 CPS enrollment status and 2012 attendance), and grade level, as well as the number of As, Bs, Cs, Ds, and Fs received. Finally, employment controls include indicators for having a valid SSN and for having any formal employment in the year before the program. We impute zeros for missing data and include indicator variables that equal one if a variable was missing, as well as indicator variables for submitting 1 or 2 duplicate applications.

19. In order for the random assignment variable, Z_{ib} , to be a valid instrument, it must be correlated with program participation, P_{ibt} , and uncorrelated with ε_{ibt} . Moreover, if treatment effects are heterogenous, it must shift participation in a uniform way across people. For example, we must assume there are no youth who would participate if assigned to the control group but not if assigned to the treatment group.

of participation from equation 2.2:

$$P_{ibt} = Z_{ib}\pi_1 + X_{ib,t-1}\pi_2 + \gamma_b + \nu_{ibt} \quad (2.2)$$

$$Y_{ibt} = \hat{P}_{ibt}\beta_1 + X_{ib,t-1}\beta_2 + \alpha_b + \varepsilon_{ibt} \quad (2.3)$$

If all youth respond the same way to participating in the program (that is, if treatment effects are constant, or homogenous, across youth), then the coefficient β_1 in the system of equations 2.2 and 2.3 is interpretable as the average treatment effect (ATE) across this population of disadvantaged youth, which will also equal the effects of treatment on the treated (TOT). If treatment effects are heterogeneous across youth, then β_1 represents the local average treatment effect (LATE), or the effect of treatment on youth who complied with random assignment (though in our case, with very little control crossover, the LATE should closely approximate the TOT). To help judge the magnitude of the LATE estimates, we will also estimate the average outcomes of those youth in the control group who would have complied with treatment had they been assigned to treatment - or the “control complier mean” (CCM) (see [38],[44]). Because the differences between the two treatment arms in the 2012 cohort are generally not statistically significant, we focus the main text on the overall treatment-control context; we report results by treatment arm in the appendix.

In any experiment testing program effects on multiple outcomes, not to mention heterogeneous treatment effects by subgroup, one might worry that the probability of Type I error increases with the number of tests conducted. We take a number of steps to ensure that our results are not just the result of data mining. First, we note that because DFSS built the program and recruiting strategy mainly to reduce youth violence, the impact on violent-crime arrests was the primary pre-specified outcome of interest.

Second, we present both unadjusted p-values and p-values which are adjusted using a free-step down permutation method. The step-down method controls the family-wise error rate (FWER), or the probability that at least one of the true null hypotheses in a family

of hypothesis tests is rejected [8, 81].²⁰ The FWER approach is useful for controlling the probability of making any Type I error, but it trades off power for this control. An alternative is to control the probability that a null rejection is a Type I error (the false discovery rate, or FDR), increasing the power of individual hypothesis tests in exchange for allowing some specified proportion of rejections to be false [14, 15]. We focus on define our families of outcomes as: 1) the four types of crime (violence, property, drug, and other, excluding total arrests since it is a linear combination of the rest), 2) the three main schooling outcomes across the subset of the sample that could still be in school (re-enrollment, days present, and GPA) plus high school graduation for those old enough to have finished, and 3) employment and earnings over the whole employment sample) as we find little to no evidence of effects on schooling or employment without these adjustments.

Third, we handle our tests for heterogeneous treatment effects in a number of ways. We report basic tests of interaction effects, focusing on one key pre-specified interaction in the 2013 cohort (whether reaching youth before they leave school and face the full-time labor market is more effective than after they have left school), as well as a few ad-hoc tests that theory and the prior literature suggests may be important (gender, prior arrest history, and prior employment). But rather than conduct more and more hypothesis tests to isolate the subgroups that benefit most, we turn to supervised machine learning algorithms to identify highly-responsive subgroups in a principled way. To this end, we focus on [80]’s Causal Forest algorithm. This algorithm estimates individual treatment effects by averaging estimates over

20. We estimate the distribution of our test statistics accounting for all of the tests within a particular family by randomly sampling permutations of treatment status within blocks and recording all of the test statistics for each permutation. Under the null hypothesis of no treatment effect, each permutation should be identically distributed. Therefore, we are able to approximate the joint distribution of our test statistics with the distribution of the test statistics across permutations. For a particular hypothesis, we are able to estimate a level α critical value, $c(\alpha)$, with the $(1 - \alpha)^{th}$ percentile of the estimated test statistic distribution. For a family of hypothesis tests, we determine the critical values using the step-down procedure outlined in [8]. Specifically, we sort the test statistics within a family of hypothesis tests from largest to smallest. Then we determine the adjusted critical value for the test with the largest test statistic using the distribution of the maximum test statistic within the family across permutations. We then drop the test with the highest test statistic and repeat the procedure for the test with the second highest test statistic. This continues until the last test in the family. We estimate the test statistic distributions using 100,000 permutations.

B regression trees. First, the sample is split in to a training and estimation sample. Then, B regression trees are generated using different subsamples of s observations randomly drawn without replacement from the training sample. Splits in each tree are determined using the splitting rule of [9] which seeks to maximize the variance of estimated conditional average treatment effects in the training sample subject to the constraint that at least k treatment and control observations are in each leaf.

2.6 Descriptive Statistics

Table 2.1 shows select baseline characteristics for 2012 (left panel) and 2013 (right panel) control groups, as well as pairwise tests of treatment-control balance for each covariate conditional on block fixed effects. No more of the differences are significant than would be expected by chance, and tests of joint significance suggest that randomization successfully balanced the two groups (pooling both samples together, $F(45,6758)=1.04$. $P=0.40$).

Youth in both cohorts are over 90 percent African-American and largely from poor households in highly disadvantaged neighborhoods: Median neighborhood income is \$33-36,000 with local unemployment rates around 19 percent. Thirty-eight percent of the 2012 cohort and all of the 2013 cohort is male. Recall that in part to test for heterogeneous program effects on a broader population of youth, the eligibility rules across program years changed. As a result, the 2013 cohort is older (18.4 versus 16.3 years old), more criminally involved (47 versus 19 percent have an arrest record), and less engaged in school (51 versus 98 percent still engaged in school before the program, and those with any attendance missed 3 months versus 6 weeks of the prior school year). Partly because of their age and school status, the 2013 youth are also about twice as likely to have been employed in the prior year (19 versus 8 percent).

Table 2.1: Baseline Balance

Program Year:	2012			2013				
	Control	Control	Control	Control	Control	Control		
	Mean	SD	Coef	SE	Mean	SD	Coef	SE
Age at Program Start	16.30	1.45	-0.05	(0.069)	18.42	1.45	0.03	(0.024)
Black	0.96	0.21	0.00	(0.010)	0.91	0.29	0.01	(0.009)
Hispanic	0.03	0.17	0.00	(0.007)	0.07	0.25	0.00	(0.008)
Any Baseline Arrest	0.19	0.39	0.02	(0.019)	0.47	0.50	.019*	(0.011)
# Arrests: Violent	0.13	0.53	0.04	(0.029)	0.62	1.30	0.03	(0.039)
# Arrests: Property	0.09	0.43	-0.01	(0.021)	0.45	1.22	0.00	(0.036)
# Arrests: Drug	0.05	0.38	0.01	(0.021)	0.69	1.83	-0.05	(0.048)
# Arrests: Other	0.15	0.70	0.02	(0.034)	1.28	2.72	-0.02	(0.077)
In CPS Data	1.00	0.00	0.00	(0.000)	0.91	0.29	0.00	(0.008)
Engaged in CPS in June	0.98	0.13	0.00	(0.006)	0.51	0.50	0.00	(0.013)
Grade in Prior School Year ¹	10.15	1.25	-0.04	(0.061)	10.57	1.01	0.03	(0.033)
Days Attended in Prior School Year ¹	136.92	30.45	0.70	(1.404)	122.78	54.28	2.50	(1.823)
Free Lunch Status ¹	0.92	0.27	0.00	(0.014)	0.86	0.35	0.01	(0.014)
GPA ²	2.37	0.88	0.00	(0.044)	1.95	0.96	-0.03	(0.046)
Has Valid SSN	0.77	0.42	0.03	(0.020)	0.67	0.47	0.02	(0.014)
Worked in Prior Year ³	0.08	0.27	-0.02	(0.014)	0.19	0.40	0.00	(0.014)
Prior Year Earnings ³	265	2047	-155	(124)	392	1696	182.941*	(109)
Census Tract: Median Income	35665	13633	-347	(660)	33752	13627	-119	(367)
Census Tract: Share HS+	72.91	15.82	-0.85	(0.790)	73.98	10.37	-0.34	(0.286)

Notes. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Characteristics are also jointly insignificant. The pooled test statistic is $F(46,6757)=1.02$ ($P=0.43$). For 2012, the test statistic is $F(40,1568)=1.03$ ($P=0.41$). For 2013, it is $F(42,5154)=0.92$ ($P=0.61$). All tests control for block fixed effects and indicator variables for submitting one or two duplicate applications for the 2013 program. The 2012 sample was 38.5% male. ¹ Defined only for the subsample of youth who were enrolled in school in the year prior to the program and were not missing information (2012 $N=1,365$ and 2013 $N=2,919$). The 2011-12 school year had 170 days and the 2012-13 school year had 181 days. ² Grades may be missing for some in school youth who attended charter schools. ³ Sample restricted to youth with valid social security number.

2.7 Participation

We have two ways to measure employment over the summer: program provider records and UI data. Provider records are specific to OSC+, so they do not capture participation in other types of summer programs or in the regular labor force. UI data do capture the latter, but they are limited in two different ways: Not all study youth could be matched to the UI data, and not all program participation shows up in the UI data (because not all program providers reported program participation to the UI system).

Our main goal is to estimate the effect of the program relative to whatever else youth would have done. As such, our first stage measures whether youth participated in the program for at least one day using provider records for the entire sample. Because the nature of the counterfactual is central to understanding what this first stage is estimating, however, we also report the proportions of treatment and control youth working in other summer jobs for the subsample with available UI data.

Table 2.2: Participation

	N	Any Days	All	Days Participants	Worked Most Days All	Participants
A. 2012 Program						
Treatment	730	0.745	26.07	34.99	0.65	0.87
Control	904	0.000	0.00	0.00	0.00	0.00
B. 2013 Program						
Treatment	2636	0.302	5.44	18.05	0.09	0.29
Control	2585	0.004	0.08	20.50	0.00	0.30
C. 2013 Extension						
Treatment	2636	0.199	3.68	18.53		
Control	2585	0.004	0.18	42.45		

Notes. "Worked Most Days" is defined as working 30 or more days in 2012 and 25 or more days in 2013.

Table 2.2 shows that 75 percent of youth offered the program actually participated in the

first program year, and participants averaged 35 days of work out of a possible 40. Of those who started the program, about 90 percent were still working during the 7th week of the program (the 8th week was an optional extension) and nearly 60 percent worked at least 35 days. In the second program year, when the maximum possible take-up rate was 38 percent by construction (see section II), actual program take-up was 30 percent. Because of how many more names were on the treatment list than available slots, not everyone in the treatment group was contacted about the program; among those whom providers report contacting, take-up was 53 percent. Participants worked an average of 18 days out of a possible 30, reflecting in part the greater challenge of recruiting and retaining a more disconnected and criminally-active population in the second year. There was no control crossover in the first cohort; 10 control youth in the second cohort (0.39 percent) participated in the program. 20 percent of the 2013 treatment group participated in any extension programming. On average, these participants attended about 18.5 days of additional programming over about a 9 month period.

Table 2.3 shows program participation and other summer employment for the sample of youth with valid SSNs who could be matched to UI data.²¹ UI data are quarterly, and the 2012 program started in the last week of June. So we define the program period as quarters 2 and 3 of 2012 (April - September) in the first study year and quarter 3 only (July - September) in the second study year, when the program started at the beginning of July. The 2012 cohort is generally less likely to be employed than the 2013 cohort: about 9 percent of the 2012 treatment group and 17 percent of the control group work in another job. In 2013, 19 percent of the treatment group and 24 percent of the control group work outside of

21. The table assumes anyone marked as a program participant actually worked in the program, even if they do not show up in the UI data. This can occur because some program providers considered program wages to be a stipend and so did not report employment to the state; the patterns of participation look almost identical when excluding the non-reporting agencies (not shown). Conversely, the table also assumes anyone who is not marked as a program participant did not participate in OSC+ (some non-participants do earn money over the summer from the same agencies that run OSC+, likely from the other summer programming those providers offer). Because not all summer programming involves wages (and because some agencies do not report into the UI system), we may understate broader participation in summer programming outside the formal labor market.

the program. This suggests that the program generates a small amount of crowd-out, though it still dramatically increases the overall proportion of youth who work over the summer: The treatment-control difference in having no job is 62 percentage points in 2012 (from a control mean of 83 percent) and 23 percentage points in 2013 (from a control mean of 75 percent).

Table 2.3: Program Participation and Other Employment

Program Only	Program + Other Job	Other Job Only	No Job	Program Only	Program + Other Job	Other Job Only	No Job
A. 2012 Cohort							
Treatment (N=586)				694 Control			
412	28	23	121	0	0	115	579
70%	5%	4%	21%	0%	0%	17%	83%
B. 2013 Cohort							
Treatment (N=1826)				1735 Control			
513	75	283	955	7	3	416	1309
28%	4%	15%	52%	0%	0%	24%	75%

2.8 Main Results

Table 2.4 shows our main crime results, which use the number of arrests of each type as the dependent variable (note coefficients and standard errors are multiplied by 100, so they represent the treatment effect per 100 youth). In order to make the estimates easy to compare across program years with different take-up rates, we focus on the LATE; the appendix shows the ITT results. As described above, we have arrest data through the 10th month of year 3 for the 2012 cohort and the 9th month of year 2 for the 2013 cohort.

Panel A of Table 2.4 pools together both study cohorts. Although the program and study populations were slightly different across years, none of the crime results are significantly different across cohorts. So Panel A maximizes our statistical power to test whether the program in general affects crime. There is a statistically significant decline in violent-crime arrests during the first year: 4.6 fewer arrests for a violent crime per 100 participants, a 29 percent decline relative to the control complier mean.²² We also see proportionally large but not statistically significant increases in property and drug arrests during year 1, such that - consistent with prior studies of youth employment programs that do not disaggregate crime by type - there are no significant changes in the number of total arrests. The decline in violent-crime arrests does not continue in the second year, when the point estimate is positive but not statistically significant.²³ Given the disproportionately large social costs of violence, as we discuss below, even a short-term decline may be socially beneficial.

One obvious concern is that only one outcome shows a statistically significant decline across four different types of crimes over two years. Since a decline in violence was the

22. Excluding program months, there is a statistically insignificant 24% decline in the pooled sample in the first year.

23. The positive point estimate is largely driven by the 2013 cohort, as shown in panel C, and may be a mechanical result of higher incarceration levels among the control youth stemming from somewhat more arrests in year 1 (incarceration temporarily reduces arrests to zero). This kind of bias might change the interpretation of the results; it is not clear whether the year 2 results reflect underlying changes in behavior or different levels of physical incapacitation. However, as long as we include the social costs of incarceration in our benefit-cost calculations, we can still ask whether spending on the program generates social benefits relative to what would have happened in the absence of the program, including the incarceration of the control group.

Table 2.4: Impact of One Summer Plus 2012 and 2013 on Arrests

Crime:	Total (1)	Violent (2)	Property (3)	Drugs (4)	Other (5)
A. Pooled Samples					
Year One	-1.201 (6.669)	-4.5397** (2.082)	2.3417 (1.753)	3.5843 (2.897)	-2.5873 (4.545)
CCM	68.577	15.528	8.115	12.011	32.924
Year Two	0.4361 (5.171)	1.2292 (1.589)	0.1481 (1.282)	-2.0394 (2.243)	1.0982 (3.422)
CCM	42.383	6.218	4.22	12.724	19.22
B. 2012 Sample					
Year One	-2.2247 (4.722)	-4.5981** (1.974)	0.9648 (1.248)	0.6267 (2.064)	0.7819 (2.495)
CCM	28.879	10.113	3.447	4.153	11.167
Year Two	-4.1174 (4.133)	0.0781 (1.590)	0.7427 (1.266)	-2.4645 (1.618)	-2.4737 (2.549)
CCM	25.625	4.701	2.566	7.428	10.93
Year Three	-0.0892 (3.805)	-1.2168 (1.336)	1.9178** (0.919)	-1.8222 (1.731)	1.032 (2.191)
CCM	15.53	3.974	-0.263	5.315	6.505
C. 2013 Sample					
Year One	-4.8423 (11.486)	-5.0755 (3.431)	3.1788 (3.028)	4.8488 (5.021)	-7.7944 (8.052)
CCM	100.36	19.851	11.475	18.192	50.842
Year Two	0.7511 (8.747)	1.7182 (2.583)	-0.913 (2.084)	-2.6749 (3.854)	2.6208 (5.894)
CCM	56.828	7.583	6.005	17.326	25.914

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Outcomes are arrests per 100 youth per month. The 2012 sample included 1,634 youth and the 2013 sample included 5,221 youth. The second year includes only 9 months of data for the 2013 sample. Standard errors are clustered by individual.

primary pre-specified hypothesis, one might argue that the risk of false positives generated by data mining here is quite low. Nonetheless, the finding is robust to different adjustments for multiple hypothesis testing within years.²⁴ In particular, the reduction in year one arrests for violent crime remains significant at the 10% level after adjusting the inference to control the FWER across the four crime categories and is nearly significant after controlling for the FDR ($q=0.116$).

Panels B and C of Table 2.4 break out the crime effects by cohort (and show the part of year 3 for which we have data in the 2012 cohort). The basic pattern of results is the same across program years - a proportionally large decline in violence - though it is not significant in the 2013 cohort (as discussed below, the decline is significant for the group of youth who look more like the 2012 cohort). There is also a large increase in property crimes in year 3 for the 2012 cohort, which might indicate some longer-term changes in behavior (e.g., if youth are working more after leaving school, they might have more opportunity for theft), or might be a byproduct of the surprisingly low control complier mean in that year. If we test the cumulative effect over all the available data for the 2012 cohort (not shown), both the violence decline and the property crime increase are significant at the 10 percent level (cumulative violent-crime effect = -5.56, SE = 3.223, CCM = 18.6; cumulative property crime effect = 3.59, SE = 2.08, CCM = 5.78). This highlights one potentially policy-relevant fact about violence: given its concentration among youth in their late teens, reducing violent crime for even a short period of time may end up having a lasting effect on violence involvement over the life course.

The 2013 cohort shows no overall statistically significant changes. The CCMs highlight that this was a much more criminally active group; over 21 months, the control compliers

24. We perform the adjustments separately by follow-up year, in part because fade-out is almost universal in social programs. This allows us to determine if the program generated any change in behavior, rather than a cumulative change in behavior. Breaking the effects down annually is useful to get a sense of the time pattern of program effects; however, we recognize that the division of effects by year is somewhat arbitrary. We take a more disciplined approach to aggregating effects over time below, where we assign social costs to outcomes and calculate the present discounted value of the future stream of effects based on when the changes occur.

averaged about 157 arrests per 100 youth (compared to about 54 in the 2012 cohort). Recall, however, that the hypothesis going into this study year was that some of the youth would respond to the program, while others might be too disconnected or too discouraged by facing the full-time labor market on their own to benefit. Below we discuss these heterogeneous effects and show that the youth who look more like the 2012 cohort show significant declines in violence as well.

Table 2.5 shows the impact on measures of formal employment. The table shows estimated program effects on the probability of being employed for the sample of youth with valid SSNs (the appendix shows that results using various imputation techniques for missing data do not change the conclusions) who were assigned to an employment provider who appears in the formal employment records. As expected, there is a large increase in formal employment, driven by greater employment at program providers, during the program quarters.

To exclude the main summer program effect, we show employment during the remainder of the first year after the program (the 4th quarter of the program year and 1st quarter of the following year in 2012 and also the 2nd quarter of the following year in 2013) as well as the full second year. The results in Table 2.5 suggest that there may have been a small increase in employment in the period just after the program, but that the improvement is likely driven by the extended programming in 2013. Overall, there is a statistically insignificant 5.4 percentage point increase in employment during the post-program quarters in year one, which is a 28 percent increase from the control complier mean of 19.9 percent. This effect is driven entirely by the 2013 cohort, which is when program activities were also offered in the post-summer school year. Column 2 suggests that the positive employment effect in the post-program year is driven largely by this program extension; there is a 4.9 percentage point increase in employment at employment providers. Among other employers, there is only a statistically insignificant 0.9 percentage point increase in employment. In other words, the immediate post-summer employment increase seems to be a mechanical effect of the program.

Table 2.5: Impact of One Summer Plus 2012 and 2013 on Formal Employment

Outcome:	Any Formal Employment	Provider Employment	Non-Provider Employment	Earnings
	(1)	(2)	(3)	(4)
A. Pooled Samples				
During Program	0.7610*** (0.032)	0.9204*** (0.019)	-0.0563* (0.029)	606.3215*** (100.030)
CCM	0.116	-0.056	0.167	147.913
Rest of Year One	0.0537 (0.033)	0.0490*** (0.010)	0.0085 (0.033)	130.8872 (196.449)
CCM	0.199	-0.002	0.200	518.733
Year Two	0.0212 (0.037)	0.1002*** (0.018)	-0.0265 (0.037)	89.1137 (321.184)
CCM	0.479	0.050	0.438	1531.513
Overall	0.4076*** (0.034)	0.8716*** (0.022)	-0.0287 (0.036)	826.3224 (538.192)
CCM	0.504	0.001	0.494	2198.159
B. 2012 Sample				
During Program	0.8147*** (0.037)	0.9858*** (0.028)	-0.0624** (0.031)	575.4548*** (141.727)
CCM	0.069	-0.105	0.169	403.365
Rest of Year One	0.0202 (0.029)	0.0202 (0.029)	-	19.0759 (144.988)
CCM	0.113	0.113		331.932
Year Two	0.0152 (0.043)	0.0413* (0.025)	0.0033 (0.042)	29.8382 (319.104)
CCM	0.457	0.048	0.405	1455.405
Overall	0.4615*** (0.038)	0.9310*** (0.031)	-0.0125 (0.042)	624.369 (547.947)
CCM	0.465	-0.037	0.468	2190.702

Table 2.5: Impact of One Summer Plus 2012 and 2013 on Formal Employment - Continued

Outcome:	Any Formal Employment	Provider Employment	Non-Provider Employment	Earnings
C. 2013 Sample				
During Program	0.7330*** (0.047)	0.8764*** (0.026)	-0.0412 (0.044)	651.9116*** (135.163)
CCM	0.141	-0.021	0.155	-23.605
Rest of Year One	0.0873* (0.051)	0.0787*** (0.015)	0.0144 (0.050)	252.2273 (306.402)
CCM	0.233	-0.006	0.237	564.605
Year Two	0.0298 (0.054)	0.1369*** (0.026)	-0.0383 (0.054)	253.6335 (482.029)
CCM	0.486	0.047	0.452	1442.572
Overall	0.3756*** (0.050)	0.8302*** (0.030)	-0.0312 (0.052)	1157.7724 (804.904)
CCM	0.528	0.030	0.502	1983.572

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample is restricted to youth with a CPS record and a valid social security number in the CPS system who were assigned to a provider that reported employment information to the Illinois Department of Employment Security. The 2012 sample includes 1,280 youth and the 2013 sample includes 3,561 youth.

We also see a substantively large increase in employment at program providers, but not at other employers, for the 2013 cohort during the following year (14 percentage points relative to a control complier mean of 5 percent). This pattern suggests that youth may have formed relationships with program providers over the course of the year, such that they continue to work in government-sponsored programs in future years. Longer follow-up data will be necessary to test whether this continued work experience translates into better future employment outcomes after youth age out of summer programming.

Table 2.6 shows results for schooling outcomes for the subset of youth with a CPS record

who had not graduated prior to the first post-program school year.²⁵ Column (1) shows there is basically no effect on the probability of attending any days of school in the first post-program school year. There is a small and insignificant decline in the probability of attending any days of school in the second post-program school year. Since youth who have already graduated are not expected to attend any days, we impute these youth as having attended at least one day.

25. Analogous results restricting the sample to youth who had not graduated prior to random assignment are in a separate appendix.

Table 2.6: Impact of One Summer Plus 2012 and 2013 on Schooling

Outcome	Any Days Present		# Days Present		GPA		Graduated		
	Imputation Method	0 if Missing (1)	None	Group Means (2)	Regression	None	Group Means (3)	Regression	None (4)
Panel A. Pooled Samples									
Year One		0.0021 (0.020)	-1.4234 (2.714)	-1.6463 (2.663)	-1.6463 (2.663)	-0.0095 (0.056)	-0.0941** (0.038)	0.1040*** (0.039)	0.0053 (0.017)
CCM		0.745	119.871	92.238	92.238	2.056	2.077	1.857	0.198
Year Two		-0.0336 (0.022)	5.3433 (3.810)	-6.0120** (2.857)	-0.5622 (2.816)	0.076 (0.082)	-0.3443*** (0.034)	0.0923** (0.041)	-0.0277 (0.021)
CCM		0.683	107.975	83.372	74.489	2.109	2.388	1.938	0.375
Panel B. 2012 Sample									
Year One		0.0047 (0.013)	-2.4905 (2.183)	-2.5504 (2.610)	-2.5504 (2.610)	-0.0577 (0.051)	-0.0737 (0.047)	-0.0597 (0.044)	0.018 (0.017)
CCM		0.943	139.653	132.783	132.783	2.275	2.292	2.224	0.281
Year Two		-0.0336 (0.022)	3.1794 (3.404)	-2.0801 (3.312)	-3.7233 (3.153)	0.1674** (0.079)	0.0629 (0.044)	0.4256*** (0.045)	-0.0076 (0.021)
CCM		0.683	133.207	123.264	114.629	2.199	2.305	2.105	0.504

Table 2.6: Impact of One Summer Plus 2012 and 2013 on Schooling - Continued

Outcome	Any Days Present		# Days Present		GPA		Graduated	
	0 if Missing (1)	None	Group Means (2)	Regression	None	Group Means (3)	Regression	None (4)
Panel C. 2013 Sample								
Year One	-0.0011 (0.036)	1.8298 (5.823)	0.0687 (4.512)	0.0687 (4.512)	0.1938 (0.148)	-0.0882 (0.059)	0.2910*** (0.063)	-0.0124 (0.028)
CCM	0.587	93.509	58.722	58.722	1.509	1.882	1.508	0.139
Year Two	-0.0457 (0.040)	11.9773 (7.974)	-9.2345** (4.541)	3.0203 (4.497)	-0.0483 (0.233)	-0.7451*** (0.055)	-0.2348*** (0.069)	-0.0489 (0.035)
CCM	0.51	72.658	51.434	41.243	1.824	2.528	1.863	0.277

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The sample is restricted to youth with a CPS record who had not graduated prior the first post-program school year. The 2012 sample includes 1,632 youth and the 2013 sample includes 4,020 youth. # Days Present is imputed with either treatment or control group means ("group means") or predicted values from separate treatment and control group regressions of the outcome on our set of baseline characteristics ("regression") for youth who graduated post-program, otherwise it is imputed as 0. GPA is imputed for all youth. Any days present is generated based on the imputed measure of attendance.

The results for the number of days attended are sensitive to how we treat graduates.²⁶ If we just restrict the sample to non-missing observations, there is an insignificant 1.4 day decline and 5.3 day increase in days attended in the first and second post-program school years, respectively. Restricting attention to non-missing cases may be undesirable as attending any days is an outcome itself, so we impute missing observations in two ways. First, we impute treatment and control group means for post-program graduates. As there are no post-program graduates at the start of year one, this approach just imputes zeros for youth missing attendance in the first post-program school year. This has little effect on the impact on days attended in the first post-program school year. However, with this imputation, the program appears to reduce attendance in the second post-program school year by 6 days. This decline is driven entirely by the 2013 cohort. Second, we use a regression based approach that imputes the predicted value from separate treatment and control group regressions of days present on baseline characteristics using the non-missing cases. Using this imputation method, we do not see evidence of an impact on the number of days attended in either the first or second post-program school years.

As with attendance, the program's impact on GPA is also sensitive to our imputation method. Here, the imputation method also affects the year one result as grades are missing for some in school youth who attend charter schools. Using non-missing cases, we see virtually no difference in GPAs in either the first or second post-program school year. If we impute group means, we find relatively large and significant declines in GPAs in both school years, but if we use the regression approach we find significant increases in GPAs.

We see no impact on graduation by the end of the first post-program school year. There is a small and statistically insignificant reduction in graduation by the end of the second post-program school year.

26. In the present set of results, we assume no youth attend private schools or public schools outside of Chicago.

2.9 Who Benefits From Summer Jobs?

Across crime, schooling, and employment outcomes, the large and significant impacts on crime likely drive any social benefits or costs of the program. Table 2.7 presents estimates of the social savings derived from the program's impact on crimes. For this analysis, we assign estimates of the social costs of crime to determine the value of the program's impacts on crime. The first column in each pair estimates the social costs of crime using contingent valuation (CV)—individuals' reported willingness to pay to avoid different types of crime [22]. The second column in each pair uses estimates of the direct costs of crime from [54]. Each pair of columns presents estimates making different adjustments to the costs of crime. Moving across columns, the four sets of results differ in whether they account for the social cost of drug use, make adjustments to convert the number of arrests in to the number of crimes, and inflate the costs of arrests to account for the collateral costs of incarceration. Differences in the estimates are driven by whether direct or CV costs are used and whether the number of arrests is converted to account for the fact that not all crimes result in an arrest.

Caution is warranted when interpreting the results as all of the estimates are quite imprecise. Using only observed crimes, the program's impacts are worth between \$1,138 and \$2,627 per participant, but these benefits fade somewhat over time. These results suggest most, but not all, of the program's costs are offset by the social savings derived from reduced crime.

Table 2.7: Estimated Social Savings per Participant from Crime Reduction

	(1)	(2)	(3)	(4)				
Year One	-2627 (2490)	-1718 (1099)	-2047 (2506)	-1138 (1140)	8756 (40414)	-9744 (13983)	9039 (41272)	-9876 (14252)
Year Two	724 (2006)	229 (874)	916 (2025)	421 (914)	32785 (25020)	4832 (8644)	33575 (25551)	4994 (8808)
Cumulative	-2278 (3533)	-1579 (1554)	-1572 (3559)	-873 (1610)	41637 (57549)	-5122 (19998)	42721 (58770)	-5094 (20381)
2012 Cumulative	-3356 (3589)	-1331 (1639)	-3538 (3615)	-1513 (1684)	-12496 (14114)	-5550 (5916)	-12728 (14403)	-5654 (6030)
2013 Cumulative	-2830 (5560)	-2573 (2411)	-1622 (5599)	-1365 (2503)	74150 (104487)	-9658 (36189)	76072 (106704)	-9613 (36881)
Source of cost estimates	CV	Direct	CV	Direct	CV	Direct	CV	Direct
Includes:								
Collateral costs of incarceration								X
Adjustment for crimes per arrest						X		X
Social cost of drug use			X					X

Notes: n =6,855, ***p<0.01, ** p<0.05, *p<0.1. Heteroskedasticity-robust standard errors in parentheses. All estimates in 2012 dollars. Columns use different social costs of crime. Homicide trimmed to cost of aggravated assault. Columns using "direct" costs based on estimates from Cohen & Piquero (2009); columns using CV based on contingent valuation estimates from Cohen (2001). See text and Appendix A for details. Columns without social cost of drug use only count direct cost of drug arrests to the criminal justice system. Benefits are dis-counted monthly at a 5% annual rate beginning in mid-September following the program.

When the number of arrests is inflated to account for the underlying number of crimes, as in specifications 3 and 4, the results flip sign depending on whether direct or CV costs are used. This is driven by the fact that the CV approach typically assigns a greater social cost to all types of crime, but property arrests for property crimes-where we see economically large but insignificant increases in arrests-are inflated much more than violent crimes. With the CV approach, the program has an adverse impact on social welfare in the first year and cumulatively. On the other hand, with direct costs, the program yields a social savings of over \$2,000 per participant after subtracting the program cost of about \$2,800 per participant.

The detrimental effects found in specifications 3 and 4 when CV costs are used are driven entirely by the 2013 cohort. Using the CV costs and specification 4, the program created an insignificant \$12,278 cumulative decrease in social costs from crime among the 2012 cohort, but a \$23,038 increase in social costs among the 2013 cohort. Given the substantial differences in the 2013 and 2012 samples, these differences in the impact on social costs across program years suggest there may be important treatment heterogeneity.

Given the imprecision of the social cost estimates and that the explicit goal of the program is violence reduction, we explore heterogeneity in the program's impact on arrests for violent crime in the first post-program year. Table 2.8 shows estimates of the impact of One Summer Plus on arrests in the first post-program year among subgroups defined by whether youth were enrolled in school in the June prior to the program, had any prior arrests at baseline, or worked in the year prior to the program. Our key pre-specified hypothesis was about school - serving youth before they face the full time labor market on their own seems to be a key difference between summer jobs programs and other programs that have all been for disconnected youth. Panel A of Table 2.8 provides suggestive evidence that prior school enrollment may be an important dimension of treatment heterogeneity. The decline in arrests for violent crime is driven entirely by in school youth for whom there is a 44% reduction in arrests for violent crime, equal to 6.5 fewer arrests per 100 youth, whereas there is virtually no difference in arrests for violent crime among out of school youth. However, these effects

are not significantly different from each other at conventional significance levels.

This decrease in violent crime among in school youth comes at the expense of a statistically significant 75 percent increase in arrests for drug crime, or 5.9 additional arrests per 100 youth, whereas there is a small decrease in arrests for drug crime among out of school youth. As with violent crime, the in and out of school drug crime effects are not significantly different from each other.

Panel B shows treatment effects separately for youth with and without a baseline arrest. In this case, most effects go in the same direction but are amplified for the youth with a baseline arrest. This is largely because youth with a baseline arrest are much more criminally involved in the first post-program year, with 123 arrests per 100 youth, compared to 13 arrests per 100 youth among youth without a baseline arrest. As with schooling, none of these effects are significantly different across the two groups.

Finally, Panel C shows how treatment effects vary with whether or not the youth had work experience in the formal sector in the year prior to the program. Here, we see that the proportionally large increase in property crime in the full sample is driven by youth with prior work experience. Youth who worked in the year prior to the program had 11.7 additional arrests for property crime per 100 youth. This effect is significant at the 10 percent level. There is no evidence of this adverse effect among youth without prior work experience. In addition to not having the adverse property crime effect, youth without prior work experience also had a statistically significant reduction of 5 violent crime arrests per 100 youth. Again, none of these subgroup differences were significantly different across groups.

While the results in Table 2.8 provide suggestive evidence that there may be several important dimensions of treatment heterogeneity, caution is warranted when doing this type of subgroup analysis. False positives, or spurious large and “significant” effects, become more likely as the number of tests increases. To better guard our analysis against false positives while still trying to provide policy relevant insights about which subgroups should optimally be offered the treatment, we adopt the Causal Forest algorithm of [80].

Table 2.8: Heterogeneity in Impacts of One Summer Plus 2012 and 2013 on Year One Crime

Sample Characteristics	Total (1)	Violent (2)	Property (3)	Drugs (4)	Other (5)
A. Treatment Heterogeneity by Baseline School Enrollment					
In School	0.5758 (6.369)	-6.4593*** (2.160)	0.5004 (1.783)	5.9077** (2.772)	0.627 (3.906)
Out of School	55.947 (17.598)	14.565 (5.110)	8.383 (4.620)	6.944 (7.903)	26.055 (12.334)
P(Difference=0)	100.832 0.6094	16.866 0.2003	7.522 0.1912	23.19 0.3556	53.253 0.2279
B. Heterogeneity by Prior Arrests					
Has Baseline Arrest	-4.6764 (14.211)	-7.9382* (4.345)	2.9203 (3.692)	6.4868 (6.146)	-6.1454 (9.850)
No Baseline Arrests	123.335 (4.054)	27.404 (1.413)	14.871 (1.215)	20.622 (1.848)	60.438 (2.205)
P(Difference=0)	13.08 0.7264	4.398 0.2058	1.166 0.7627	2.162 0.3946	5.354 0.5446
C. Heterogeneity by Prior Work Experience					
Worked Prior Year	20.8411 (23.173)	0.4062 (7.378)	11.7123* (6.501)	4.9025 (10.051)	3.8201 (14.565)
Did Not Work Prior Year	46.574 (7.756)	13.654 (2.484)	-1.352 (2.075)	11.451 (3.508)	22.82 (4.955)
P(Difference=0)	-0.9649 66.425 0.3744	-5.0620** 15.368 0.4792	0.6419 10.343 0.1104	4.9925 10.863 0.9932	-1.5373 29.851 0.7308

Notes: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Panel A includes youth with a CPS record. Panel B includes all youth. Panel C includes youth a valid SSN.

The Causal Forest algorithm adapts regression tree algorithms from the supervised machine learning literature to the problem of causal inference. Regression trees build potentially complex non-linear models of an outcome as a function of a large set of covariates, or features, by recursively partitioning the data a single covariate at a time. Potential splits are evaluated by a penalized in-sample goodness-of-fit criterion like Mean Squared Error, where the penalty parameter is chosen using cross-validation in order to maximize the out-of-sample predictive accuracy of the selected model. In traditional applications, regression trees are used to model observable outcomes so predictive accuracy can be measured directly. When applied to causal inference, predictive accuracy cannot be directly assessed since treatment effects are not observed for any individual. [9] propose several novel in- and out-of-sample goodness-of-fit measures to generalize this approach to estimating conditional average treatment effects. [80] extend this methodology to random forests, derive the limiting distribution of the estimates, and provide a method for constructing asymptotically valid confidence intervals.

Specifically, [80]’s causal forests algorithm yields consistent estimates of individual estimates by generating B regression trees using subsamples of the training sample of size s and averages the predictions across the B trees. Asymptotically valid point-wise standard error estimates are estimated using the infinitesimal jackknife [41, 79, 80]. These standard errors are proportional to the covariance of individual estimates across trees with weights that vary according to whether an observation was included in the subsample used to generate the estimate associated with a particular tree. We implement the causal forests algorithm using a beta version of [80]’s **causalForest** package in R with $B=1000$ trees and subsamples of size $s=n/2$, where n is the size of the training sample.

We restrict attention to covariates which policy makers could reasonably incorporate in to eligibility restrictions: age and gender, number of violent, property, drug, and other prior arrests, grade and attendance if youth attended school in the school year before the program, and any formal work experience in the year before the program for youth with a valid SSN.

We follow the recommendation of [67] and estimate separate “reduced” causal forests using non-missing covariates for four subgroups with different missing data patterns: 1) youth who were in school prior to the program and had a valid SSN number; 2) youth who were in school prior to the program and did not have a valid SSN number; 3) youth who were out of school and had a valid SSN number; and 4) youth who were out of school and did not have a valid SSN number.

To account for different treatment probabilities across blocks, we scale treatment outcomes by $\frac{1}{p_b}$ and control outcomes by $\frac{1}{1-p_b}$, where p_b is the block specific treatment probability. Estimates are interpretable as intent-to-treat estimates as we do not adjust for differential participation rates across subgroups.

Figure 2.1: Density of Causal Forest Year One Violent Crime Estimates

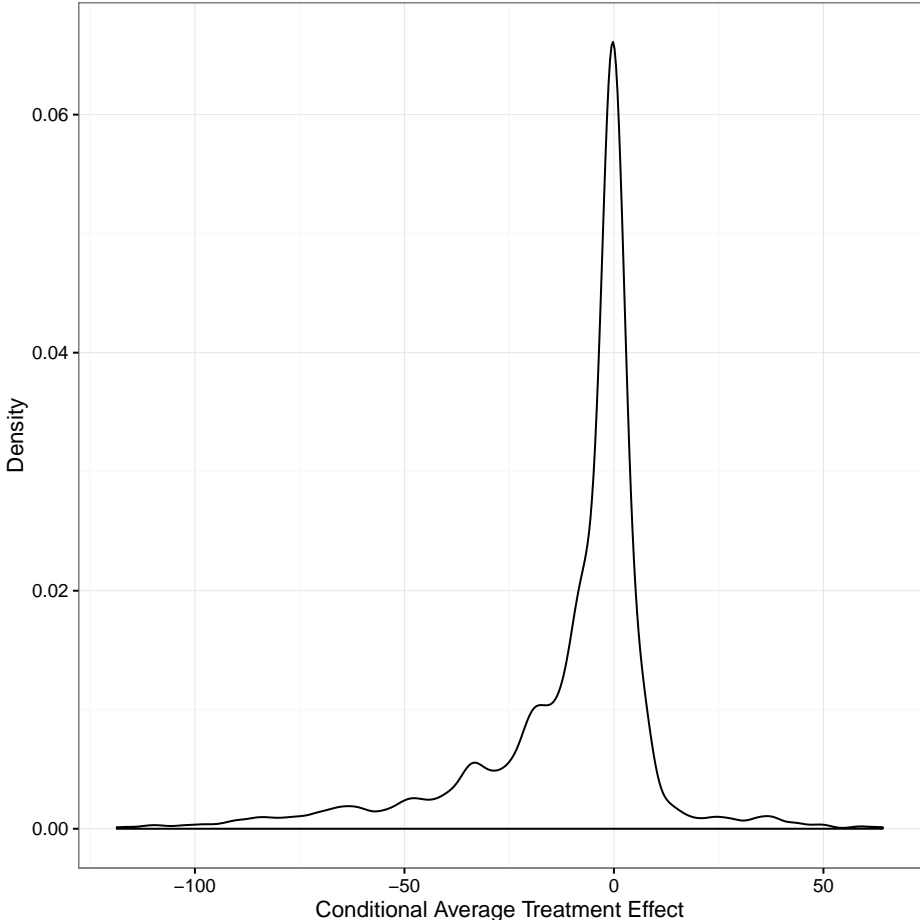


Figure 2.1 shows the density of the conditional average treatment effect (CATE) point

estimates. While the modal value of the density is approximately zero, The average effect is -9.84 and the median effect is -2.88 These are both larger than the comparable year one violent crime ITT estimate of -1.84. The density is skewed towards negative values; the estimated conditional average treatment effect estimates are positive for 19.7% of individuals in the prediction sample and negative for the other remaining 80.3% of individuals. Figure 2.2 shows the point estimates - sorted from least to greatest - and 95% confidence intervals. Point estimates which are significant at the 5% level are in blue, insignificant estimates are in red. Most of the point estimates are imprecise. 5.7 percent and 1.3 percent of coefficients are significantly negative and positive, respectively. We can reject the null hypothesis of no heterogeneity in conditional average treatment effects. As a conservative test for treatment heterogeneity, we test whether any of the 3,457 conditional average treatment effects in the prediction sample are significantly different than the average effect of -9.84 using a Bonferroni corrected critical value.²⁷

Table 2.9 shows estimates of a linear regression of the estimated CATEs and indicators for whether the CATEs are negative, significantly negative, or significantly positive on the same set of covariates used in the causal forests. These estimates provide a descriptive summary of some of the important dimensions of heterogeneity in the data, but surely miss many important interactions between covariates. The R2 of the linear CATE model is only 0.42. If the less restrictive causal forest estimates were actually linear we would expect an R2 of 1. On average, arrests for violent crime appear to be particularly responsive among males, youth with arrests for violent or other crimes, in school youth with relatively poor attendance, and youth with prior work experience.

To better take advantage of the added complexity allowed by the causal forest algorithm, we examine characteristics of youth with significantly positive and negative CATE estimates. The CATE estimates are only significantly positive-meaning an adverse effect on arrests for violent crime-for 44 youth. 43 of these youth are males. 77 percent are 19-21 years old

27. To maintain a maximum 5% probability of false rejection across 3,457 tests, the critical value is 4.34.

Table 2.9: Selected Partial Correlations with Conditional Average Treatment Effects

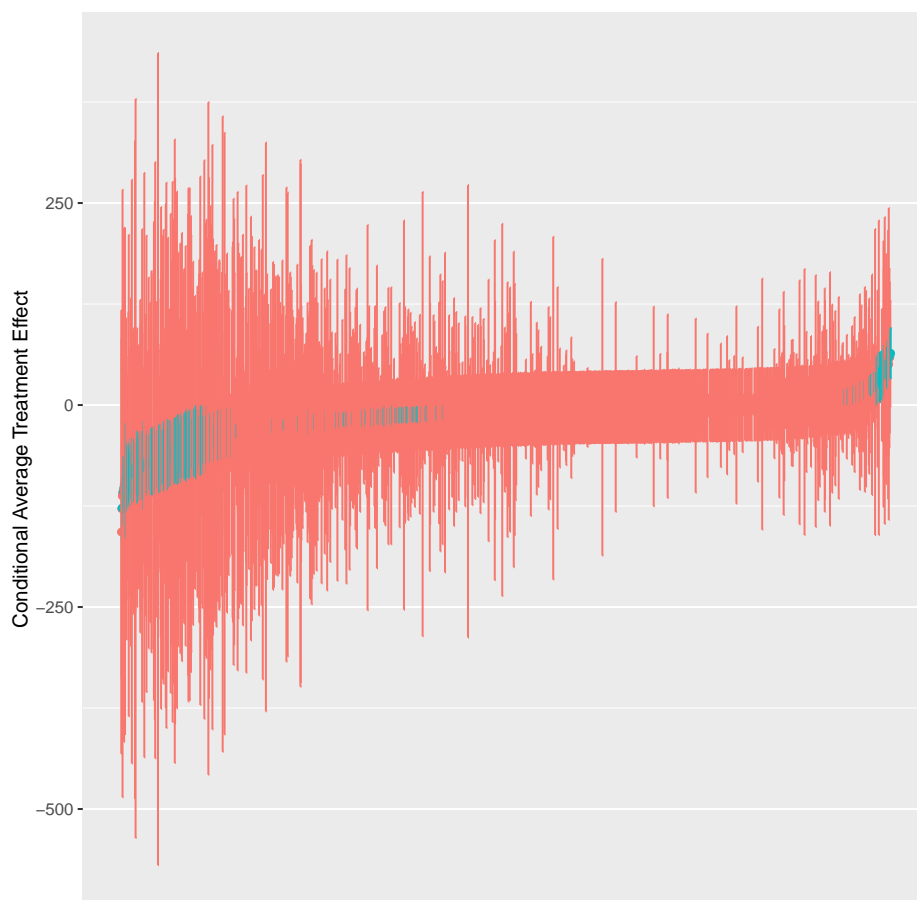
	CATE	CATE<0	Sig. Negative	Sig. Positive
	(1)	(2)	(3)	(4)
Male	-3.7073*** 0.6139	5.0374** 2.295	1.2694 1.0337	-0.9516*** 0.2421
Violent Arrests: 1	-8.9284*** (1.046)	8.5143*** (1.861)	7.4912*** (1.455)	0.3729 (0.611)
Violent Arrests: 2+	-18.1778*** (1.386)	8.0076*** (2.059)	12.5879*** (2.082)	3.4447*** (0.982)
Property Arrests: 1	-1.1123 (1.054)	-4.5925** (2.081)	0.0927 (1.529)	-0.0754 (0.769)
Property Arrests: 2+	-1.9636 (1.521)	-7.2435*** (2.428)	-1.5791 (2.034)	0.9645 (1.194)
Drug Arrests: 1	-4.6077*** (1.277)	4.6954** (2.064)	0.3989 (1.771)	-0.9619 (0.631)
Drug Arrests: 2+	1.0264 (1.529)	-12.3171*** (2.249)	0.6664 (2.032)	3.4172*** (1.048)
Other Arrests: 1	-2.2544** (0.951)	-2.7741 (2.284)	1.9358 (1.392)	-0.463 (0.381)
Other Arrests: 2+	0.1107 (1.207)	-6.5092*** (2.172)	3.1287* (1.774)	1.9460*** (0.727)

Table 2.9: Selected Partial Correlations - Continued

Grade: <9	-14.4588*** (3.721)	5.1861 (8.972)	3.6968 (4.739)	-5.2263** (2.200)
Grade: 9	-19.2421*** (1.915)	22.0155*** (3.382)	7.9644*** (2.676)	-2.8505*** (0.607)
Grade: 10	-9.5303*** (1.376)	15.5755*** (3.075)	-1.524 (1.783)	-1.7890*** (0.497)
Grade: 11	-5.6097*** (1.048)	15.5808*** (3.057)	-1.0192 (1.397)	-1.1116*** (0.380)
Grade: 12	-3.3987*** (0.894)	9.5222*** (3.065)	-2.6623** (1.047)	-0.9681** (0.472)
Attendance Quantile: 1	-13.6582*** (1.328)	16.2278*** (2.677)	2.8597 (1.867)	-2.9427*** (0.697)
Attendance Quantile: 2	-12.0817*** (1.084)	15.4063*** (2.460)	4.7104*** (1.465)	-1.5721*** (0.317)
Attendance Quantile: 3	-4.8492*** (0.824)	4.5209 (2.771)	2.3754** (1.136)	-0.8728*** (0.210)
Attendance Quantile: 4	-1.8919*** (0.675)	1.3375 (2.829)	2.5789*** (0.937)	-0.1423 (0.297)
Worked Prior Year	-1.8887** (0.748)	-3.4148 (2.251)	-0.1518 (0.924)	-1.6931*** (0.424)

Notes. The sample includes the 3,457 observations in the Causal Forest prediction sample. Outcomes are generated using Causal Forest estimates of conditional average treatment effects. Indicators for age and missing gender also included.

Figure 2.2: Causal Forest Point Estimates with 95% Confidence Intervals



and all had at least one prior arrest, with an average of 18.5 prior arrests. 7 percent of the males were in school prior to the program.²⁸ Only one youth worked in the year before the program.

The CATE estimates are significantly negative for 198 youth, 171 males and 21 females. 62 percent of the females had no prior arrests. All of the females were in school since they were only in the 2012 sample. They attended 146 days of school on average. 17 of the females were freshmen. Among the 171 males, all but 6 had at least one prior arrest (97%) and, on average, these males had 7.8 arrests. 138 were enrolled in school in the year prior to the program—112 freshmen or sophomores and 23 juniors or seniors, but they only attended

28. All females in the sample were in school since the 2013 sample is all male.

90 days of school, on average. The 39 males who were not involved in school look somewhat similar to the subset of youth with significantly positive effects only slightly older—19.7 versus 17.5 years old—and less criminally involved - 7.8 versus 19 prior arrests, on average.

Table 2.10 explores the potential impact of different eligibility criteria on average effects. Given the imprecision of most of the individual estimates, these estimates are also quite imprecise and should be interpreted with caution. The first two rows show the estimated effects using the actual Plus 2012 and 2013 samples. The average effect is nearly twice as large among the 2013 cohort, -8.6, than among the 2012 cohort, -4.4, but the standard errors are several times larger than either of the estimates. Next, we see that if all of the individuals who would be expected to have reduction in arrests for violent crime were offered the program, the average effect would increase to -13.9 fewer violent crime arrests per 100 treated youth. The next two rows demonstrate that, in line with our pre-specified hypothesis, in school youths' violent crime arrests are nearly twice as responsive as those of out of school youth with a reduction of 9.3 arrests per 100 in school youth compared to a reduction of 4 arrests per 100 out of school youth. Further narrowing eligibility to the subgroups who seem most responsive based on the above analysis, criminally-involved males with poor school attendance in particular, further increases the program's reduction in arrests for violent crime.

2.10 Conclusion

This paper uses medium-term results from two randomized controlled trials of Chicago's One Summer Plus summer jobs program to better understand why summer jobs reduce violence and for whom such programs work best. We see almost identical decreases in arrests for violent crime- equal to 5 arrests per 100 youth- in both the initial and follow up study, but we find no effect on other types of crime. We find limited effects on schooling, including a marginally significant reduction in graduation two years after the program. We find an increase in post-program formal employment with the program providers, but not among

Table 2.10: Average Effects Under Different Eligibility Criteria

Hypothetical Eligibility Criteria	Average Effect
Plus 2012	-4.354 (23.896)
Plus 2013	-8.602 (32.012)
CATE<0	-13.862 (34.906)
In School Only	-9.346 (33.850)
Out of School Only	-4.094 (21.421)
In School Males with a Prior Arrest	-19.231 (46.625)
In School Males with a Prior Arrest and <100 Days of Attendance	-24.264 (55.542)
In School Males with a Prior Arrest or Female Freshmen with <=1 Arrest	-18.065 (44.883)

Notes. Estimates based on individual CATE estimates. The sample is restricted to the causal forest prediction sample. Standard errors calculated under assumption that individual estimates are independent.

other employers.

One hypothesis for the program's success is that it targets many youth prior to school exit, acting as unemployment prevention rather than remediation. We use new methods from the supervised machine learning literature to explore which subgroups' violent crime rates are most responsive to the intervention. We find that criminally involved males who are still enrolled in school but have poor attendance and minimally criminally involved females early in their high school career benefit most. We find that more disconnected males who are older, more criminally involved and do not have recent formal work experience are most adversely affected by the program. As this group is often the target of similar social programs, this has potentially important implications about optimally targeting similar "light touch" social programs.

CHAPTER 3

DESIGNING ORGANIZATIONAL VERSUS PUBLIC MARKETS (W/ B. PABLO MONTAGNES)

3.1 Introduction

In this paper, we show that matching problems within an organization¹ are distinct from traditional applications in public markets. Our main finding is that assignment mechanisms which are responsive to market participants' preferences do not always best serve the organization's objective, even in situations that seem favorable to a decentralized approach. In some cases, an organization can better achieve its objectives by ignoring preferences and randomly choosing assignments, even when market participants have preferences aligned with the organization's preferences, have outside options, and have private information about match qualities.

Many markets are composed of autonomous agents who are free to contract or match outside of a centralized mechanism. We refer to such markets as *public markets*. As a result, market designers almost always restrict attention to mechanisms which determine matches using agents' preferences, and often they restrict attention to mechanisms which yield pairwise stable matches² in the sense that no unmatched pair of agents would prefer to be matched together over their assigned match. In their seminal paper documenting the redesign of the National Residency Matching Program, [64] explain: "Perhaps the most important and least controversial empirical finding about centralized matching algorithms is that they are most often successful if the matchings they produce are stable" (p. 752). [42] consider a situation where the government can prohibit blocking pairs but argue that "an assignment that completely ignores participants' preferences would be undesirable" (p. 77) and argue that stability is justified for normative reasons.

1. We refer to an "organization" instead of a "firm or organization" for ease of exposition.

2. Following [47] we refer to such mechanisms as "stable mechanisms".

The success of market design in public markets has led to greater interest in the use of economic design to solve a diverse set of organizational problems. The International Monetary Fund uses the deferred acceptance algorithm [30] to assign new economists to research teams³ [12], Teach For America uses a variant of the deferred acceptance algorithm to assign teachers to schools in some regions [25], and the United States Military Academy uses a cumulative offer mechanism to assign cadets to branches [73, 72]. In all of these applications, matches are determined entirely by agents' preferences with no consideration of the optimal allocation from the organization's perspective.

Individuals choose to join an organization and may have to pay adjustment costs to switch to a new organization (for example, searching for and changing to a new job). As a result, organizations may have more flexibility in choosing an assignment mechanism than clearinghouses in public markets. An extreme case is the military which can court martial cadets for insubordination. In general, organizations do not have complete authority over their members and cannot prevent their members from exiting the organization, but they can forbid or sanction participants in internal two-sided matching markets from matching outside of the centralized assignment process. More concretely, organizational market design problems are constrained by individual rationality (IR) constraints. In many real world applications, organizations enforce an assignment without any input from the agents. For example, employees are often assigned to a supervisor or a group within a company as the company sees fit.⁴

The appeal of adopting the insights of market design within organizations has merit. Organizations may find it desirable to determine matches using a decentralized mechanism in order to bring some of the benefits of markets in to the organization. One might anticipate advantages to this approach when:

3. [12] propose the use of deferred acceptance. We confirmed that it was implemented with Vardy on December 1, 2014.

4. [65] motivate their exploration of the relationship between matching and assignment problems with the example of matching workers to supervisors.

1. agents and organizations value similar match features;
2. agents possess more information than the organization; and
3. agents have outside options.

When agents and organizations value similar match features there appears to be little tension between the organization's goals and its members' preferences. Dispersed private information is a classic argument in favor of decentralization [35], with the logic being that the decision maker who best understands the benefits and costs of a decision will be best able to optimally make the decision. A feature of decentralized mechanisms is their ability to aggregate this dispersed information. When agents have stronger outside options organizations must keep them happier in order to compete for their ongoing membership, perhaps by providing them with more autonomy over their matches. We show that none of the three conditions above, alone or in combination, is guaranteed to improve the performance of preference respecting mechanisms. In some cases, they make decentralized outcomes worse.

The rest of this paper is organized as follows. Section 2 presents a model of organizational market design problems. Section 3 presents our results. Section 4 concludes.

3.2 Model

We consider a variant of a two-sided assignment game [71, 66]. Call one side of the market "teachers" and the other side "schools". Suppose there are M teachers and N schools to be matched in a market without prices or transfers. Each of the possible $M \times N$ matches is associated with an indivisible output:

$$\alpha_{ij} = f(i, j), \tag{3.1}$$

where $f(\cdot)$ is an arbitrary production function, i indexes teachers and j indexes schools.⁵ This yields the following potential output matrix:

$$A = \begin{bmatrix} \alpha_{11} & \dots & \alpha_{1N} \\ \dots & & \dots \\ \alpha_{M1} & \dots & \alpha_{MN} \end{bmatrix}.$$

Teachers and schools get the following utility from each potential match, respectively:

$$u_{ij} = \alpha_{ij} + \mu_{ij} - \underline{u}_i, \quad (3.3)$$

$$v_{ij} = \alpha_{ij} + \nu_{ij}, \quad (3.4)$$

where \underline{u}_i captures a teacher specific outside option. We assume that a match is acceptable to a teacher or school if $u_{ij} \geq 0$ or $v_{ij} \geq 0$, respectively. Teacher preferences can be written in matrix form as $U - \underline{U}$, where U represents their preferences over matches and \underline{U} captures each teacher's outside option. Similarly, school preferences can be denoted V . Teachers and schools observe each component of their utility, including the actual output of the match, α_{ij} .

An *Organizational Market Design Problem* is a two-sided assignment game without prices or transfers with two distinguishing features. First, a risk neutral organization provides a technology that makes matches productive. We assume that matches are unproductive without organizational support. Second, the organization supporting the match is invested in the resulting assignment. Throughout this paper we assume the organization's objective is to assign teachers to schools in order to maximize total output:

5. [40] studied teacher-school match qualities empirically by modeling $f()$ with a variant of the following Cobb-Douglas production function:

$$\alpha_{ij} = \gamma + \delta_i + \beta_j + \eta_{ij}, \quad (3.2)$$

where δ_i is a teacher-specific contribution to the match output, β_j is a school specific contribution to the match output, γ represents the organization's contribution the match output, and η_{ij} is a match specific component of productivity.

$$\begin{aligned}
& \max_{x_{ij}} \sum_{j=1}^N \sum_{i=1}^M \alpha_{ij} x_{ij}, \\
\text{subject to } & \sum_{j=1}^N x_{ij} \leq 1 \text{ for } i = 1, \dots, M, \\
& \sum_{i=1}^M x_{ij} \leq 1 \text{ for } j = 1, \dots, N,
\end{aligned} \tag{3.5}$$

$$x_{ij} \in \{0, 1\}, \tag{3.6}$$

$$u_{ij} x_{ij} \geq 0, \tag{3.7}$$

$$v_{ij} x_{ij} \geq 0 \text{ for } i = 1, \dots, M, j = 1, \dots, N.$$

In general, organizations may have different objective functions. For example, a pharmaceutical company trying to match scientists to labs may care about the maximum match quality rather than the sum of all match productivities since the best match is most likely to result in a scientific breakthrough.⁶

Organizations have imperfect information about match qualities. Prior to the match, the organization observes \hat{A}_{ij}^k , where for each i and j :

$$\hat{a}_{ij} = \alpha_{ij} + \zeta_{ij}, \tag{3.8}$$

where ζ_{ij} is a random variable assumed to be independent of α_{ij} . Organizations and market participants have asymmetric information whenever $\text{Var}(\zeta_{ij}) \neq 0$.

The organization must decide whether to implement a match based on its noisy information about the match outputs, \hat{A}_{ij} or to delegate the assignment decision to the market participants. For example, the organization could delegate assignments using the deferred ac-

6. [51] present an example where optimal assignments within a firm are determined by the absolute advantage of workers. Since stability is determined by absolute advantage, stable mechanisms are likely to be optimal in their example.

ceptance algorithm [30] or with other mechanisms that will not always yield stable matches⁷, like Top Trading Cycles [70, 3] or a random serial dictatorship for one side of the market. We say that an assignment mechanism *respects agents' preferences* whenever two agents who most prefer each other are always matched by the mechanism. Respect for agents' preferences is a weaker notion than stability. All stable mechanisms respect agents' preferences since the pair who most prefer each other are a blocking pair if not matched together. However, other non-stable mechanisms, like top trading cycles, also respect agents' preferences.

In general, organizations can assign some matches and delegate others. The organization's problem amounts to selecting a matching algorithm and a constraint matrix, C :⁸

$$C = \begin{bmatrix} c_{11} & \dots & c_{1N} \\ \dots & & \dots \\ c_{M1} & \dots & c_{mn} \end{bmatrix},$$

where, for all i and j , c_{ij} is 1 if the organization allows the match and 0 otherwise. The constrained assignment problem is then delegated to market participants. Teacher and school preferences over the constrained set are given by $U \circ C - \underline{U}$ and $V \circ C$, respectively, where \circ denotes the Hadamard Product.⁹

3.3 Results

We begin our analysis by showing that in general an organization will not want to delegate the assignment decision to market participants using an assignment mechanism that respects agents' preferences when the organization has perfect information about the match productivity matrix, A , and teacher and school preferences. With perfect information, the

7. A match is stable if there is no unmatched school-teacher pair where both the school and the teacher would prefer to be matched together over their assigned match.

8. There are $2^{N \times M}$ possible constraint matrices. Consequently, it becomes quite difficult to determine the optimal constraint matrix as the size of the market grows. In other work, we investigate the optimal choice of constraint matrix.

9. The Hadamard Product multiplies matrices component by component.

organization’s problem is a classical assignment problem as specified by [50]. An optimal assignment exists and can be solved relatively easily using well-known solution methods, such as the Primal-Dual Transportation Algorithm. Therefore, with complete knowledge about A and participation constraints, the organization can do no better than to implement the optimal assignment without input from the teachers and schools. However, some organizations may prefer to give their members autonomy over matches if the resulting match is still optimal. Result 1 shows that in general assignment mechanisms that respect agents’ preferences will not yield the optimal assignment.

Result 1. There exists no assignment mechanism that respects agents’ preferences that always selects an organization’s optimal assignment.

We demonstrate Result 1 using a counterexample.

Example 3. [49] (LLW) show that minority and poor students are frequently taught by the least skilled teachers. This could be particularly troubling from a social welfare perspective if disadvantaged students, who likely have less support and fewer resources at home on average, benefit most from having a high quality teacher. LLW conclude that any salary differentials across urban and suburban schools are insufficient to compensate teachers for the difficulties associated with teaching in an urban school.

In this example, we present a simple model of the scenario LLW describe. Suppose the market consists of one good and one bad teacher, one wealthy and one poor school, and a single organization supporting matches. Moreover, suppose teaching at the wealthy school is easier because wealthy students have more enriching home environments. Consequently, both the good and bad teacher are more productive at the wealthy school, but the bad teacher’s comparative advantage at the wealthy school is greater since students at the wealthy school are generally successful. The outputs associated with the four possible matches are given by:

$$\begin{array}{rcc}
& & \textit{Wealthy School} & \textit{Poor School} \\
A = & \textit{Good Teacher} & \alpha & 1 \\
& \textit{Bad Teacher} & 1 & 0
\end{array} \quad ,$$

where $1 < \alpha < 2$. The organization's optimal assignment is:

$$\{(\textit{Good Teacher}, \textit{Poor School}), (\textit{Bad Teacher}, \textit{Wealthy School})\},$$

which yields a total output of 2.

Suppose schools only value the productivity of the matches, but teacher preferences are given by:

$$\begin{array}{rcc}
& & \textit{Wealthy School} & \textit{Poor School} \\
U = & \textit{Good Teacher} & \alpha & 1 \pm \zeta_{12} \\
& \textit{Bad Teacher} & 1 & 0
\end{array} \quad ,$$

where α is the same as in A and ζ_{12} equals 1 half the time and -1 the other half of the time. As an example, ζ_{12} could represent how happy the good teacher was when she arrived at her interview with the poor school.

If the organization delegates the decision to the market participants using a mechanism which respects agents' preferences, the mechanism will select the optimal assignment when $\zeta_{12} = 1$, since the Good Teacher and Poor School most prefer each other. This assignment yields a total output of 2. When $\zeta_{12} = -1$, the mechanism selects the following assignment:

$$\{(\textit{Good Teacher}, \textit{Wealthy School}), (\textit{Bad Teacher}, \textit{Poor School})\},$$

since the Good Teacher and Wealthy School most prefer each other. This assignment yields a total output of α . Therefore, the expected output of the delegated assignment is $\frac{1}{2}2 + \frac{1}{2}\alpha$ which is less than the output of the optimal assignment.

The key insight underlying Result 1 and Example 3 is that ordinal preferences are determined by teachers’ and schools’ absolute advantage in matches, rather than their comparative advantage. Consequently, a match that respects agents’ preferences will only be optimal when absolute advantage and comparative advantage coincide. Choosing according to absolute instead of comparative advantage ignores the effect of a match on the productivity of other matches and therefore imposes a *displacement externality* on the organization and the other members of the organization. Importantly, a match must be supported by an organization in order to be productive, so an organization is free to choose any assignment that is acceptable to its members.

3.3.1 *Aligned Preferences*

In the introduction, we hypothesized that an organization may find it beneficial to delegate the assignment when market participants value the same match features as the organization, the organization has imperfect information, and market participants have outside options. Result 2 shows that the first of these conditions can actually make things worse for an organization’s bottom line.

Result 2. Assignment mechanisms which respect agents’ preferences may yield worse outcomes for an organization as market participants’ preferences become more aligned with an organization’s objectives.

We demonstrate Result 2 using another counterexample. Example 4 shows that assignment mechanisms that respect agents’ preferences will not always select the organization’s optimal assignment even in environments where the market participants’ preferences are highly aligned with the organizations. We formalize what we mean by highly aligned preferences with the following definition.

Definition 1. Teachers and schools are said to have preferences **aligned**¹⁰ with the orga-

10. This is a special case of the definition used by [57] where the ordinal potential is given by the productivity matrix, A .

nization if their preferences over matches are determined entirely by the productivity of the matches, or more precisely:

$$u_{ij} = \alpha_{ij} - u_i,$$

$$v_{ij} = \alpha_{ij}.$$

Assuming that agents have this type of aligned preferences should be favorable to a decentralized approach. Without aligned preferences, one could easily construct examples where a decentralized approach does terribly because agents want to minimize the organization's productivity. Combined with our assumption that the organization would like to maximize the sum of all match productivities, assuming market participants have aligned preferences is equivalent to assuming that the organization would like to select the utilitarian optimal match for its members.

The next example demonstrates where assignments from preference respecting mechanisms can go wrong from the organization's perspective when preferences are aligned.

Example 4. *Consider again the setup of Example 3. Assume preferences are aligned, so teachers and schools only value the productivity of the matches. Also, assume the utility from all agents' outside options is 0. The organization's optimal assignment is still:*

$$\{(Good\ Teacher, Poor\ School), (Bad\ Teacher, Wealthy\ School)\},$$

which yields a total output of 2. If the organization delegates the decision to the market participants using a mechanism which respects agents' preferences, the assignment will be:

$$\{(Good\ Teacher, Wealthy\ School), (Bad\ Teacher, Poor\ School)\},$$

since the good teacher and wealthy school most prefer each other. This assignment yields a total output of α where $\alpha < 2$. By comparison, when the wealthy teacher also values how

happy she is at her interview with the poor school as in Example 3, the expected output of the delegated assignment is $\frac{1}{2}2 + \frac{1}{2}\alpha$. Therefore, the expected output of a delegated mechanism is actually lower when teachers are assumed to value the exact same match features as the organization.

Example 4 demonstrates that mechanisms that respect agents' preferences will not always yield better matches from an organization's perspective when market participants value something similar to the organization. In fact, these mechanisms may yield worse matches from the organization's perspective when preferences are aligned.

Ordinal preferences are determined by teachers' and schools' absolute advantage in matches. When preferences are aligned but absolute and comparative advantage are maximized by different matches, the match from a mechanism that respects agents' preferences cannot be optimal. However, if preferences are not aligned, as in Example 3, the noise in teacher preferences can be beneficial to the organization if it causes the perceived absolute advantage to coincide with comparative advantage.

3.3.2 Imperfect Information

Examples 3 and 4 were extreme because they made the unrealistic assumption that the organization has perfect information about match productivities. In this section, we consider another, perhaps more realistic, extreme case, by assuming the organization has no information about match qualities. The organization can either impose a randomly chosen assignment or capitalize on the private information of the market participants by delegating the match with an assignment mechanism that respects preferences at the cost of ruling out organizationally stable matches that may conflict with agents' preferences. Result 3 shows that preference respecting mechanisms do not always dominate random assignment.

Result 3. Random assignment can dominate an assignment mechanism that respects agents' preferences from a risk-neutral organization's perspective even if preferences are aligned.

We demonstrate Result 3 using a counterexample. Example 5 shows that delegated matches may not perform well even when the organization has no information about the matches and market participants value the same match features as the organization.

Example 5. *Consider the setup of Example 4. Assume preferences are aligned and that the utility from all agents' outside options is 0. Suppose the market consists of a good and bad teacher, a wealthy and poor school, and a single organization supporting matches. The true outputs associated with the four possible matches are identical to the outputs in Example 4.*

If the organization delegates the decision to the market participants using a mechanism that respects agents' preferences, the assignment is:

$$\{(Good\ Teacher, Wealthy\ School), (Bad\ Teacher, Poor\ School)\},$$

which yields a total output of α . If instead, the organization randomly selects the assignment, the assignment will be:

$$\{(Good\ Teacher, Wealthy\ School), (Bad\ Teacher, Poor\ School)\},$$

with a total output of α half the time and:

$$\{(Good\ Teacher, Poor\ School), (Bad\ Teacher, Wealthy\ School)\},$$

with a total output of 2 the other half of the time. The expected match quality is therefore $1 + \frac{\alpha}{2}$ which is greater than α because $\alpha < 2$. The random assignment mechanism dominates the delegated mechanism.

Example 5 demonstrates that delegating the assignment may not be optimal even in situations with highly dispersed information and aligned preferences.

This result is driven by two features of Example 5. Like Example 4, absolute and comparative advantage are maximized by different matches, so the preference respecting match

is not optimal. Since there are only two possible matches that do not leave either teacher or school unmatched, the preference respecting match is the worst possible match. As a result, the delegated assignment mechanism is dominated in expectation by any mechanism that does not always respect preferences, including random assignment.

3.3.3 *Outside Options*

Another argument for delegating assignments is individual rationality (IR) constraints. Intuition suggests that the benefit of giving market participants more input in assignments should be greater when organizations must compete for membership. Result 4 shows that this is not always the case. In some cases, delegated assignments perform worse as outside options improve.

Result 4. Preference respecting mechanisms may perform worse when outside options improve.

We demonstrate Result 4 using another counterexample. Unlike our previous examples, we construct Example 6 so that absolute and comparative advantage coincide when there are no viable outside options. However, when outside options improve, matches with relatively low absolute advantage become unacceptable. Only considering absolute rather than comparative advantage ignores the effect of a match on other match productivities. This imposes a displacement externality on the organization and induces the teacher with the lower match quality to leave the organization. In short, preference respecting mechanisms may cater too much to some members at the expense of other members.

Example 6. *Consider a similar setup to Example 4. Again, assume preferences are aligned and that the utility from all agents' outside options is 0. Suppose the market includes a good teacher and a bad teacher, a wealthy school and a poor school, and a single organization supporting matches. The outputs associated with the four possible matches are given by:*

	<i>Wealthy School</i>	<i>Poor School</i>
$A =$	<i>Good Teacher</i>	1.5
	<i>Bad Teacher</i>	1
		0.75.

If the organization delegates the decision to the market participants using a preference respecting mechanism, the unique assignment is:

$$\{(Good\ Teacher, Wealthy\ School), (Bad\ Teacher, Poor\ School)\},$$

which yields a total output of 2.25. This is also the optimal assignment.

Now, suppose the utility from both teachers' outside options improves from 0 to 1. The bad teacher no longer finds being matched to the poor school acceptable and will quit, so the assignment's output falls to 1.5. If instead, the organization randomly selects the assignment, the assignment will be:

$$\{(Good\ Teacher, Wealthy\ School), (Bad\ Teacher, Poor\ School)\},$$

with a total output of 1.5 half the time and:

$$\{(Good\ Teacher, Poor\ School), (Bad\ Teacher, Wealthy\ School)\},$$

with a total output of 2 the other half of the time. The random assignment mechanism dominates the delegated mechanism with the improved outside options.

While Result 4 shows that better outside options are not necessarily favorable to a decentralized approach from an organization's perspective, this is in part due to our assumption of a fixed pool of potential members. To see the limitations of this assumption, consider the assignment of airline seat upgrades [48]. This could be considered a type of organizational market design problem since ticketed passengers on a flight have already purchased their

tickets for the flight and face a large cost of switching airlines for that flight. Consequently, the airline is constrained by IR constraints when assigning seat upgrades for that flight and does not necessarily need to assign upgrades in accordance with passengers' preferences. However, the airline would be naive to ignore customer satisfaction because the passengers can easily switch to a new airline for their next flight. Moreover, future frequent travelers may avoid that airline if they know they are more likely to receive upgrades from a competing airline.

3.3.4 More General Solutions

The previous examples show that preference respecting assignment mechanisms do not always yield assignments that are desirable to an organization, even under conditions that seem quite favorable to a decentralized approach. More generally, organizations are not limited to a choice between choosing all assignments or delegating all assignments using a preference respecting mechanism. Instead, organizations can determine assignments using a decentralized, but constrained, solution, where matches are determined using a preference respecting mechanism, but market participants have constrained choice sets. The following example demonstrates how this approach may be optimal for an organization.

Example 7. *Suppose now that the market includes a single good teacher, two bad teachers, two wealthy schools, and one poor school. The true output matrix is given by:*

$$A = \begin{array}{rcccl} & & \textit{Wealthy School 1} & \textit{Wealthy School 2} & \textit{Poor School 3} \\ \textit{Good Teacher 1} & & 12 + \eta_{11} & 12 + \eta_{12} & 9 + \eta_{13} \\ \textit{Bad Teacher 2} & & 9 + \eta_{21} & 9 + \eta_{22} & 3 + \eta_{23} \\ \textit{Bad Teacher 3} & & 9 + \eta_{31} & 9 + \eta_{32} & 3 + \eta_{33} \end{array},$$

where the η_{ij} are independent and identically distributed with $P(\eta_{ij} = 0.5) = \frac{1}{2}$ and $P(\eta_{ij} = -0.5) = \frac{1}{2}$. The organization knows the mean of each match quality and the

distribution of η_{ij} , but does not observe the η_{ij} . All expected outputs mentioned below are derived in Appendix 1.

If the organization delegates the assignment, the assignment will match the good teacher and one of the bad teachers with the wealthy schools and the remaining bad teacher with a poor school. In expectation, this will result in an aggregate output of 24.5. In contrast, the expected output of a randomly selected assignment is 25. Again, the random assignment mechanism outperforms the delegated mechanism.

However, the organization can do better than random assignment since it knows the expected productivity of each match. In particular, notice that the good teacher's comparative advantage is always greatest at the poor school.¹¹ Since the optimal assignment will always place teachers in the position where their comparative advantage is greatest, the organization should optimally assign the good teacher to the poor school.

After assigning the good teacher to the poor school, the organization must decide whether to randomly assign matches between the bad teachers and wealthy schools or to delegate these assignment decisions. The expected output from randomly choosing how bad teachers and wealthy schools are matched is 27, whereas the expected output of the delegated assignment is $27\frac{5}{16}$. The delegated assignment nearly achieves the optimal match's expected output of $27\frac{3}{8}$.

The organization's optimal solution to this Organizational Market Design Problem is to select the constraint matrix, C^* :

$$C^* = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix},$$

and allow the market participants to select matches from their constrained choice sets using a delegated mechanism.

11. The maximum comparative advantage the good teacher can have at the wealthy school is $12.5 - 8.5 = 4$, whereas the minimum comparative advantage she can have at the poor school is $8.5 - 3.5 = 5$.

Our previous examples demonstrated that preference respecting assignments may be suboptimal because choosing assignments according to absolute advantage rather than comparative advantage imposes a displacement externality on the organization and the other members of the organization. In some cases, an organization may recognize that comparative advantage is maximized for a subset of its members by a particular match. In these cases, the organization can require the aforementioned subset of its members to match but allow its other members to match using a decentralized assignment mechanism. Example 7 demonstrates that while delegated assignments may perform worse than a completely random assignment in these cases, a constrained version of the delegated assignment may still be optimal.

3.4 Conclusion

Since the successful redesign of the National Residency Matching Program [64] the tools of market design have increasingly been adopted as solutions to real world matching problems in situations without prices. Recent applications include the use of ordinal assignment mechanisms to solve matching problems within organizations [12, 73, 72]. In this paper, we show that matching problems within an organization are distinct from traditional applications in public markets. In particular, organizations have flexibility in choosing an assignment mechanism that best achieves its objectives and may benefit from disregarding agents' preferences when picking assignments.

Our main finding is that assignments resulting from mechanisms that respect agents' preferences, like the deferred acceptance algorithm and top trading cycles, do not always best serve the organization's objective, even in situations that seem favorable to the decentralized approach. We illustrate examples where an organization is better off randomly choosing assignments even when market participants value the same types of things as the organization, have more information about match qualities than the organization, and have increasingly viable outside options.

REFERENCES

- [1] The Cost of Teacher Turnover. Technical report, Texas State Board for Educator Certification, 2000.
- [2] Atila Abdulkadiroglu, Parag A. Pathak, Alvin E. Roth, and Tayfun Sonmez. The Boston Public School Match. *American Economic Review*, 95(2):368–371, 2005.
- [3] Atila Abdulkadiroglu and Tayfun Sonmez. House Allocation with Existing Tenants. *Journal of Economic Theory*, 88:233–260, 1999.
- [4] Nikhil Agarwal. An Empirical Model of the Medical Match. *American Economic Review*, 105(7):1939–1978, 2015.
- [5] Nikhil Agarwal and William Diamond. Identification and Estimation in Two-Sided Matching Markets. 2014.
- [6] Sandro Ambuehl and Vivienne Groves. Unraveling Over Time. 2015.
- [7] Center for Disease Control and Prevention. Web-based Injury Statistics Query and Reporting System (WISQARS), 2014.
- [8] Michael L. Anderson. Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484):1481–1495, 2008.
- [9] Susan Athey and Guido W Imbens. Machine Learning Methods for Estimating Heterogeneous Causal Effects. 2015.
- [10] Christopher Avery, Christine Jolls, Richard a. Posner, and Alvin E. Roth. The Market for Federal Judicial Law Clerks. *The University of Chicago Law Review*, 68:793–902, 2001.
- [11] Christopher Avery, Christine Jolls, Richard A. Posner, and Alvin E. Roth. The New Market for Federal Judicial Law Clerks. *The University of Chicago Law Review*, 447:447–486, 2007.
- [12] Greg Barron and Felix Várdy. The Internal Job Market of the IMF’s Economist Program. 2004.
- [13] Robert M Bell and Daniel F McCaffrey. Bias Reduction in Standard Errors for Linear Regression with Multi-Stage Samples. *Survey Methodology*, 28(2):169–181, 2002.
- [14] Yoav Benjamini and Yosef Hochberg. Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1):289–300, 1995.
- [15] Yoav Benjamini, Abba M. Krieger, and Daniel Yekutieli. Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491–507, 2006.

- [16] Howard S Bloom, Larry L. Orr, Stephen H. Bell, George Cave, Fred Doolittle, Winston Lin, and Johannes M. Bos. The Benefits and Costs of JTPA Title II-A Programs: Key Findings from the National Job Training Partnership Act Study. *The Journal of Human Resources*, 32(3):549–576, 1997.
- [17] Donald Boyd, Hamilton Lankford, Susanna Loeb, and James Wyckoff. Analyzing the Determinants of the Matching of Public School Teachers to Jobs : Disentangling the Preferences of Teachers and Employers. *Journal of Labor Economics*, 31(1):83–117, 2013.
- [18] A. Colin Cameron, Jonah B. Gelbach, and Douglas L. Miller. Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3):414–427, 2008.
- [19] George Cave, Hans Bos, Fred Doolittle, and Cyril Toussaint. JOBSTART: Final Report on a Program for School Dropouts. Technical Report October, MDRC, 1993.
- [20] Center for Labor Market Studies and Northeastern University. The Depression in the Teen Labor Market in Illinois in Recent Year. Technical Report January, 2012.
- [21] Raj Chetty, John N Friedman, and Jonah E Rockoff. The long-term impacts of teachers: Teacher value-added and student outcomes in adulthood. *American Economic Review*, 104(9):2633–2679, 2014.
- [22] Mark A. Cohen and Alex R. Piquero. New Evidence on the Monetary Value of Saving a High Risk Youth. *Journal of Quantitative Criminology*, 25:25–49, 2009.
- [23] Julien Combe, Olivier Tercieux, and Camille Terrier. The Design of Teacher Assignment : Theory and Evidence. 2015.
- [24] Ettore Damiano, Wing Suen, and Hao Li. Unravelling of Dynamic Sorting. *Review of Economic Studies*, 72(4):1057–1076, 2005.
- [25] Jonathan M.V. Davis. Do Assignment Mechanisms Matter? Quasi-Experimental Evidence from a Teach for America Pilot. 2015.
- [26] Umut Dur and Onur Kesten. Sequential Assignment Problems and Two Applications. 2013.
- [27] Clayton R Featherstone. A rank-based refinement of ordinal efficiency and a new (but familiar) class of ordinal assignment mechanisms. pages 1–68, 2011.
- [28] Guillaume R Fréchette, Alvin E Roth, and M Utku Ünver. Unraveling yields inefficient matchings: evidence from post-season college football bowls. *RAND Journal of Economics*, 38(4):967–982, 2007.
- [29] Milton Friedman and Rose Friedman. *Free to Choose: A Personal Statement*. Harcourt, Inc., Orlando, FL, 1990.

- [30] David Gale and Lloyd S . Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.
- [31] Alexander Gelber, Adam Isen, and Judd B. Kessler. The Effects of Youth Employment: Evidence From New York City Lotteries. *The Quarterly Journal of Economics*, 131:423–460, 2015.
- [32] Steven Glazerman, Daniel Mayer, and Paul Decker. Alternative Routes to Teaching: The Impacts of Teach for America on Student Achievement and Other Outcomes. *Journal of Policy Analysis and Management*, 25(1):75–96, 2006.
- [33] Alice Goffman. *On The Run: Fugitive Life in an American City*. Macmillan, 2015.
- [34] Hanna Halaburda. Unravelling in Two-Sided Matching Markets and Similarity of Preferences. *Games and Economic Behavior*, 69(2):365–393, 2010.
- [35] F.A. Hayek. The Use of Knowledge in Society. *The American Economic Review*, 35(4):519–530, 1945.
- [36] James J Heckman. Investing in Disadvantaged Children. *Science*, 312:1900–1902, 2006.
- [37] Sara B Heller. Summer jobs reduce violence among disadvantaged youth. *Science*, 346(6214):1219–1223, 2014.
- [38] Sara B. Heller, Anuj K. Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold A. Pollack. Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago. 2015.
- [39] Jennifer Imazeki. Teacher salaries and teacher attrition. *Economics of Education Review*, 24(4):431–449, 2005.
- [40] C. Kirabo Jackson. Match Quality, Worker Productivity, and Worker Mobility: Direct Evidence from Teachers. *The Review of Economics and Statistics*, 95(4):1096–1116, 2013.
- [41] Louis A. Jaeckel. The Infinitesimal Jackknife. Technical report, 1972.
- [42] Yuichiro Kamada and Fuhito Kojima. Efficient Matching Under Distributional Concerns: Theory and Applications. *American Economic Review*, 105(1):67–99, 2015.
- [43] Thomas J Kane and Douglas O Staiger. Using Imperfect Information to Identify Effective Teachers. 2005.
- [44] Lawrence F. Katz, Jeffrey R. Kling, and Jeffrey B. Liebman. Moving to opportunity in boston: early results of a randomized mobility experiment* 1. *The Quarterly Journal of Economics*, 116(May):607–654, 2001.
- [45] James J Kemple. Career Academies: Long-Term Impacts on Labor Market Outcomes, Educational Attainment, and Transitions to Adulthood. Technical report, MDRC, 2008.

- [46] Christopher T King and Carolyn J Heinrich. How Effective Are Workforce Development Programs? Implications for U.S. Workforce Policies. Technical Report November, 2011.
- [47] Fuhito Kojima. School choice : Impossibilities for affirmative action. *Games and Economic Behavior*, 75:685–693, 2012.
- [48] Scott Duke Kominers and Tayfun Sonmez. Matching with Slot-Specific Priorities : Theory. 2014.
- [49] Hamilton Lankford, Susanna Loeb, and James Wyckoff. Teacher Sorting and the Plight of Urban Schools : A Descriptive Analysis. *Educational Evaluation and Policy Analysis*, 24(1):37–62, 2002.
- [50] David G. Luenberger. *Linear and Nonlinear Programming: Second Edition*. Springer, New York, 2003.
- [51] Glenn M. MacDonald and James R. Markusen. A Rehabilitation of Absolute Advantage. *Journal of Political Economy*, 93(2):277–297, 1985.
- [52] Sheena McConnell and Steven Glazerman. National Job Corps Study: The Benefits and Costs of Job Corps. Technical report, Mathematica Policy Research, Inc., 2001.
- [53] Megan Millenky, Dan Bloom, Sara Muller-Ravett, and Joseph Broadus. Staying on Course: Three-Year Results of the National Guard Youth ChalleNGe Evaluation. Technical Report June, MDRC, 2011.
- [54] Ted R. Miller, Mark A. Cohen, and Brian Wiersema. Victim Costs and Consequences: A New Look. Technical report, U.S. Department of Justice, Office of Justice Programs, National Institute of Justice, 1996.
- [55] Muriel Niederle and Alvin E. Roth. Relationship Between Wages and Presence of a Match in Medical Fellowships. *Journal of the American Medical Association*, 290(9):1153–1154, 2003.
- [56] Muriel Niederle and Alvin E. Roth. Unraveling Reduces Mobility in a Labor Market: Gastroenterology with and without a Centralized Match. *Journal of Political Economy*, 111(6):1342–1352, 2003.
- [57] Muriel Niederle and Leeat Yariv. Decentralized Matching with Aligned Preferences. 2009.
- [58] Parag A. Pathak and Tayfun Sönmez. Leveling the playing field: Sincere and sophisticated players in the Boston mechanism. *American Economic Review*, 98(4):1636–1652, 2008.
- [59] Juan Sebastián Pereyra. A dynamic school choice model. *Games and Economic Behavior*, 80:100–114, 2013.

- [60] Francisco Perez-Arce, Louay Constant, David S. Loughran, and Lynn A. Karoly. A Cost-Benefit Analysis of the National Guard Youth ChalleNGe Program. Technical report, RAND Corporation, 2012.
- [61] Alvin E Roth. A Natural Experiment in the Organization of Entry-Level Labor Markets : Regional Markets for New Physicians and Surgeons in the United Kingdom. *The American Economic Review*, 81(3):415–440, 1991.
- [62] Alvin E. Roth. Deferred acceptance algorithms: history, theory, practice, and open questions. *International Journal of Game Theory*, 36(3-4):537–569, jan 2008.
- [63] Alvin E Roth and Elliott Peranson. The Redesign of the Matching Market for American Physicians : Some Engineering Aspects of Economic Design. *The American Economic Review*, 89(4):748–780, 1999.
- [64] Alvin E Roth and Elliott Peranson. The Redesign of the Matching Market for American Physicians : Some Engineering Aspects of Economic Design. *The American Economic Review*, 89(4):748–780, 1999.
- [65] Alvin E Roth, Uriel G Rothblum, and John H Vande Vate. Stable Matchings, Optimal Assignments, and Linear Programming. *Mathematics of Operations Research*, 18(4):803–828, 1993.
- [66] Alvin E. Roth and Marilda Sotomayor. Two-Sided Matching. In R.J. Aumann and S. Hart, editors, *Handbook of Game Theory*, volume 1, chapter 16, pages 485–541. Elsevier Science Publishers B.V., 1992.
- [67] Maytal Saar-Tsechansky and Foster Provost. Handling Missing Values when Applying Classification Models. *Journal of Machine Learning Research*, 8:1625–1657, 2007.
- [68] Peter Z. Schochet, John Burghardt, and Sheena McConnell. National Job Corps Study and Longer-Term Follow- Up Study: Impact and Benefit-Cost Findings Using Survey and Summary Earnings Records Data. Technical report, Mathematica Policy Research, Inc., 2006.
- [69] Peter Z. Schochet, John Burghardt, and Sheena McConnell. Does Job Corps work? Impact findings from the national Job Corps study. *American Economic Review*, 98(5):1864–1886, 2008.
- [70] Lloyd Shapley and Herbert Scarf. On Cores and Indivisibility. *Journal of Mathematical Economics*, 1:23–37, 1974.
- [71] Lloyd S. Shapley and Martin Shubik. The Assignment Game I: The Core. *International Journal of Game Theory*, 1:111–130, 1972.
- [72] Tayfun Sonmez. Bidding for Army Career Specialties : Improving the ROTC Branching Mechanism. *Journal of Political Economy*, 121(1):186–219, 2013.

- [73] Tayfun Sonmez and Tobias B Switzer. Matching With (Branch-of-Choice) Contracts at the United States Military Academy. *Econometrica*, 81(2):451–488, 2013.
- [74] Douglas O Staiger and Jonah E Rockoff. Searching for Effective Teachers with Imperfect Information. *The Journal of Economic Perspectives*, 24(3):97–117, 2010.
- [75] Bureau of Labor Statistics. Employment and Unemployment Among Youth - Summer 2015. Technical Report August, 2015.
- [76] Christopher Uggen. Work as a Turning Point in the Life Course of Criminals: A Duration Model of Age, Employment, and Recidivism. *American Sociological Review*, 65(4):529–546, 2000.
- [77] Emiliana Vegas, Miguel Urquiola, and Pedro Cerdán-Infantes. Teacher assignment , mobility and their impact on equity and quality of education in Uruguay. 2006.
- [78] Sudhir Alladi Venkatesh. *Off the Books: The Underground Economy of the Urban Poor*. Harvard University Press, 2006.
- [79] Stefan Wager. Asymptotic Theory for Random Forests. 2014.
- [80] Stefan Wager and Susan Athey. Estimation and Inference of Heterogeneous Treatment Effects using Random Forests. 2015.
- [81] Peter H. Westfall and S. Stanley Young. *Resampling-Based Multiple Testing: Examples and Methods for p-Value Adjustment*. Wiley-Interscience, 1993.
- [82] Takirra Winfield. Teach for America Welcomes Most Diverse Talent in 25-Year History, 2014.