UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Essays on Impact Evaluation in Labor and Development Economics

Permalink
https://escholarship.org/uc/item/3zc2h924

Author
Christensen, Garret Smyth

Publication Date
2011

Peer reviewed|Thesis/dissertation

Essays on Impact Evaluation in Labor and Development Economics

by

Garret Smyth Christensen

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Edward Miguel, Chair
Professor David Card
Professor Justin McCrary

Spring 2011

Essays on Impact Evaluation in Labor and Development Economics

Abstract

Essays on Impact Evaluation in Labor and Development Economics

by

Garret Smyth Christensen

Doctor of Philosophy in Economics

University of California, Berkeley

Professor Edward Miguel, Chair

This dissertation studies examples of applied econometrics for causal inference in labor and development economics. One of the fundamental problems in applied fields of economics is causal inference. Merely observing that event B occurred after event A is not enough to claim that A caused B. The field of economics, and the social sciences in general, are limited by ethics and practicality in their ability to conduct randomized field experiments, the gold standard for causality in other fields. Several statistical methods have been devised to obtain causal estimates from "natural" or "quasi" experimental settings—settings where plausibly exogenous variation in a treatment effect of interest can be found and exploited to produce an unbiased estimate of causal effects. Some of these methods include panel data with fixed effects, nearest-neighbor matching, and regression discontinuity. This dissertation explores applications of these econometric methods, as well as an actual randomized controlled trial, in issues of labor and development economics.

The first chapter uses panel data, and causal estimates are identified using a series of fixed effects to control for unmeasurable characteristics that could be correlated with both dependent and independent variables. The subject matter is the recruiting task of the United States military, which is the largest employer in the nation and spends over $4 billion each year to recruit roughly 200,000 new soldiers to maintain its troop levels. This recruiting task has become more expensive since the beginning of the wars in Iraq and Afghanistan. I use a detailed new dataset of all US military applicants over several recent years and find that deaths in Iraq of US soldiers had a significant deterrent effect on recruiting in the home county of the soldiers who were killed. The deterrent effect of local deaths is significantly larger than the deterrent from a death from outside the county. The deterrent exhibits significant heterogeneity across characteristics of deaths, recruits, and locations. Deaths from Iraq decrease recruiting, while deaths from Afghanistan actually increase recruiting. Recruits with higher test scores are more deterred by deaths, and the deterrent is larger and more negative in less populous and more racially diverse counties, but is significantly smaller and in many cases even positive in counties that voted for George W. Bush in the 2004 presidential election. The findings provide strong evidence that recruits are over-emphasizing local

information and have war-specific tastes and preferences that makes enlistment decisions more complicated than a full-information utility-maximization model of risk and monetary compensation would predict.

The second chapter uses nearest-neighbor matching techniques to look at performance of Major League Baseball players after they win awards in order to shed light on the more general question of how rational agents perform after they have been rewarded for good behavior up to that point. Comparing individual player's performance after winning major awards to their performance before winning shows that although players do perform significantly better in the year in which they win the award, performance after the award is generally indistinguishable from pre-award performance. Matching methods based on both baseball writer voting and performance statistics also indicate the likely absence of any sort of "curse" from winning awards for the winners themselves, their teams, and their teammates.

The third chapter, which is co-authored work with Michael Kremer and Edward Miguel, uses data from a randomized controlled trial, the Girls Scholarship Program (GSP), as well as the Kenya Life Panel Survey (KLPS) to conduct three types of analysis of bursary programs. We evaluate the effect of different targeting rules for secondary school scholarships, we estimate the impact of attending a primary school that took part in a scholarship program, and we estimate the effect of winning a scholarship from the program. Giving scholarships based on KCPE alone would lead to under representation of children whose parents have no secondary education and girls relative to their proportion of the population. Distributing the scholarships to the top students in each school as opposed to each district does little to alleviate this discrepancy. Analysis of the medium-run impacts of the Girls Scholarship Program, gave largely inconclusive but suggestive evidence that there were moderate benefits from attending a scholarship program school on the order of one half of the benefits observed in the original study held immediately after the scholarship program. The evidence indicates that scholarship winners did not benefit greatly from the award itself.

To my sister Emily

"Just because you feel it doesn't mean it's there."—Thom Yorke

Contents

List of Figures

List of Tables

Acknowledgments

I owe thanks to many people for helping me to complete my dissertation and PhD.

With regards to the first chapter, many thanks to Dan Acland, John Warner, Alex Gelber and the personnel of the Defense Manpower Data Center for the data and assistance they provided. Thanks to my ever-helpful advisor Ted Miguel, Justin McCrary, David Card, as well as Pat Kline, Fred Finan, and Jesse Rothstein for help with this paper. Also, I appreciate the helpful suggestions of Mark Borgschulte, Justin Gallagher, Willa Friedman and UC Berkeley labor lunch and development lunch participants for suggestions.

In addition to those above I would also like to thank Ulrike Malmendier, Stefano DellaVigna, and Alex Rothenburg for help with my chapter on baseball. I would also like to thank Ted Miguel again, as well as Michael Kremer, Anthony Keats, Karen Levy, and my wonderful field officers in Busia Kenya who made my field work in Kenya and my third chapter possible.

First, I wouldn't have been able to complete this work without taking the excellent courses of Ted Miguel, David Lee, and Ken Chay. I also appreciate the opportunity to teach for Martha Olney, and I would like to express thanks to CEGA, Ted Miguel yet again, and Temina Madon for the opportunity to help design and teach a course on Impact Evaluation.

I also owe thanks to my wonderful friends, family, office-mates, and Ashley Langer, Scott Williamson, Nitro, Nano, Gazelle, and other running and hiking partners for keeping me sane through this whole ordeal. Emily, the Rutmans, Marcus, Amy, Steve, Mark, and Lanwei, I wouldn't have survived without you.

Chapter 1

War Deaths and the Military Labor Supply

1.1 Introduction

This dissertation studies examples of applied econometrics in labor and development economics and attempts to obtain well-identified estimates of causal effects of government and social programs. The field of economics, and the social sciences in general, are limited by ethics and practicality in their ability to conduct randomized field experiments, the gold standard for causality in other fields. Several statistical methods have been devised to obtain causal estimates from "natural" or "quasi" experimental settings—settings where plausibly exogenous variation in a treatment effect of interest can be found and exploited to produce an unbiased estimate of causal effects. Some of these methods include panel data with fixed effects, nearest-neighbor matching, and regression discontinuity. This dissertation explores applications of these econometric methods, as well as an actual randomized controlled trial, in issues of labor and development economics.

This first chapter investigates the question of the effect of war deaths on the US military labor supply. The military plays a very large role in the United States economy. In 2010, President Obama signed a bill authorizing \$680 billion in military spending, making up nearly 20% of total federal expenditures.[1] Of this, \$130 billion was for the wars in Iraq and Afghanistantan, which have so far claimed the lives of nearly 7,000 US and coalition soldiers.[2][3] Another \$177.5 billion of this was to be spent on direct compensation to military personnel and family.[4] As of July 2010, this money was being used to pay and support over 1.4 million active duty men and women in uniform, and another 1.4 million National Guard

---

[1]http://www.nytimes.com/2009/10/29/business/29defense.html
[2]http://www.gpoaccess.gov/usbudget/fy10/pdf/fy10-newera.pdf
[3]http://www.cnn.com/SPECIALS/war.casualties/index.html
[4]http://www.defenselink.mil/news/2010%20Budget%20Proposal.pdf

and Reserve troops.

Recruiting this many troops costs a great deal. Most of the soldiers in the military serve for only a few years, so the military needs to recruit approximately 200,000 new troops every year. A 2003 GAO report lists the Defense Department annual recruiting budget as $4 billion, roughly $20,000 per recruit with over $1900 per recruit spent on advertising alone.[5] Part of this recruiting budget is spent on the salaries of production recruiters, active-duty men and women whose job it is to find new recruits. As of 2010, the Army employed over 8,000 soldiers as recruiters, and the Navy 4,897.[6]

The purpose of this paper is to determine how the difficulties these recruiters face change when soldiers are killed in war. The direction of the effect of deaths on recruiting is not obvious. One possibility is that the brothers and friends of a soldier killed in Iraq will come to disapprove of the military and become far less likely to join. Conversely, if one is personally convinced of the virtue or necessity of the war in which someone is killed, then a sense of duty or patriotism might lead one to become more likely to join the military after a solider from the local area has been killed. As anecdotal evidence of this, 61-year-old orthopedic surgeon Bill Krissoff acquired an age waiver and enlisted in the Navy Medical Corps after his son Nathan was killed in the Marines in Iraq.[7] Since some of the Marine Corps' support operations are provided by the Navy, serving in the Navy Medical Corps enabled Krisoff to give medical care to those with whom his son had fought and died. This paper is an empirical test of whether the deterrent or incentive effect is larger.

My analysis draws on a valuable new dataset obtained through the Freedom of Information Act consisting of the complete set of active duty enlisted recruits from 2001 to 2006, matched with detailed data on every death of a US soldier that occurred in Operation Iraqi Freedom and Operation Enduring Freedom during the same period. With very detailed geographic and date information, I am able to analyze recruiting at the county-month level, a significant improvement upon much of the literature. I use data on the home locations of recently killed troops and correlate the death rate of soldiers from the local area with the local rates of recruiting while flexibly controlling for the underlying characteristics of counties as well as nation-wide changes over time using county and monthly fixed effects and state time trends. After controlling for these underlying characteristics, the hometown of the casualty is arguably exogenous, and I use this source of variation to analyze the causal effects of local deaths on local military recruiting.

Using Poisson maximum likelihood estimates of semi-elasticities, I find that when a soldier died in Iraq or Afghanistan, that soldier's home county saw a decrease in recruiting of almost one percent. This effect is similar for recruits in both of the stages of the recruiting process that I test. I also obtain very similar semi-elasticity estimates using both Poisson regression and least squares weighted by county population, and the estimates are very stable across

---

[5]http://www.gao.gov/new.items/d031005.pdf

[6]http://www.2k.army.mil/faqs.htm, http://www.cnrc.navy.mil/PAO/facts_stats.htm

[7]http://www.npr.org/templates/story/story.php?storyId=17013597

different sets of fixed effects. I find that the effect is very localized—deaths in neighboring counties and deaths in counties in the same media market produce no deterrent effect. There is also considerable heterogeneity in the deterrent effect of local deaths. It is significantly larger in counties with higher than average African-American populations and is significantly smaller (and even positive) in more populous counties, counties with higher unemployment, and counties that voted for George W. Bush in 2000 or 2004. I also find that the effect is different for different types of recruits. It is larger for recruits to the Marine Corps and smaller for recruits to the Air Force. It is also significantly higher for recruits of the highest quality as measured by Armed Forces Qualification Test (AFQT) score and educational attainment.

Deaths at the national level are also of interest and likely have an effect on the difficulty the military services have in recruiting. National recruitment rates and deaths tend to fluctuate quite a bit, so I exploit this variation and briefly investigate this question. I find a negative correlation between total national deaths and national recruiting. A one percent increase in deaths is correlated with a two percent decrease in national recruits. However, whether this relationship is a causal one is of course a difficult question. While a nation-wide causal relationship may in fact exist, controlling for all the unobserved factors in addition to total deaths that determine how many people in the entire nation are interested in joining the military is likely an impossible task. Public support for the war, the state of the economy, and the state of the war effort itself are all extremely difficulty to quantify, so omitted variable and endogeneity issues would likely be prohibitively difficult to overcome. This paper thus does not address the question of the causal effect of national deaths.

My findings on local deaths do, however, provide evidence that people are not behaving in the way that a simple full-information utility-maximization model would predict. I show that one's hometown when enlisting has little to do with the likelihood of death given enlistment. Thus if a potential recruit were to observe that a soldier from his or her county had been killed in a war, that soldier's death has no more bearing on his or her own risk from joining the military than the death of a soldier from halfway across the country. If a potential recruit were basing his enlistment decision on standard factors such as monetary compensation and the risk of death, a county death would have the same effect as a state death. The data shows that this is clearly not the case. Thus individuals must either be updating their priors with incomplete information and misperceiving the actual risk or basing the non-pecuniary benefits they receive from military service on the proximity of deaths, and not just the number.

These findings should also be of great interest to the military and those who determine its funding. By analyzing the characteristics of a county, the military could produce a detailed estimate of the effect of deaths on recruiting in that county. If the military desires a wide geographic recruiting base, or if they desire to minimize costs, they could use the findings in this paper to help in their decision to reallocate funding and recruiting manpower.

The rest of the paper is as follows: section 1.2 describes the military recruitment process and places my work in the literature. Section 1.3 presents a very simple model of occupational

choice that helps to frame the empirical results, section 1.4 describes the data used for my analysis, section 1.5 presents the analysis, and section 1.6 concludes.

1.2 Background on Military Recruiting

All recruiting of enlisted members of the military is handled by military recruiters. Being a recruiter is very similar to other military occupational specialties—recruiters are mostly enlisted men and women who are assigned to a specific location for a three year stint, without absolute control over where they are assigned. Recruiters work out of offices spread all over the country, often in shopping malls or heavily trafficked areas. Whenever anyone enlists in the military, it is through such a recruiting office.

Potential recruits in the first stage of the process are referred to as "applicants." When a potential recruit first calls on the telephone or walks in the door and expresses interest in joining the service, the recruiter will make sure the candidate meets certain medical and legal requirements, for instance, he or she can have no felonies, cannot be on probation, and cannot be a single parent. The recruiter will enter data on the potential recruit into the database system as soon as possible after the initial expression of interest. The interested party will typically take a short (30-minute) practice version of the Armed Services Vocational Aptitude Battery (ASVAB). If they don't perform very well, perhaps they'll be told to study for a bit before taking the actual 3-hour ASVAB, but those who seem prepared would soon travel to a regional processing center (at a location other than the storefront recruiting center they've been visiting) and take the ASVAB, in some cases as soon as the day after expressing initial interest. Four of the 11 sections of the ASVAB are used as the AFQT. Assuming that a potential soldier passes the examination (by scoring a 31 or higher for Army enlistments) the applicant will then return to the recruiting center and be shown by their recruiter what jobs are available and when. Potential recruits are entered into my data set as soon as they have taken the ASVAB.

Once a potential recruit has taken the exam, chosen a military career (the availability of which depends on their test score) and is assigned a departure and enrollment date, they can sign a contract and take the oath of military service. This is the point at which a potential recruit is recorded as a "contract" in the data. When a contracted recruit finally ships off to training, they are recorded as an "accession." These accessions are the most commonly reported figures in the media and in Defense Department press releases pertaining to the military having reached its recruiting goals, but it is common in the literature to use data on contracts, since the accession date is more under the influence of the needs of the military, and is thus more demand-constrained and exhibits very strong seasonal fluctuations.

Although a good deal of research has been conducted on the labor supply elasticity of the all-volunteer US military, little, if any, has analyzed the effect of war-time deaths on the labor supply. Most of the existing research has focused on the supply elasticity with respect to salaries.

The literature on the military labor supply is vast, starting with 1960's estimates of the military in the absence of a draft, such as Altman and Fechter [1967]. These typically regress the number of recruits of certain types against unemployment, military and civilian wages by region, and regional demographic controls. They find that recruits who fell into higher quality "mental groups" based on test results were less affected by military pay and unemployment than those who scored lower on entrance tests. They also report based on survey data that higher quality recruits on average would have been less likely to have joined the military had there not been a draft in place [Altman, 1969].

Some early work such as Ash, Udis, and McNown [1983] found that unemployment did not have a significant effect on recruiting and that the military labor supply had a low elasticity with respect to pay. McNown et al. [1980] went as far as saying "the evidence...is consistent in rejecting unemployment as a significant determinant of total accessions." This was contradicted in later work by Dale and Gilroy [1985] which showed that while earlier work used data on military accessions (the final stage in the recruiting process, defined as shipping off to boot camp), a slightly earlier stage in the recruiting process referred to as "contracts" is a much better dependent variable to use, because recruits can sign a binding contract to join the military up to a year in advance through the Delayed Entry Program. When they actually ship out is very seasonal and highest in the summer. Contract signing is smoother over time and when this data is used, recruitment has a high elasticity with respect to unemployment, from 0.8 to 0.9 in Dale and Gilroy [1985], or from 0.4 to 0.8 in Brown [1985].

More recent work has attempted to control for both the demand side of the process, acknowledging changes in the military's need for recruits, as well as the supply side. Some work that controls for the demand-side by using the quantity of production recruiters in a given location and time such as Dale and Gilroy [1984] has not found significantly different estimates once the extra controls were added. Yet other work by Dertouzos [1985] controls for endogeneity by using recruiter quotas and explicitly models the tradeoff a recruiter faces between high and low quality recruits and finds that elasticities with respect to pay and unemployment are even higher once these extra controls are included.

One of the more interesting data sets is that of Hanssens and Levien [1983], which has monthly advertising and recruiting spending for each of the 43 Navy Recruiting Districts from January 1976-December 1978. They find that television advertising does serve to increase recruiting, but less than the number of active recruiters and the money allocated to their efforts (appearing at schools, posting flyers, handing out brochures) and their effort (as defined by their monthly recruiting quotas). Yet the number and effort of recruiters was found to matter less than economic or environmental conditions such as unemployment. Using similarly detailed advertising budgets for the Army from 1981 to 1984 with detailed data on the number of recruiters, in addition to modeling recruiter effort using quotas for low and high-quality recruits, Dertouzos and Polich [1989] find very similar estimates—unemployment and wages have high supply elasticities. The number of recruiters is also associated with large increases in recruits, but advertising has very small elasticities. How-

ever, the marginal cost of an additional recruit through extra recruiters is found to be similar to the cost through advertising ($6000), which is far cheaper than the marginal cost through enlistment bonuses ($16,000). When adjusted to today's dollars, their estimated costs of a new recruit are roughly in line with the average cost today.

One attempt to estimate the effects of changes on the demand side of the recruiting process is Asch [1990]. The data set covers a five-month period for the Chicago Navy Recruiting District–monthly observations of 125 recruiters, with data on individual recruiter successes and their quotas and points based on a complex reward system that was in place at the time. She finds that recruiter effort does appear closely based on the potential of winning awards, as recruiters exhibit different amounts of effort over the different stages in the 12-month award cycle. However, Oken and Asch [1997] show in a retrospective history of all four service branches that both the quota and reward systems have varied significantly across both time and service branches, and at times both the Army and the Navy did not even set individual recruiter goals, although they continued to set nationwide and recruiting district goals. More recent work has also focused on accounting for the demand side of the recruiting process, either using quotas and a control function approach as in Asch and Heaton [2008] or by instrumenting for the military wage using the legislative formula for annual increases, as in Gelber [2007].

Some limited research has been done analyzing the effect of Iraq and Afghanistan casualties at the national level—Asch et al. [2010] estimates that casualties in Iraq were responsible for anywhere from a 6 to 60 percent decline in Army recruiting depending on methodology, and Simon and Warner [2007] find that an additional 400 casualties per year in Iraq is associated with a decrease of high quality Army recruits of 6 percent. Yet no papers have been published using the spatial variation in US military combat deaths in Iraq and Afghanistan (or those from any other war, for that matter) to examine the effect on recruiting. Still, some researchers have used spatial variation in the deaths for other purposes. For example, Karol and Miguel [2007], used the plausibly exogenous variation in the geography of the deaths to examine the effect on changes in voteshare for George W. Bush between the presidential elections of 2000 and 2004. They find strong negative localized effects of deaths–without the deaths, Bush might have won an additional two percent of the national vote. Earlier work by Gartner and Segura [1998] and Gartner et al. [1997]use the geographic variation in casualties from the Vietnam War to show that local casualties have a very strong relation to public approval of the President and his handling of the war. They find that this effect is strongest during the first half of the war, when casualties are accelerating, while in the second half of the war, as the rate of casualties is decreasing, socio-demographic characteristics play a more important role than local casualties.

There is also recent research by Condra et al. [2010] that uses the flip side of the coin of US military combat deaths—detailed military data on spatial variation of Afghan and Iraqi civilian casualties. This research shows that local civilian deaths lead to more incidence of local violent attacks in Afghanistan, but not in Iraq. Perhaps future research will be able to synthesize or generalize the effects of war deaths on different types of individuals on all

sides of battles.

1.3 A Simple Model

In order to better frame the empirical analysis in economic terms, I will briefly discuss a model of occupational choice, adapted from standard models in Roy [1951] and Rosen [1986], which have previously been used to discuss the military in Warner and Asch [1995] and Fisher [1969]. Additional insights are adapted from behavioral models as discussed in DellaVigna [2009].

Assume that individuals are choosing between two occupation types, military (M) and civilian (C), and utility depends on wages (w) as well as a taste parameter (τ). Thus,

$$u^C = w^C + \tau^C$$

and

$$u^M = w^M + \tau^M. \tag{1.1}$$

Individuals will choose to enlist if $u^M > u^C$, or

$$(w^M - w^C) > \tau = (\tau^C - \tau^M) \tag{1.2}$$

that is, if the pay differential is greater than their relative preference for civilian life. Taste for military employment is a function of both the perceived risk of death an individual would face when employed by the military and an innate desire to serve in the military for cultural, patriotic, or other psychological reasons, which is itself a function of perceived risk of death. I write:

$$\tau^M = \tau^M(p(\hat{r}), \hat{r}) \tag{1.3}$$

where $p(\hat{r})$ is the level of patriotism or innate desire for a given individual and $\hat{r} \in (0,1)$ is the perceived risk of death in the military. I assume that p and τ^M are differentiable functions that vary across individuals in the population, creating the potential for different outcomes for different individuals. The empirical analysis in section 1.5.3 provides strong evidence that counties (and presumably, the individuals within those counties) have heterogeneous responses to risk and death depending on characteristics such as racial demographics and political preferences.

If I were to assume that patriotism were fixed for each individual instead of being a function of risk, that is, $\tau^M(p, \hat{r})$, then theory would predict that $\frac{\partial \tau^M}{\partial \hat{r}} < 0$, that more relative risk would make an occupation type less desirable, since higher risk of death or injury would lower future expected earnings. But by allowing patriotism to be a function of perceived risk, and by allowing for the possibility that $\frac{dp}{d\hat{r}} > 0$, it thus becomes a possibility

that $\frac{d\tau^M}{d\hat{r}} > 0$, that the military becomes more attractive as it becomes more dangerous. As with the anecdotal example of Bill Krisoff mentioned earlier, additional deaths, which are signals of potential for future danger, may increase an individual's sense of duty, revenge, patriotism, or honor and make military employment more preferred. There is also anecdotal evidence that recruiting stations were overwhelmed with potential recruits after 9/11, but it is a goal of this paper to empirically determine whether increased risk actually led to more or fewer recruits.[8] To estimate $\frac{d\tau^M}{d\hat{r}}$, I assume the other terms in (1.2)—the preference for civilian employment, and the military and civilian wages—are all constant with respect to risk of death in the military. These and other identifying assumptions are discussed further in section 1.5.

In addition to empirically determining the sign and magnitude of the partial derivative mentioned above, the central economic question analyzed in this paper is whether individuals accurately perceive the increased risk they would face by enlisting in the military. Observing that the nation is at war and that soldiers are dying, potential recruits are assumed to infer some likelihood of their own death several months out into the future if they were to enlist. But information acquisition may be costly. Media coverage may be biased towards local events, so individuals with limited resources available for information acquisition may not be as well informed about deaths of soldiers from more distant locales. Or even if equally well informed, distant deaths may somehow seem less emotionally salient than deaths of soldiers from the local area. Thus I write $\hat{r} = r(d^{local} + (1-\theta)d^{distant})$, where $\theta \in (0,1)$ is the degree of inattention paid to distant deaths, a function of salience and competing stimuli. If in actuality deaths from distant locales have no more effect than local deaths on the risk of death, a standard model of full information would assume $\theta = 0$. A major purpose of this paper is to see if, for whatever reason, $\theta > 0$ in the observed data, and individuals are responding differently to local deaths than they are to distant deaths. The empirical results in section 1.5 consistently show that potential recruits are responding far more strongly to local deaths than to distant deaths. [9]

---

[8]Compare http://www.nytimes.com/2001/11/12/us/nation-challenged-recruit-self-described-slacker-decides-he-s-ready-be-soldier.html, which describes an individual motivated to enlist to http://www.nytimes.com/2001/09/16/us/after-attacks-military-despite-national-rush-emotion-recruiting-centers-aren-t.html, in which recruiters claim not to have seen a significant increase in qualified recruits.

[9]I find that the recruiting response is always significantly different for local deaths (deaths in the same county) than for deaths from more distant locales (deaths from outside the county but in the same state, and deaths from outside the state). Coefficients on local deaths are typically on the order of five times larger for local deaths than for distant deaths. It is ongoing work to use this fact to estimate the actual θ parameter. See Chetty et al. [2009] for estimation of a very similar parameter, or the companion Chetty et al. [2007], which develops a structural interpretation for θ using bounded rationality.

1.4 Data

The data used in this paper is a rich new set with valuable information. The military has typically not released or maintained publicly available datasets of the deaths of its soldiers in the last two decades, which were numerous even in times of peace. The onset of the wars in Iraq and Afghanistan has changed this, making data on a large number of deaths available to the public with relative ease. The recruiting data used in the literature has also typically been analyzed at the quarterly or yearly level, often at the state level, while my data contains the exact dates of applications and the ZIP code for each applicant, which I have aggregated to the monthly-county level.

The recruiting data used in this paper was obtained through Freedom of Information requests to the office of the Secretary of Defense and handled by the Defense Manpower Data Center. It consists of three distinct sets of individuals: "applicants," "contracts," and "accessions" (explained above) and contains the date on which these individuals were recorded as starting one of the three specific parts of the recruitment process, ZIP code, AFQT score, educational attainment, and branch and component of the military to which the potential recruit was applying. The same data is available for applicants and contracts, but the data are stored separately for each step in the recruiting process and are unfortunately not linked by individuals across datasets.

I have recruiting data for fiscal years 1990-2006. (The military operates on an October 1-September 30 fiscal year). The applicants data set contains 6.4 million active duty observations, the contracts data set has 3.6 million active duty observations, and the accessions data set has 3.0 million active duty observations. There are nearly 50% more observations when one includes reserve and guard recruits (I observe roughly 9 million total applicants), however it appears that much of the contracts data for Reservists and Army and Air National Guard are missing (the data contain only 375 Army Reservist contracts, an implausibly low number over a 17-year period). To account for this, all of the analysis is run using only the applicants or contracts to the active duty components of the military. I am able to match roughly 96% of the applicant observations by ZIP code to a US county.

The casualty data come from a public list compiled by the Statistical Information Analysis Division at the Defense Manpower Data Center and freely obtained at http://siadapp.dmdc.osd.mil/personnel/CASUALTY/castop.htm. Starting October 7, 2001, every fatality in Operation Iraqi Freedom and Operation Enduring Freedom (in Afghanistan) is listed, and includes the service branch, component (active/reserve/guard), name, rank, pay grade, date of death, hostile status of death, age, gender, home of record city, home of record county, home of record state, home of record country, unit, incident geographic code, casualty geographic code, casualty county, city of loss, and race/ethnicity of the deceased. An important point to note here is that the data include "home of record" which is where the soldier lived on the day they joined the service. This is important with regards to my claim of plausible random assignment of death with respect to county after controlling for military population levels—the data is not tainted by service-men and women with very dangerous

military professions buying homes near their duty-base and changing their legal residence to the county in which the base resides.

Unfortunately this data does not include the fairly common deaths that occur in routine military training such as helicopter crashes. Outside of deaths directly related to the wars in Iraq and Afghanistan, the only data publicly available are annual figures without any geographic information or detailed information on date or cause. Yearly figures indicate that in the period for which I have recruiting data (1990-2006), the year 2000 had the lowest number of military deaths–zero from hostile action, but 758 from accident, illness, homicide, or suicide. 1991 had 1787 military deaths, only 147 of which were from hostile action in the Persian Gulf War. Table 1.2 shows total annual military deaths and the subsets of those recorded as hostile action and those considered part of the Iraqi/Enduring Freedom operations as reported by the DMDC. One can see that at most just under 50% of the deaths of active duty US military members are included in the Iraq/Afghanistan data. It would be interesting to analyze the effects of military deaths not directly related to war efforts (helicopter training, suicides, homicides, etc.), but unfortunately these are not recorded in this data. I have filed a FOIA request for *all* military deaths from 1990 to 2006, and will incorporate this into my research as soon as I am granted access, but at present all analysis is limited to the period from October 2001 to August 2006. During this period there were 2886 deaths, 2725 of which I have been able to link to the home county of record of the deceased soldier.

Although my analysis primarily rests on the panel nature of the data and the inclusion of area and time fixed effects to identify the effect of local deaths, I have also included time varying characteristics of counties to the extent that they are available. These include unemployment at the national, state, and county level as reported by the Bureau of Labor Statistics, and mortality for young males age 18-24 from the Multiple Cause of Death files at the National Center for Health Statistics National Vital Statistics System. Unfortunately, this data is not available in a useful format after 2004. Starting in 2005 the publicly available data is stripped of all geographic identifiers below the national level. I have applied to the NCHS for access to the restricted detailed data and hope to be able to include this data at the county level for 2005 and 2006 soon. Statewide numbers of recruiters by service branch have also been included in certain specifications.

1.5 Analysis

Although this paper focuses on the effects of local deaths on local recruiting, it is worthwhile to think about the effects of total national deaths on national recruiting. Figure 1.1 shows monthly total national combat deaths and monthly total applicants to the military from October 2001 through July 2006, and Table 1.1 shows a simple regression analysis of the this time series. Graphically, spikes in deaths after the initial invasion of Iraq and the first and second battles for Fallujah (the obvious high points in the figure) are very clearly followed by

decreases in recruits. With one observation for every month nation-wide, there are only 58 observations, but there is still a strong and consistent negative correlation between deaths in the current and/or previous month and recruits. In terms of elasticities, as shown in the table, a one percent increase in deaths is associated with a 1.5 to 2 percent decrease in applicants and a 2.0 to 2.7 percent reduction in contracts. (Standard non-logged OLS regressions show that deaths are associated with 60-90 fewer applicants and 27 to 43 fewer contracted recruits.) So it would seem that deaths in the military are followed by an overall decrease in the national number of recruits. This should not necessarily be given a causal interpretation, as a simple linear time trend is not nearly enough to control for all the unobserved changes that occurred in the country over this nearly five-year period, all of which could be biasing the estimate up or down.

The same potential problem does not exist once I narrow the analysis to a finer geographic level and use repeated observations from multiple states or counties over time, where I can use fixed effects to control for unobserved characteristics. It seems prima facie obvious that there are significant differences between states and counties that might be correlated with both the number of deaths and the number of recruits. County population jumps to mind. A naive regression of counties' recruits on death without accounting for population would have an upward-biased estimate due to omitted variable bias, since with equal recruiting rates, the higher a county's population, the more recruits from that county, and mechanically, the more deaths from that county. With an obvious measure like population, it is possible to get rough estimates of the population that change over time, but with many more subtle county characteristics, such as abstract support for the military, it is not possible to get even one estimate for the county, let alone multiple measurements over time, thus the need for fixed effects to flexibly control for all immeasurable county characteristics and eliminate omitted variable bias.

A more subtle characteristic that may have an effect of the interpretation of my results is the specific type of military career for which the residents of certain counties or states are likely to sign up. It seems likely that those in the infantry are more likely to be killed than those in ancillary support operations. And it is possible that recruits from certain states are more likely than others to sign up for the more dangerous occupations. To my knowledge, data on military occupational specialties is not publicly available. Instead I can make a comparison of the likelihood of death for a recruit from each of the states. Figure 1.2 shows that, as one would expect, there are clearly differences in state populations. The horizontal axis is a state's percentage of the nation's male 18-24 year-old population, and the vertical axis shows a state's percentage of the applicants over the entire 16-year period for which I have data. With one observation for each state, a constant national propensity to enlist would yield a population weighted OLS regression coefficient of 1, which can be easily rejected statistically. Figure 1.3 shows a state's percentage of the young male population and its percentage of the deaths in Iraq and Afghanistan. This slope is also statistically different from 1, indicating that there is not one constant national likelihood of death in the military. This finding may be unsurprising, and is shown only since it may be interesting to

know in and of itself which states have higher proportions of their young men enlist and die in the military.

Regarding the issue of whether people from a certain state are more likely to work in dangerous military specialties, since I lack data on military occupational specialties, I can compare the number of recruits from a state to the number of deaths from the same state. Figure 1.4 shows histograms of the ratio of active duty deaths to total active duty applicants for each state over the whole period for which I have data. The ratios are centered around 0.3%, but are clearly not all the same. I have repeated this exercise including both active duty and reserve and guard deaths (since service and death in the reserve and guard duty is clearly correlated with where one lives, including them might lead to complications) for both applicants and contracts, using both unweighted and population-weighted means. The coefficients of variation for each of these eight methods of calculating the risk of death by state are relatively small, ranging from 0.143 to 0.319. Simply put, recruits have about the same likelihood of dying, regardless of where they are from.

Another way to analyze this is to look at each individual observation, and check the likelihood that it came from a binomial distribution with the hazard rate equal to that of the overall national hazard rate. (The number of active-duty deaths divided by the total number of active duty applicants was roughly .003.) I then tested the likelihood that each observation came from a binomial distribution with this hazard rate of p=.003. Figure 1.5 displays two histograms of the p-value for each state, one using active-duty deaths and active duty applicants, the other using active-duty deaths and active-duty contracts. The histogram displays the p-values as calculated, but to interpret, one should use Bonferroni or Sidak corrected p-values (i.e., divide the cutoff for significance by the number of tests, 51, thus replacing a cutoff of .05 with 0.05/51=0.0009). Only two of the state observations (Florida and Massachusetts) reject the null hypothesis that their true probability is in fact .003 using applicant data, and only one, Massachusetts, rejects using contracts data. So it is possible that recruits from certain states are more likely to enter dangerous military occupations, but according to my evidence, the idea that the risk of death is the same across all states can only *barely* be rejected.

This is not necessarily a reason to worry about omitted variable bias, since fixed effects for each state will still be able to control for this underlying characteristic of the state, however, it does give a slightly different meaning to the estimates I will develop in the next few pages. If deaths in Iraq and Afghanistan were truly uniformly distributed amongst all the troops, regardless of where they came from, then the fact that a soldier from a given county, say Fairfax County, Virginia, had died would provide no more information regarding the risk of death to a potential recruit from Fairfax County upon enlisting than would the death of a soldier from Maricopa County, Arizona. Any extra deterrent to enlisting because this death occurred locally would thus be an emotional or behavioral response and not an accurate updating of preferences based on risk. However, if a recruit from Fairfax County were more likely to sign up for front-line occupations, and those are the soldiers who were dying, then this death would actually contain useful information as to the risk of death,

and an extra deterrent effect might be warranted for those reasons. Looking at the above histograms of the rates of death by state, it seems that soldiers from different states have only slightly different hazard rates, and it is not the case that one state or another with a supposed reputation for strong military support or lack thereof has a vastly different rate of death of its soldiers.

1.5.1   County-Level Least Squares Analysis

The previous discussion focused on national or state-by-state comparison, but since I possess data on ZIP code for recruits and data on home city for deaths, I have aggregated the data at the county level. County-level least squares analysis uses the regression equation : $Recruits_{ct} = \beta_1 CountyDeaths_{c,t-1} + \beta_2 CountyUnemployment_{ct} + x'_{ct}\eta + \alpha_c + \gamma_t + u_{ct}$, where x includes in-state (but out of county) deaths as well as state unemployment. γ_t is a fixed effects for every month, so national characteristics that are the same across counties in any given time period such as the total national number of deaths, national unemployment rate, or the military wage rate cannot also be estimated. Results at the county level are shown next in Table 1.3. The left half of these regressions show the analysis done for applicants, the right hand side for contracts, one step further in the recruiting process. All observations are weighted by county population (I have run un-weighted regressions, in which the results are equally significant, if only about 1/3 as large). One can clearly see that with a full set of fixed effects, in-county deaths in the previous month are associated with 15 fewer applicants and nearly 10 fewer contracts. In-state deaths are associated with 0.3 to 1.1 fewer applicants, and 0.2 to 0.7 contracted recruits fewer, although the estimate is sometimes positive and not always statistically different from zero. The coefficient on lagged in-county deaths is the main estimate of interest, and is remarkably stable across different sets of fixed effects after county-level fixed effects are included.

Unemployment levels are also closely correlated with recruiting levels. State unemployment has fairly volatile estimates depending on which fixed effects are included, but coefficients for county unemployment are stable; an one percentage point increase in county unemployment leads to two to four more recruits. A simple comparison of the coefficients on lagged county deaths and county unemployment indicates that one fewer county death would cause the same increase in recruiting as a 4 to 5 point increase in county unemployment.

One could conceivably be concerned that county unemployment was measured somewhat inaccurately and in a way that somehow biased the estimates, especially given the relatively small sample size used to calculate each county's unemployment rate and its mechanical relationship to state and national unemployment levels. I have tested this by adding normally distributed noise to the county unemployment estimates; none of the other estimates changed significantly.

The idea that potential recruits are responding to county and state unemployment above and beyond the national unemployment level is in accordance with rational utility-maximizing individuals, assuming that moving across county or state lines (or finding em-

ployment across county or state lines) is costly, as county and state unemployment levels directly affect one's likelihood of employment, and thus income and utility. Deaths of active duty soldiers from one's own county or state are unrelated to one's own likelihood of dying in the service, since the Army operates at a national level and recruits are put into military careers irrespective of their state or county of origin (at least to the extent discussed above). Clearly this is not quite the case with Reserve and National Guard troops, as Reservists simply report to the nearest base for their one weekend a month and two weeks a year of training, but their recruiting numbers are not included in this analysis.

I have calculated, but do not show estimates of the elasticity of the recruiting response to death based on least squares regressions. They indicate that a one percent increase in the number of in-county deaths leads to a 2.7% reduction in in-county recruits in the next month. However, I put less emphasis on this result for two reasons. First, the vast majority of county-months do not observe a death, and thus one cannot take the natural log of the data for simple elasticity calculations. Second, since most county-months have zero deaths, when any number of deaths is observed, there has actually been an infinite percent increase in the number of deaths. There are of course methods to deal with the first problem and other ways to calculate elasticities, however the second reason makes the interpretation of the 2.7 coefficient fairly odd. It's hard to know what sort of effect is implied by a $2.7 * \infty\%$ decrease in recruits. Thus I have chosen to focus on semi-elasticity estimates, where the right-hand side variable is in levels and the left hand side is logged. This also makes the OLS estimates more easily comparable to the Poisson estimates in the next section.

Table 1.4 shows these semi-elasticity estimates for both applicants and contracts. The dependent variable is the log of active duty recruits. Observations have been weighted by county population. The semi-elasticity estimates show that an in-county death in the previous month leads to a 0.8 to 2% reduction in both applicants and recruits. Out-of-county deaths lead to small increases in recruiting, although the effect is slightly more volatile than that for in-county deaths and even becomes negative when state*year interacted fixed effects are added.

1.5.2 Basic Poisson Analysis

Since data on recruits is in count form, the Poisson model is perhaps a more natural one to fit to the data. Poisson regression fits a generalized linear model of the form $log(\mu_i) = x_i'\beta$, so $\mu_i = exp(x_i'\beta)$ and a one unit increase in x_j multiplies μ_j by $exp(\beta_j)$. However, if I am modeling an underlying rate of enlistment, $\mu_i = e^{x_i'\beta}$, then the observed number of recruits would be the rate times the exposure, which in my case is the population of young males. If R_i is the expected number of recruits, then $R_i = Population_i * e^{x_i'\beta} = e^{ln(Population_i) + x_i'\beta}$. Thus all the Poisson models have been fitted with a coefficient constrained to 1 for the county's log young male population. Table 1.5 shows the results when Poisson regression is used to analyze the data at the county-month level. Fixed effects for each county and for each month are included, and state time trends in certain specifications. Observations are

weighted by county population, and standard errors are clustered by county. The results indicate that one additional in-county death is followed in the next month by a 0.79% to 0.90% decrease in applicants and a 0.77% to 0.84% reduction in contracts. Note that all death figures have been divided by 100 to make more useful digits of the coefficients visible, and thus all coefficients for deaths should be interpreted as percents and not fractions (i.e. 0.4 is 0.4 percent, not 40 percent). Deaths from in-state but out-of-county appear to have a small positive effect, from 0.1% to 0.2%, indicating that potential recruits are differentially affected by deaths from different areas, and that deaths from further away may increase the likelihood of joining the military. Unemployment at the state level has a negative effect: a one percentage point increase in unemployment leads to a small but not always significant decrease in recruiting, while a one percentage point increase in county unemployment leads to a 1.6% increase in recruiting.

This analysis has been done using both active duty and reserve and guard duty deaths. I do this because the main emphasis of my analysis is to determine the magnitude of the observed reaction to deaths. It may be true that the response to deaths of local soldiers from reserve and guards units is a rational response based on an updated assessment of the risk of death, but still, the magnitude of the observed deterrence effect, rational or not, would be what is of interest to policy makers. As a robustness check, however, I have run the analysis using only the active-duty deaths, and under this specification, in-county deaths are followed by a 1.0 to 1.1% reduction in recruits in the next month. The coefficients for out-of-county deaths and unemployment remain very similar.

Despite the inclusion of fixed effects, the potential for omitted variable bias still exists. One of the most obvious ways this might occur is through the action of the military's production recruiters. It seems likely that the number of production recruiters is positively correlated with the number of recruits, and at the extreme this is clearly true mechanically. If the number of recruiters (or their level of effort) were also correlated with the number of deaths, my estimates would be biased. Given that recruiters serve for three years in one place, I find it highly unlikely that the military is relocating them in a way that is correlated with monthly deaths. Without being relocated, however, recruiters may change their level of effort. To account for this, I hope to eventually include recruiting goals or quotas as a proxy for effort. I have filed FOIA requests for this data, but currently all I possess is the number of recruiters by state and quarter until halfway through 2004, so I have included this extra control variable for that portion of the sample. I also have detailed mortality data, but only through 2004. It is conceivable that deaths unrelated to the military would play a role in determining recruiting (for example, young men in a crime-ridden community may be anxious to join the military to escape) thus I include monthly male 18-24 year-old mortality figures as well. Table 1.6 shows these results. The analysis is done for both applicants and contracts, with the observations limited to those from October 2001 to June 2004. County and monthly fixed effects as well as state trends are included. One can see that the estimates of the effect of a death do not change very much (from -0.980 to -0.993 for applicants and -0.415 to -0.468 for contracts) when I add the extra controls. The estimates for contracts

are only about half of what they are using the full data set, indicating that the response to deaths is not constant over the entire sample.

Another interesting test of these results is shown in Table 1.7. Here I have included the number of deaths that occurred in contiguous counties and the number of deaths that occurred in counties that share the same media market as the main county of interest. County contiguity is defined using the 1991 ICPSR contiguous county file. [10] Media markets are defined using the Nielsen Media Research's Designated Market Area (DMA). The regressions show that deaths in nearby areas, whether defined using county borders or media market, do not effect recruiting in a given county. The deterrent effect of deaths appears to be very localized.

1.5.3 Poisson Analysis with Interactions

The analysis in the previous two subsections makes it clear that in-county deaths result in a significant decrease in county recruiting. An important corollary question is regarding the heterogeneity of this effect. All counties are unlikely to observe the same deterrent effect of death. Here I investigate the recruiting response to deaths based on a county's demographic and cultural makeup, specifically, its population, unemployment, racial makeup, rural/urban status, and political alignment. Table 1.8 displays these regressions. They all include monthly and county fixed effects, out of county but in-state deaths, as well as unemployment, and I have added county characteristics interacted with lagged in-county deaths. All variables to be interacted have had the population weighted mean subtracted. In a standard OLS regression this de-meaning would mean we could expect the main coefficient on lagged county deaths to remain fairly constant, but the weighting scheme used in Poisson regressions does not give the same sort of results. I have tested this and found that the main coefficient does stay very stable in least squares regressions regardless of what sort of weighted de-meaned interaction variables are included. I have also run these regressions using least squares and the results are qualitatively similar.

I have interacted lagged county deaths with county population. This interaction is deaths/population, so the interpretation is slightly different than all the other interactions, which are multiplied by the number of deaths. Also included are interactions with the monthly county unemployment figure (the only county characteristic I have that changes over time and thus is not collinear with the fixed effects and can be included in the regression by itself), percent African-American population as measured in 2005, racial fractionalization using percent white, black, Native American, Asian, and Pacific Islander, a binary measure of rurality using the USDA's Economic Research Service classification, and the percent of the county that voted for George Bush in 2004. The regressions are run for applicants and

---

[10]U.S. Dept. of Commerce, Bureau of the Census. CONTIGUOUS COUNTY FILE, 1991: [UNITED STATES] [Computer file]. Washington, DC: U.S. Dept. of Commerce, Bureau of the Census [producer], 1992. Ann Arbor, MI: Inter- university Consortium for Political and Social Research [distributor], 1992. doi:10.3886/ICPSR09835

contracts, the first column with percent black, and the second with racial fractionalization. I have also run regressions including interactions of all these same variables, but interacted with all four counts of deaths (in and out of county, lagged and current) the coefficients on the original interaction are very similar, and the coefficients for the interactions with out-of-county deaths and current in-county deaths all either go in the same direction as the ones shown in Table 1.8 or are statistically not different than zero.

(Deaths/100)/Male 18-24 year-old County Population has a rather large and negative coefficient. The coefficient implies that there is a level effect of deaths, and then an additional effect of deaths per capita, from -2086 to -2215 for applicants and a statistically insignificant -1510 to -1704 for contracts. This means that an increase of one death for every one young male (obviously improbably high) would result in a decrease of recruits by 1500 to 2000 percent (also improbably high). Simpler to imagine is that for every additional 1500 to 2000 young males in a county, the deterrent effect of deaths is one percent smaller (closer to zero). More-populated counties have a smaller percentage recruiting response to deaths than less-populated counties. With more people, perhaps other young men are less likely to hear the news of the death of a soldier, and if they hear it, perhaps they are less likely to have known the soldier who was killed and thus be relatively undeterred by his death. It is slightly puzzling, however, why this differential effect would be so strongly significant for applicants and not significant for contracts.

Death * County Unemployment yields positive but insignificant estimates for applicants and significant estimates from 0.779 to 0.842 for contracts, indicating that deaths in counties with higher unemployment are not as large a deterrent effect, and can even make the recruiting response to deaths positive. Under the first specification for contracts, a county with the weighted average level of county unemployment (5.5% in the sample) would have a 0.5% reduction in recruits for every death. A county with 6.5% unemployment would actually see a 0.5+0.8=1.3% increase in recruiting with every death (not to mention the 1.6% increase in recruiting the level effect of county unemployment) .

The percentage of county population that is African-American increases the size of the effect of a death. A county with the weighted average proportion of the population (13%) African-American would see a 0.5% reduction in recruits for every death, a county one standard deviation (13%) higher African-American population would see a -0.5+(13%*-0.06)=1.28% reduction in recruits for each death. The estimates are of the same order of magnitude for both applicants and contracts. I have also done the analysis using a more detailed description of the racial makeup of counties instead of simply percent African-American, racial fractionalization, but those estimates do not turn out significant.

Rural is a binary measure of whether the county is rural, and is highly correlated with my measure of population, so even though my population measure is for young males only, it's not surprising that the interaction of rurality with deaths is not significant. The sign does seem to go in the same direction, however, indicating that rural counties (with lower populations) would have larger negative recruiting responses to deaths. Running the regressions without the rural interaction does not change the coefficient on the population interaction

significantly.

Finally, I have interacted the county percent of the vote that went to George Bush in 2004 with deaths. The coefficient estimates range from 0.08% for applicants to 0.166% for contracts. This indicates that a county with the weighted mean Bush voteshare would see a decrease in recruiting of 0.5 to 0.7% for every death, but a county with one percentage point higher vote for Bush would see a 0.09% smaller (closer to zero) decrease in recruiting. This indicates that a county with roughly 6 to 8 percentage point higher than the weighted average Bush vote would see increases in recruiting after deaths. The average Bush vote share is 50.6%, and the standard deviation is nearly 14 percentage points. Well over half the counties had a Bush vote share over 57%. As shown in Karol and Miguel [2007], at least at the state level, war deaths led to poorer Bush election performance in 2004. As a robustness check I have replaced the Bush '04 vote share with Bush '00 county vote-share, which was obviously unaffected by Iraq and Afghanistan combat deaths. The estimates are nearly identical.

These estimates all show that county characteristics are very important in determining the response of a county's potential recruits to the news of a death. More heavily populated counties appear to have a smaller proportional response, as do counties with higher unemployment. Counties with higher fractions of African-American population have a larger response to deaths, as did counties that voted against George W. Bush (in either 2000 or 2004).

1.5.4 Recruiting Response for Different Types of Recruits

In this section I examine how the individual services fare in their recruiting. There are clearly differences between the services in terms of their operations; perhaps stemming from this there are often assumed to be significant cultural differences between the branches, leading to a difference in the type of people who make up the potential applicant pool for each of the services, and the possibility for a difference in the potential applicants' response to a local death. Table 1.9 shows these estimates, again using Poison regression analysis at the month-county level, with county and monthly fixed effects and state time trends. It appears that Marine recruiting decreases at a rate of over 2% for every death, while the Air Force reaction to death is a statistically insignificant 0.05% reduction in recruits. Deaths at the state level and state and county unemployment seem to have effects that are less clearly distinguishable across service branches. This makes sense given that very few of the deaths in Iraq and Afghanistan have come from the Air Force, while many have come from the Marines. But an even larger proportion have come from the Army, so if potential recruits were simply steering away from the branches of the military with the most deaths, the Army would be the branch with the largest recruiting response to deaths, which is not the case. I have repeated this exercise using weighted least squares regressions of the log of the recruits of each service branch, and the deterrent effects by service branch maintain the same ratios relative to one another, exhibiting further evidence that the recruiting deterrent is greatest

for the Marine Corps.

In addition to the demands for people in each individual service branch, the military, like any other organization, has a strong interest in recruiting high quality people. The services have generally held "high quality" to mean a person in possession of a high school degree and a score of 50 or higher on the AFQT. The services have often had separate quotas for high and low quality enlistees, and they have generally required that a high percentage of their recruits fall into the high category, although these requirements have changed over time with the needs of the services.

Table 1.10 shows results that are analogous to those in Table1.9, except with recruits of different quality levels. I have broken recruits into four groups. Low Quality recruits either scored below 50 on the AFQT or do not have a high school degree. High Quality have both a 50 or higher on the AFQT as well as a high school degree. High Quality-Alt scored 50 or higher on the AFQT but may still be in their senior year of high school (many recruits sign contracts while they are still in school, but join through the Delayed Entry Program, so they do not actually ship out until they graduate and are considered high quality recruits by the military.) Very High Quality recruits is not a specific distinction used by the military, but is meant to identify the most sought after recruits–those who have a 75 or higher on the AFQT and have taken at least some college courses. The results indicate that all but the Very High Quality recruits have roughly the same response to deaths: a slightly smaller than 1% reduction in recruits for every death. Amongst Very High Quality recruits, the effect is almost 2.5%.[11] Another interesting difference is the effect of unemployment on different types of recruits. *A priori* it is unclear how unemployment would affect different types of recruits. Higher unemployment could raise low quality enlistment because low quality individuals have fewer outside options, or it could hurt low quality individuals, because high quality individuals have their outside options eliminated, then they join the military in greater numbers, and there isn't enough demand remaining for low quality individuals to enlist. (While High Quality can typically enlist in the military at any time with no cap on demand, low quality recruiting is frequently subject to both demand and supply constraints.) The table seems to indicate that county unemployment leads to a 2.5% increase in high quality recruiting and a 1.5% increase in low quality recruits. Again, I've tested whether the results are the same when done using weighted least squares analysis, and the results exhibit the same patterns for deaths–very high quality recruits are more deterred by deaths than other

---

[11]It may be slightly surprising that higher-quality recruits are more deterred by local deaths if one interprets the response to a county death as an "over-response" compared to deaths from out of county or out of state, since higher quality recruits are better-educated and might be expected to read national newspapers or acquire information about distant deaths with lower cost. Indeed regressions not shown indicate that the response to out-of-county and out-of-state deaths is no larger for higher quality recruits than for lower quality recruits. However, the results are consistent with a story of the local-death-deterrent being due to personal knowledge of the soldier who was killed, since evidence indicates that those with more education are likely to have larger social networks. As written in Glaeser et al. [2002], "The connection between social capital and human capital is one of the most robust empirical regularities in the social capital literature."

types of recruits.

1.5.5   Recruiting Response for Different Types of Deaths

Amongst the four branches of the military, the deaths in Iraq and Afghanistan have been highly concentrated amongst soldiers in the Army and the Marines. In the data used in this paper, roughly 2000 deaths are from the Army, 800 from the Marines, 85 from the Navy, and 50 from the Air Force. Since there have been 40 times more deaths in the Army than the Air Force, it's entirely possible that potential Army recruits have instead gone on to join the Air Force instead. To test this, I have rearranged the data into county-month-service branch observations and test the recruiting response to a specific service branch after deaths from the same service branch and from other service branches. Table 1.11 shows the results of these tests. Poisson regressions shows the percent response to a death in the same service branch as the recruit, and in any of the other three service branches. As before, I have controlled for monthly fixed effects, state trends, and state and county unemployment. In these regressions I have included interacted county*service branch fixed effects as well. The bottom of the table shows pair-wise comparisons of the corresponding same-service and other-service death coefficients. For example, for applicants, a lagged same-service in-county death leads to a 0.7% reduction in recruits in that service, while a lagged other-service in-county death leads to a 1.1% reduction in recruits for that service. In all of the specifications, tests fail to reject the hypothesis that the effect of a lagged same-service in-county death is statistically identical to that of a lagged other-service in-county death. The same can be said of lagged same-service out-of-county deaths and lagged other-service out-of-county deaths. However, for several of the current-month deaths, statistical test reject equivalence. They indicate that same-service deaths have a larger (and positive effect) than do other-service deaths. *A priori* I would have assumed that if anything, same-service deaths would lead to a larger decrease in recruits, since recruits could either not sign up or substitute to a different service, but this does not seem to be the case. The results are further evidence that potential recruits are not using the information contained in a death (in this case, the service branch in which the death occurred) in a sophisticated or strictly risk-based manner.

A slightly different way to get at the question of deaths by service branch is to lump all the recruits together but show the specific response to deaths in a given service. These estimates are shown in Table 1.12. The coefficients are positive for Air Force and the Navy in-county deaths, however they are not statistically significant. I have also run specifications with the in-state but out-of-county deaths split by service branch, and I have run tests of equivalence of the coefficients for all the service branches. The hypotheses of equality cannot be rejected for county deaths, but can be for state deaths.

I have run similar analysis comparing deaths across several other dimensions, specifically, the gender of the casualty, the classification of the death by the military as hostile or non-hostile, the race of the casualty, and the war in which the deceased was killed (Iraq or Afghanistan). Casualties are found to have no significantly different deterrent effect based

on gender, hostility-status, and race.[12] However, the war in which the death occurred has a significant effect of the recruiting response. Table 1.13 shows that county deaths from Iraq lead to a 1.6 to 1.7% decrease in recruiting in the following month, while county deaths from Afghanistan lead to an increase in recruiting of 2.3 to 2.9% in the following month. These effects are significant with high confidence, and tests of the equality of the coefficients are easily rejected. This seems to be further evidence that recruits are responding not only to the risk of death, but are also exhibiting a response based on a subjective valuation of the circumstances of the death.

1.5.6   Lags of Deaths and Unemployment

My main empirical method thus far has been to compare county recruits in a given calendar month to county-wide and state-wide deaths in the previous month. It is possible that potential recruits initially deterred from enlisting by a death eventually "forget" about local deaths and join the military. Table 1.14 shows Poisson regressions with cumulative death and unemployment lags of two, four, six, and twelve months–that is, the sum of current deaths plus all the deaths that occurred in the previous number of months. The results indicate that deaths from previous months have, on average, a significantly smaller deterrent effect on recruiting than more recent deaths. Earlier regressions have shown a deterrent effect of nearly one percent for deaths in the previous month; these regressions show a much smaller average deterrent effect, decreasing to one-half a percent, down to one-third or even a statistically insignificant one-sixth of a percent deterrent effect for twelve months of lagged deaths. Weighted least squares regressions produce very similar semi-elasticity estimates.

1.6   Conclusion

The military's job of recruiting is a difficult and expensive task. I have shown in this paper that military deaths make this task significantly more complicated. At the national level, a one percent increase in the death rate is associated with a 1.5 to 2.5 percent decrease in national recruiting in the following month. This should not necessarily be given a causal interpretation, thanks to the potential for omitted variable bias. However, I make a strong case that panel data regression analysis at the county level warrants a causal interpretation, as I can flexibly control for county characteristics that are fixed across time, national trends that are constant across different counties, and even state-level time trends. Using both

---

[12]Although statistical tests cannot reject that the race of the death does not affect the size of the deterrent effect, I have also run regressions interacting the race of the death with county racial characteristics. The coefficient of interaction between black deaths and black population is twice the magnitude of the coefficient on the interaction of white deaths and black population, inidicating that perhaps counties with more blacks are even more deterred by black soldier deaths, although the difference between these two interactions is again not significant.

weighted least squares and Poisson regression shows remarkably similar and stable estimates of the effect of deaths of local soldiers on local recruiting. Each in-county death leads to a one percent decrease in that county's recruiting in the next month, and this finding is robust across several specifications. Thus a large fraction of the overall deterrent effect of deaths appears to be due to local deaths. I have also shown that the local effect is in fact quite concentrated—deaths in contiguous counties and deaths in counties in the same media market do not have cause a decrease in recruiting.

The localized deterrent effect also exhibits heterogeneity in interesting fashions. Counties with higher than average young male populations see smaller decreases in recruiting, as do counties with higher than average unemployment. Counties that voted for George W. Bush in 2000 or 2004 see very different, and even positive recruiting responses to local military deaths. Counties with higher than average African-American populations see significantly more negative responses to local deaths. The effect of a death leads to a larger decrease in recruiting for the Marines, and the Air Force sees smaller reactions. The military also sees the largest reduction in recruits of the highest quality (as measured by AFQT score and educational attainment) after a local death. However, it does not appear that recruits are responding significantly differently to deaths in their own service branch than to deaths from a different service branch.

Still, it is puzzling to the economist who assumes actors have full information and are completely rational why there would be any difference in the response to a local death than to a death from further away. I have documented that the likelihood of dying is not related to the location in which one enlists, so this paper provides evidence of a larger response to local matters than is justified based on calculation of risk alone. Models of non-standard decision making that include a salience parameter such as Chetty et al. [2009] or Hossain and Morgan [2006] may be able to better explain the observed recruiting phenomenon.

Figure 1.1: Graph of Monthly Recruits and Monthly Iraq/Afghanistan Combat Deaths
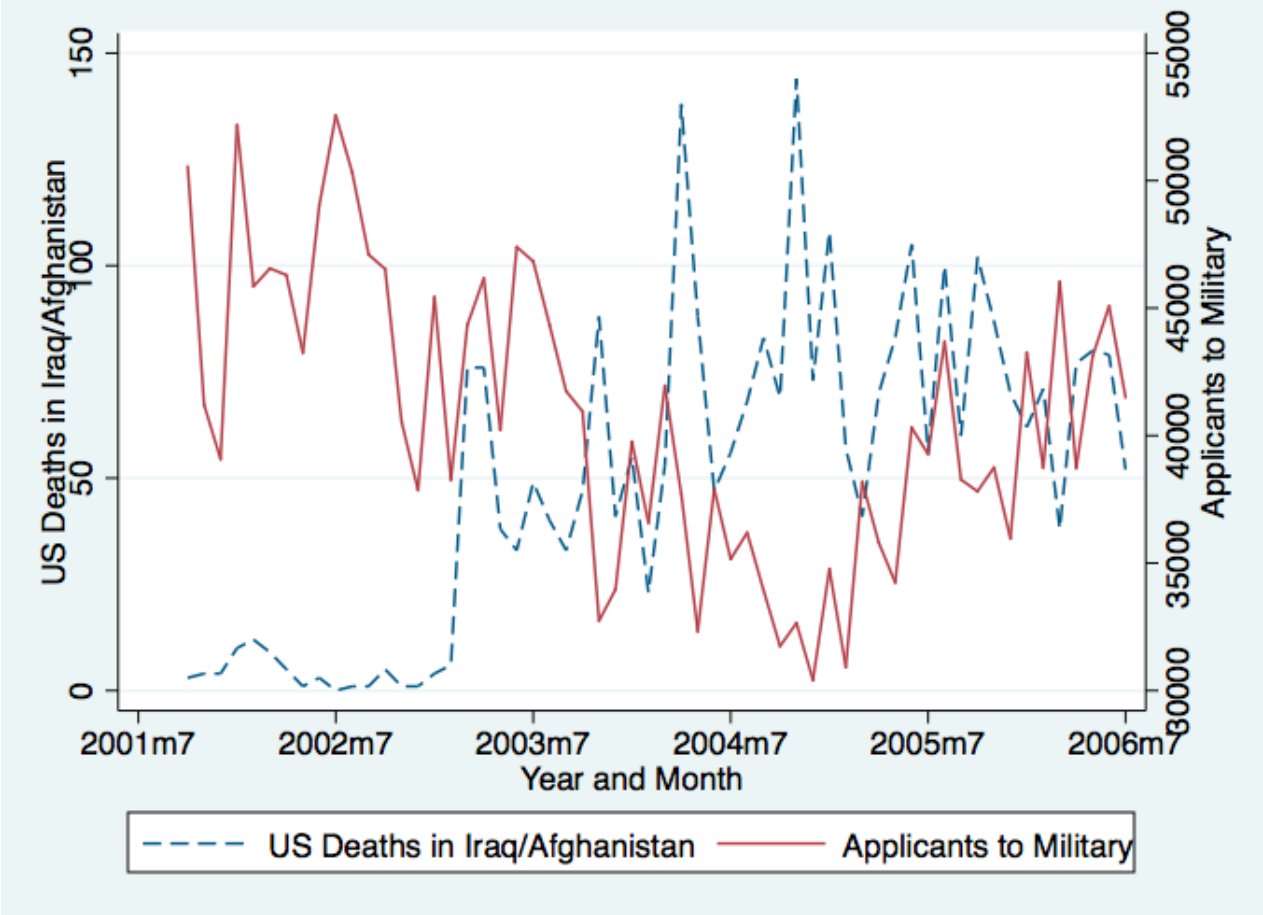
Figure 1.2: State Percentage of Recruits by State Percentage of Young Male Population
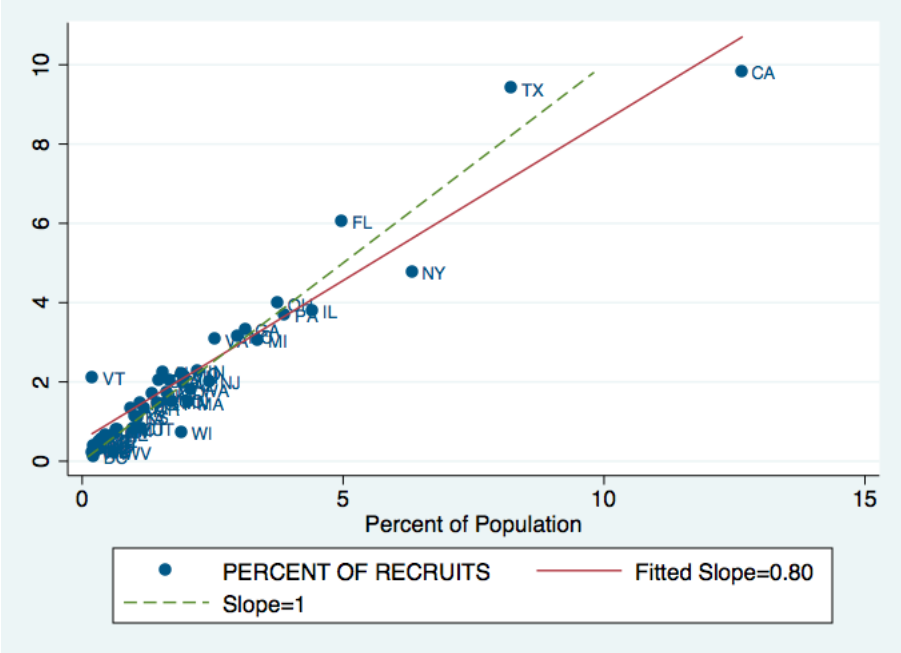
Figure 1.3: State Percentage of Population vs. State Percentage of Deaths
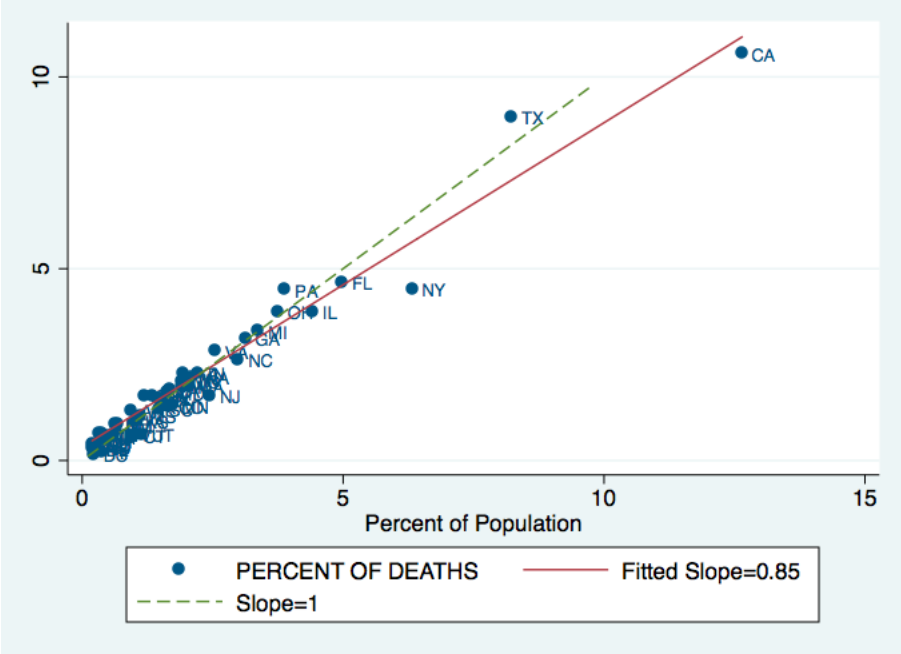
Figure 1.4: Total Applicants and Total Deaths
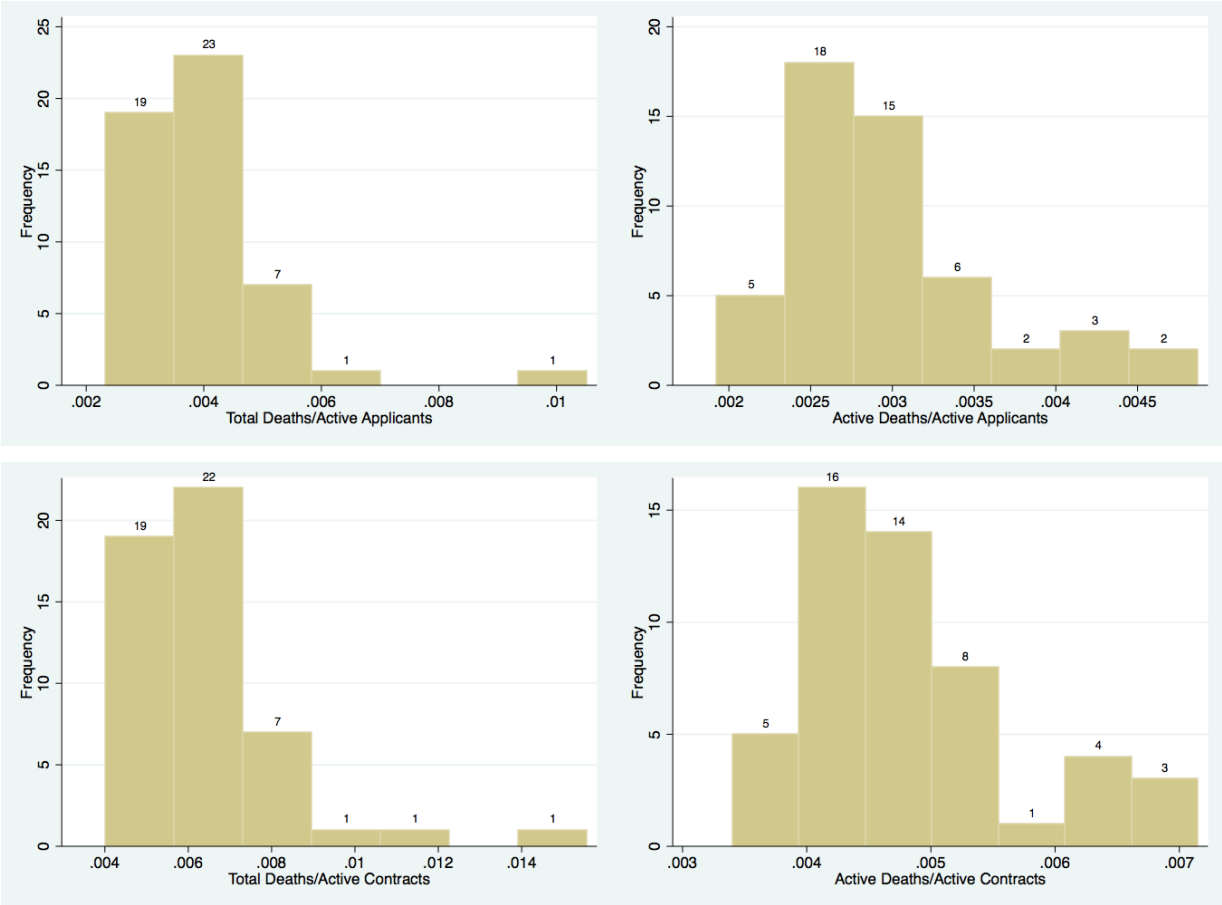
Figure 1.5: Histogram of P-Values Testing Whether State Deaths All Came from the Same Binomial Distribution
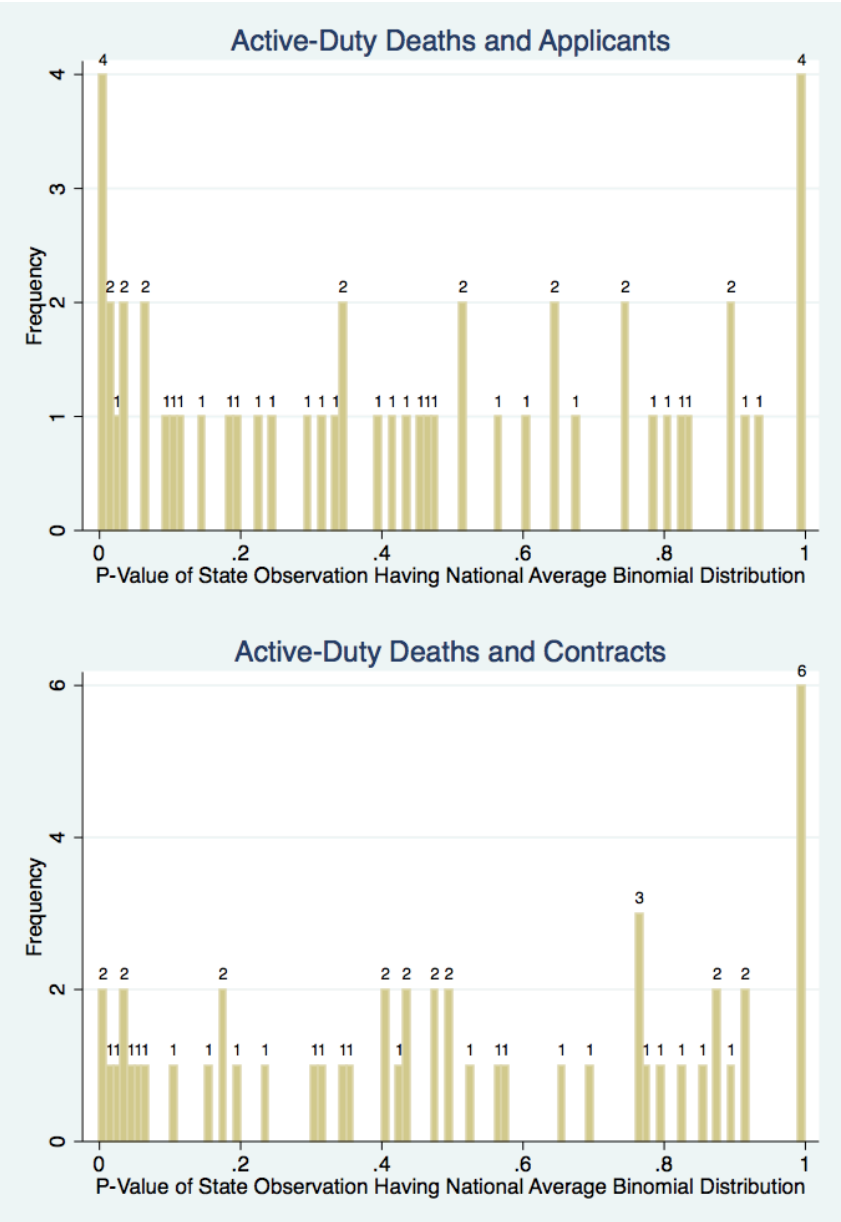
Table 1.1: Regression of Monthly Total National Recruits on Monthly Total Deaths

Regression of Monthly Total National Recruits on Monthly Total Deaths

	Applicants					
	(1)	(2)	(3)	(4)	(5)	(6)
Log National Deaths	-0.022***	-0.022***	-0.015***	-0.012***		
	[0.006]	[0.006]	[0.003]	[0.003]		
Log Lag National Deaths			-0.015***	-0.012***	-0.021***	-0.015**
			[0.004]	[0.004]	[0.006]	[0.006]
Time Trend	NO	NO	NO	YES	NO	YES
Observations	58	58	57	57	57	57
R-squared	0.214					

	Contracts					
	(7)	(8)	(9)	(10)	(11)	(12)
Log National Deaths	-0.027***	-0.027***	-0.019***	-0.013***		
	[0.007]	[0.006]	[0.003]	[0.003]		
Log Lag National Deaths			-0.020***	-0.014***	-0.027***	-0.016***
			[0.004]	[0.003]	[0.007]	[0.004]
Time Trend	NO	NO	NO	YES	NO	YES
Observations	58	58	57	57	57	57
R-squared	0.236					

Notes: Dependent variable is the log of the number of recruits, and right-hand side variables are lagged so coefficients can be interpreted as elasticities. Standard errors in brackets. Standard Errors are stand OLS standard errors in the first columns and Newey-West auto-correlation robust standard errors allowing for 4 lags in the rest of the specifications. Unit of observation is a calendar month at the national level. One month in the sample had zero deaths, so this is replaced with an arbitrarily small number for the sake of the logged (elasticity) estimates. Time Trend refers to inclusion of a linear time trend.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.2: Annual US Military Deaths

Annual US Military Deaths

Year	Miltary FTE	Deaths per 100K	Total Deaths	Hostile Action	Deaths in Iraq/Afghanistan in Data
1990	2,258,324	67	1507		
1991	2,198,189	81	1787	147	
1992	1,953,337	66	1293		
1993	1,849,537	66	1213		
1994	1,746,482	62	1075		
1995	1,661,928	63	1040		
1996	1,613,675	60	974	1	
1997	1,578,382	52	817		
1998	1,538,570	54	827		
1999	1,525,942	52	796		
2000	1,530,430	50	758		
2001	1,552,096	57	891	3	11
2002	1,627,142	61	999	18	49
2003	1,732,632	81	1410	339	531
2004	1,711,916	109	1873	738	897
2005	1,664,014	117	1940	739	939
2006	1,611,533	117	1882	769	915
2007	1,608,226	121	1953	847	1019
2008	1,683,144	85	1439	352	467
2009	1,640,751	92	1517	338	457

Notes: Data from http://siadapp.dmdc.osd.mil/personnel/CASUALTY/castop.htm

Table 1.3: Basic County-Level Least Squares Analysis of Deaths and Unemployment

Basic County-Level Least Squares Analysis of Deaths and Unemployment

VARIABLES	Applicants				Contracts			
	(1) Basic	(2) w/State	(3) w/State Trend	(4) w/State*Year FE	(5) Basic	(6) w/State	(7) w/State Trend	(8) w/State*Year FE
In-County Deaths	-14.551***	-13.845***	-12.174***	-12.553***	-9.298***	-8.867***	-7.835***	-8.132***
	[4.710]	[4.515]	[3.177]	[3.360]	[3.307]	[3.174]	[2.339]	[2.453]
Lag In-County Deaths	-17.691***	-16.601***	-14.790***	-15.344***	-10.591***	-9.947***	-8.827***	-9.165***
	[4.193]	[3.724]	[2.521]	[2.758]	[2.930]	[2.632]	[1.815]	[2.014]
Out-of-County Deaths		-0.231**	0.541	0.836		-0.175**	0.228	0.384
		[0.105]	[0.392]	[0.527]		[0.070]	[0.247]	[0.345]
Lag Out-of-County Deaths		-1.139**	-0.298*	-0.399***		-0.667*	-0.233**	-0.147
		[0.562]	[0.170]	[0.119]		[0.355]	[0.108]	[0.091]
County Unemployment		2.550***	2.471***	4.100***		1.928***	1.860***	2.333***
		[0.890]	[0.792]	[1.205]		[0.607]	[0.549]	[0.680]
State Unemployment		5.665**	-1.934*	0.079		2.871**	-1.238*	0.669
		[2.268]	[1.135]	[1.248]		[1.397]	[0.677]	[0.894]
Constant	91.650***	46.975***	89.360***	65.840***	53.118***	27.096***	50.180***	34.593***
	[1.921]	[14.041]	[3.921]	[7.561]	[1.273]	[9.122]	[2.668]	[5.747]
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES	YES	YES
State Time Trends	NO	NO	YES	NO	NO	NO	YES	NO
State*Year FE	NO	NO	NO	YES	NO	NO	NO	YES
Observations	178,809	178,739	178,739	178,739	178,809	178,739	178,739	178,739
R-squared	0.968	0.968	0.971	0.968	0.963	0.964	0.966	0.963

Notes: Dependent variable is the number of recruits in a county-month. Robust standard errors in brackets, clustered by county. Observations are weighted by county population. The left four columns are for applicants, the right are for recruits. Fixed effects are by county, monthly, linear state time trend, or interacted state*year.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.4: Least Squares Deaths and Unemployment as Semi-Elasticities

Least Squares Deaths and Unemployment as Semi-Elasticities

	Applicants				Contracts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Basic	w/State	w/State Trend	w/State*Year FE	Basic	w/State	w/State Trend	w/State*Year FE
In-County Deaths/100	-0.182	-0.312	-0.18	-1.071**	-0.1	-0.236	-0.153	-1.374***
	[0.408]	[0.404]	[0.305]	[0.420]	[0.387]	[0.364]	[0.264]	[0.435]
Lag In-County Deaths/100	-0.821***	-0.940***	-0.795***	-2.017***	-0.887***	-0.970***	-0.877***	-2.084***
	[0.185]	[0.196]	[0.268]	[0.335]	[0.240]	[0.244]	[0.301]	[0.359]
Out-of-County Deaths/100		0.194***	0.193***	0.039		0.194**	0.139*	-0.195**
		[0.067]	[0.067]	[0.074]		[0.075]	[0.076]	[0.093]
Lag Out-of-County Deaths/100		0.141*	0.148**	-0.773***		0.075	0.015	-0.654***
		[0.073]	[0.064]	[0.064]		[0.091]	[0.075]	[0.075]
County Unemployment		0.014**	0.012**	0.037***		0.016***	0.015***	0.034***
		[0.006]	[0.006]	[0.006]		[0.006]	[0.005]	[0.005]
State Unemployment		0.002	-0.023***	-0.008		-0.003	-0.024***	0.011
		[0.007]	[0.007]	[0.011]		[0.008]	[0.009]	[0.016]
Constant	3.488***	3.402***	3.548***	3.298***	3.015***	2.945***	3.066***	2.745***
	[0.009]	[0.039]	[0.044]	[0.068]	[0.011]	[0.044]	[0.049]	[0.090]
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES	YES	YES
State Time Trend	NO	NO	YES	NO	NO	NO	YES	NO
State*Year FE	NO	NO	NO	YES	NO	NO	NO	YES
Observations	137,595	137,541	137,541	137,541	124,334	124,281	124,281	124,281
R-squared	0.959	0.96	0.961	0.957	0.952	0.952	0.953	0.947

Notes: Dependent variable is the natural log of active duty recruits in a county month. Coefficients can be interpreted as semi-elasticities. Deaths on the right hand side have been divided by 100, so a -0.82 coefficient on the four death variables would imply a 0.82 percent (not 82 percent) decrease. Robust standard errors in brackets, clustered by county. Observations have been weighted by county populations.
*** p<0.01, ** p<0.05, * p<0.1

Table 1.5: Poisson Regression of County Recruits on Deaths and Unemployment

Poisson Regression of County Recruits on Deaths and Unemployment

VARIABLES	Applicants			Contracts		
	(1) Basic	(2) w/State	(3) w/State Trend	(4) Basic	(5) w/State	(6) w/State Trend
In-County Deaths/100	-0.045	-0.18	-0.089	0.027	-0.146	-0.086
	[0.436]	[0.433]	[0.320]	[0.448]	[0.431]	[0.310]
Lag In-County Deaths/100	-0.756***	-0.898***	-0.793***	-0.646**	-0.836***	-0.774***
	[0.184]	[0.195]	[0.244]	[0.279]	[0.273]	[0.259]
Out-of-County Deaths/100		0.218***	0.183***		0.230***	0.139**
		[0.064]	[0.062]		[0.070]	[0.071]
Lag Out-of-County Deaths/100		0.179***	0.155***		0.202***	0.103
		[0.066]	[0.058]		[0.078]	[0.064]
State Unemployment		-0.003	-0.025***		-0.01	-0.025***
		[0.006]	[0.006]		[0.007]	[0.006]
County Unemployment		0.016***	0.016***		0.016***	0.016***
		[0.005]	[0.004]		[0.005]	[0.004]
Population Offset	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES
State Time Trend	NO	NO	YES	NO	NO	YES
Observations	178,239	178,169	178,169	178,182	178,112	178,112
Number of Counties	3,127	3,127	3,127	3,126	3,126	3,126
Likelihood	-332831	-331962	-331200	-278656	-278173	-277676

Notes: Dependent variable is the number of recruits in a county-month. The right hand side variables represent deaths in a month divided by 100, so coefficients are interpretable as semi-elasticities, and .8 refers to 0.8 percent. Out of county deaths refers only to in-state but out-of-state deaths. Robust standard errors in brackets, which allow for clustering by county. All regressions include an offset of county young male population (using Stata's exposure option). Fixed effects included are for every county and month. State-specific time trends are also included.
*** p<0.01, ** p<0.05, * p<0.1

Table 1.6: Poisson Regression with Recruiter and Mortality Controls

Poisson Regression with Recruiter and Mortality Controls

VARIABLES	Applicants (1) Basic	(2) w/ Controls	Contracts (3) Basic	(4) w/ Controls
In-County Deaths/100	-0.494*	-0.484	-0.012	-0.075
	[0.297]	[0.314]	[0.308]	[0.315]
Lag In-County Deaths/100	-0.980***	-0.993***	-0.415	-0.468
	[0.373]	[0.362]	[0.410]	[0.400]
Out-of-County Deaths/100	0.088	0.156	0.034	0.103
	[0.095]	[0.099]	[0.091]	[0.091]
Lag Out-of-County Deaths/100	0.175*	0.207**	0.096	0.142
	[0.103]	[0.104]	[0.102]	[0.101]
State Unemployment	-0.038***	-0.038***	-0.026**	-0.025**
	[0.010]	[0.010]	[0.012]	[0.012]
County Unemployment	0.021***	0.021***	0.019***	0.019***
	[0.002]	[0.002]	[0.003]	[0.003]
Recruiters/100		0.033***		0.043***
		[0.011]		[0.012]
Lag State Mortality/100		0.037		-0.007
		[0.039]		[0.037]
Lag County Mortality/100		0.031**		0.021*
		[0.013]		[0.013]
Population Offset	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES
State Time Trend	YES	YES	YES	YES
Observations	97,353	97,353	97,274	97,274
Number of Counties	3,122	3,122	3,119	3,119
Likelihood	-180856	-180838	-151557	-151543

Notes: Dependent variable is the number of recruits in a county-month. The right hand side variables represent deaths in a month divided by 100, so coefficients are interpretable as semi-elasticities, and .8 refers to 0.8 percent. Out of county deaths refers only to in-state but out-of-state deaths. Robust standard errors in brackets, which allow for clustering by county. All regressions include an offset of county young male population (using the exposure option). Fixed effects included are for every county and month. State-specific time trends are also included in all specifications.
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.7: Poisson Regression of Media-Market and Contiguous County Deaths

Poisson Regression of Media-Market and Contiguous County Deaths

	Applicants				Contracts			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
VARIABLES	Neighboring		Media		Neighboring		Media	
In-County Deaths/100		-0.107		-0.039		-0.085		-0.029
		[0.342]		[0.323]		[0.322]		[0.307]
Lag In-County Deaths/100		-0.691***		-0.720***		-0.644***		-0.693***
		[0.222]		[0.247]		[0.241]		[0.267]
Contiguous County Deaths/100	0.191	0.227			0.085	0.118		
	[0.165]	[0.183]			[0.189]	[0.207]		
Lag Contiguous County Deaths/100	-0.154	-0.096			-0.207	-0.154		
	[0.216]	[0.220]			[0.212]	[0.212]		
Media Market Deaths/100			0.135	0.148			-0.014	-0.002
			[0.164]	[0.166]			[0.183]	[0.185]
Lag Media Market Deaths/100			-0.164	-0.137			-0.318	-0.292
			[0.183]	[0.183]			[0.206]	[0.204]
State Unemployment	-0.025***	-0.025***	-0.024***	-0.024***	-0.025***	-0.025***	-0.025***	-0.025***
	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]	[0.006]
County Unemployment	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***	0.016***
	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]	[0.004]
Population Offset	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES	YES	YES
State Time Trend	YES	YES	YES	YES	YES	YES	YES	YES
Observations	178,055	178,055	178,169	178,169	177,998	177,998	178,112	178,112
Number of Counties	3,125	3,125	3,127	3,127	3,124	3,124	3,126	3,126

Notes: Robust standard errors in brackets, clustered by county. Neighboring counties are defined using the ICPSR contiguous county file. Media Market deaths are defined using Nielsen Media Research's DMA, data courtesy of James Snyder. All regressions include an offset of county young male population (using Stata's exposure option). Fixed effects included are for every county and month. State-specific time trends are also included.

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.8: Poisson Regression with Multiple Interactions

Poisson Regression with Multiple Interactions

VARIABLES	Applicants		Contracts	
	(1)	(2)	(3)	(4)
	w/ %Black	w/ Racial Fractionalization	w/ %Black	w/ Racial Fractionalization
County Deaths/100	-0.227	-0.206	-0.205	-0.169
	[0.422]	[0.423]	[0.419]	[0.423]
Lag County Deaths/100	-0.739	-0.525	-0.509	-0.222
	[0.462]	[0.497]	[0.519]	[0.559]
Out of County Deaths/100	0.221***	0.221***	0.226***	0.225***
	[0.063]	[0.063]	[0.069]	[0.069]
Lag Out of County Deaths/100	0.172***	0.177***	0.184**	0.189**
	[0.066]	[0.066]	[0.079]	[0.078]
State Unemployment	-0.003	-0.003	-0.01	-0.01
	[0.006]	[0.006]	[0.006]	[0.006]
County Unemployment	0.016***	0.016***	0.015***	0.015***
	[0.005]	[0.005]	[0.005]	[0.005]
Death * County Unemployment	0.42	0.377	0.842**	0.779**
	[0.417]	[0.419]	[0.380]	[0.388]
Death / County Population	-2,085.880***	-2,215.198***	-1,509.78	-1,703.94
	[511.203]	[601.024]	[1,088.919]	[1,148.970]
Death * %Population Black	-0.061**		-0.100***	
	[0.030]		[0.038]	
Death * Racial Fractionalization		-2.306		-0.957
		[3.215]		[3.524]
Death * %Bush Vote '04	0.088***	0.088***	0.151***	0.166***
	[0.027]	[0.033]	[0.025]	[0.034]
Death * Rural	-4.139	-4.419	-5.998	-6.081
	[3.851]	[3.847]	[4.205]	[4.204]
Population Offset	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Observations	176,801	176,801	176,744	176,744
Number of Counties	3,103	3,103	3,102	3,102
Likelihood	-330331	-330334	-276845	-276850

Notes: Dependent variable is the number of recruits in a county-month. The right hand side variables represent deaths in a month divided by 100, so coefficients are interpretable as semi-elasticities, and .8 refers to 0.8 percent. Percent Bush vote and percent population Black are on 0-100 scale. Out of county deaths refers only to in-state but out-of-state deaths. Robust standard errors in brackets, which allow for clustering by county. All regressions include an offset of county population (using the exposure option). Fixed effects included are for every county and month.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.9: Poisson Regressions of Recruits by Service Branch

Poisson Regressions of Recruits by Service Branch

VARIABLES	Applicants				Contracts			
	(1) Army	(2) Air Force	(3) Marines	(4) Navy	(5) Army	(6) Air Force	(7) Marines	(8) Navy
In-County Deaths/100	0.549	-1.345**	-1.002	-0.382	0.79	-1.560**	-0.14	-0.712
	[0.405]	[0.649]	[0.616]	[0.609]	[0.514]	[0.755]	[0.516]	[0.582]
Lag In-County Deaths/100	-0.617*	-0.049	-2.528***	-0.637	-0.585	-0.936	-1.113*	-0.924
	[0.340]	[1.088]	[0.562]	[0.469]	[0.473]	[0.625]	[0.606]	[0.582]
Out-of-County Deaths/100	0.205**	0.181	0.026	0.230**	0.279*	0.011	0.025	0.12
	[0.104]	[0.135]	[0.124]	[0.108]	[0.143]	[0.146]	[0.140]	[0.120]
Lag Out-of-County Deaths/100	0.025	0.174	0.039	0.436***	0.055	0.224	-0.055	0.210*
	[0.094]	[0.142]	[0.127]	[0.121]	[0.110]	[0.166]	[0.122]	[0.119]
State Unemployment	-0.017*	-0.018	-0.026**	-0.039***	-0.033***	-0.003	-0.036***	-0.022**
	[0.009]	[0.011]	[0.011]	[0.010]	[0.011]	[0.012]	[0.011]	[0.011]
County Unemployment	0.014***	0.017***	0.020***	0.016***	0.019***	0.011*	0.020***	0.014***
	[0.004]	[0.006]	[0.005]	[0.006]	[0.005]	[0.006]	[0.005]	[0.005]
Population Offset	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES	YES	YES
State Time Trend	YES	YES	YES	YES	YES	YES	YES	YES
Observations	177,029	174,578	173,894	175,604	175,661	173,039	171,842	174,578
Number of Counties	3,107	3,064	3,052	3,082	3,083	3,037	3,016	3,064

Notes: Dependent variable is the number of active duty recruits to the service branch indicated at the top. Robust standard errors which allow for clustering by county are in brackets. All regressions include county young male population as offset (using exposure option). Monthly and county fixed effects as well as state time trends are included in all specifications.
*** p<0.01, ** p<0.05, * p<0.1

Table 1.10: Poisson Regressions of Recruits by Quality Level

Poisson Regressions of Recruits by Quality Level

VARIABLES	Applicants				Contracts			
	(1) Low	(2) High	(3) High-Alt	(4) Very High	(5) Low	(6) High	(7) High-Alt	(8) Very High
In-County Deaths/100	-0.03	-0.161	0.065	-0.341	-0.287	0.185	0.285	2.405*
	[0.318]	[0.480]	[0.432]	[0.983]	[0.313]	[0.406]	[0.365]	[1.347]
Lag In-County Deaths/100	-0.839***	-0.687*	-0.793***	-2.465***	-1.020***	-0.44	-0.867***	-2.491
	[0.289]	[0.370]	[0.277]	[0.834]	[0.359]	[0.289]	[0.278]	[1.615]
Out-of-County Deaths/100	0.238***	0.083	0.184**	0.323	0.171**	0.092	0.149*	0.412
	[0.077]	[0.092]	[0.077]	[0.295]	[0.087]	[0.103]	[0.089]	[0.372]
Lag Out-of-County Deaths/100	0.165**	0.136	0.285***	0.011	0.084	0.124	0.190**	0.024
	[0.074]	[0.084]	[0.073]	[0.260]	[0.094]	[0.095]	[0.080]	[0.360]
State Unemployment	-0.033***	-0.013	-0.007	-0.029	-0.033***	-0.017*	-0.013*	-0.045
	[0.006]	[0.008]	[0.007]	[0.022]	[0.007]	[0.009]	[0.007]	[0.028]
County Unemployment	0.012***	0.025***	0.014***	0.013	0.011***	0.025***	0.016***	0.023**
	[0.004]	[0.005]	[0.004]	[0.009]	[0.004]	[0.005]	[0.004]	[0.011]
Population Offset	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES	YES	YES
State Time Trend	YES	YES	YES	YES	YES	YES	YES	YES
Observations	178,055	176,573	177,884	140,207	177,713	175,889	177,998	131,999
Number of Counties	3,125	3,099	3,122	2,461	3,119	3,087	3,124	2,317

Notes: Dependent variable is the number of active duty recruits of the quality-level indicated at the top. Regression is Poisson, and deaths have been divided by 100, so coefficients can be interpreted as semi-elasticities. All regressions include county young male population as an offset (using exposure option). Low quality is a recruit who scored below 50 on the AFQT or does not have a high school diploma. High Quality have both a 50 or higher and a high school degree. High Quality-Alt have a 50+ and have either finished high school or are in their senior year. Very High is a 75+ AFQT and at least some college. Robust standard errors which allow for clustering by county are in brackets. Monthly and county fixed effects as well as state time trends are included in all specifications.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.11: Poisson Regressions of County-Month-Service Branch Recruits on Same Service and Other Service Deaths

Poisson Regressions of County-Month-Service Branch Recruits on Same Service and Other Service Deaths

| VARIABLES | Applicants | | Contracts | |
	(1) County*Unit FE	(2) w/ Trend	(3) County*Unit FE	(4) w/Trend
Same Service In-County Deaths/100	0.412	0.484	1.888***	1.975***
	[0.683]	[0.691]	[0.683]	[0.712]
Other Service In-County Deaths/100	-0.165	-0.085	-0.613	-0.563
	[0.522]	[0.486]	[0.599]	[0.546]
Lag Same Service In-County Deaths/100	-0.741	-0.667	-0.479	-0.403
	[0.679]	[0.697]	[0.810]	[0.850]
Lag Other Service In-County Deaths/100	-1.155***	-1.063***	-1.265**	-1.220***
	[0.409]	[0.388]	[0.492]	[0.444]
Same Service Out-of-County Deaths/100	0.519***	0.463***	0.688***	0.572***
	[0.148]	[0.133]	[0.157]	[0.144]
Other Service Out-of-County Deaths/100	0.013	-0.026	0.066	-0.026
	[0.086]	[0.084]	[0.089]	[0.092]
Lag Same Service Out-of-County Deaths/100	0.196	0.139	0.233	0.096
	[0.162]	[0.158]	[0.203]	[0.182]
Lag Other Service Out-of-County Deaths/100	0.160*	0.121	0.156*	0.05
	[0.088]	[0.083]	[0.090]	[0.089]
Population Offset	YES	YES	YES	YES
County & State Unemployment	YES	YES	YES	YES
County*Service Branch FE	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES
State Time Trend	NO	YES	NO	YES
Observations	701,105	701,105	695,120	695,120
Number of County*Service Branch	12,305	12,305	12,200	12,200
Likelihood	-784693	-783929	-634157	-633660
In-County Test	0.465	0.451	0.0054	0.00378
Lag In-County Test	0.605	0.606	0.428	0.394
Out-of-County Test	0.00528	0.00434	0.00122	0.000942
Lag Out-of-County Test	0.855	0.924	0.74	0.834

Notes: Dependent variable is the number of recruits in a given county-month-service branch. Robust standard errors which allow for clustering at the county-service branch level are in brackets. All regressions include an offset of young male county population. County and State Unemployment, County*Service Branch fixed effects, monthly fixed effects, and in certain specifications, state time trends are all included. Beneath the regression are p-values of tests that the Same-Service coefficient is equal to the corresponding Other-Service coefficient.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.12: Poisson Regression of County Recruits and Deaths by Service Branch

Poisson Regression of County Recruits and Deaths by Service Branch

VARIABLES	Applicants		Contracts	
	(1) County	(2) State	(3) County	(4) State
In-county Deaths/100	-0.086	-0.084	-0.091	-0.088
	[0.317]	[0.313]	[0.302]	[0.302]
Lag In-County Army Deaths/100	-0.926**	-0.983**	-1.094*	-1.130**
	[0.453]	[0.439]	[0.582]	[0.568]
Lag In-County Air Force Deaths/100	2.948	2.572	2.726	2.233
	[2.632]	[2.516]	[2.980]	[2.958]
Lag In-County Marine Corps Deaths/100	-1.140**	-1.337***	-0.922**	-1.096***
	[0.473]	[0.456]	[0.396]	[0.413]
Lag In-County Navy Deaths/100	0.692	0.623	0.281	0.211
	[1.173]	[1.185]	[1.783]	[1.815]
Out-of-County Deaths/100	0.186***	0.186***	0.142**	0.145**
	[0.062]	[0.062]	[0.070]	[0.071]
Lag Out-of-County Total Deaths/100	0.155***		0.102	
	[0.058]		[0.064]	
Lag Out-of-County Army Deaths/100		0.103		0.014
		[0.105]		[0.111]
Lag Out-of-County Air Force Deaths/100		-0.657		-0.892
		[0.620]		[0.638]
Lag Out-of-County Marine Corps Deaths/100		0.315***		0.229**
		[0.093]		[0.113]
Lag Out-of-County Navy Deaths/100		1.313***		1.378***
		[0.450]		[0.479]
State Unemployment	-0.024***	-0.025***	-0.025***	-0.025***
	[0.006]	[0.005]	[0.006]	[0.006]
County Unemployment	0.016***	0.016***	0.016***	0.016***
	[0.004]	[0.004]	[0.004]	[0.004]
Population Offset	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES
State Time Trends	YES	YES	YES	YES
Observations	178,169	178,169	178,112	178,112
Number of Counties	3,127	3,127	3,126	3,126
Test Lag County Deaths	0.222	0.207	0.563	0.626
Test Lag State		0.0382		0.021

Notes: Dependent variable is the number of active duty recruits in a county-month. Robust standard errors in brackets, clustered by county. Regressions include county young male population as an offset (using exposure option). County and monthly fixed effects are included, as well as state time trends. Beneath the regression are the p-values from tests of equality of the coefficients for the deaths of all four service branches, separately by state and county.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.13: Poisson Regression of County Recruits with Deaths by Specific War

Poisson Regression of County Recruits with Deaths by Specific War

	Applicants		**Contracts**	
	(1)	(2)	(3)	(4)
VARIABLES	County	State	County	State
In-County Deaths/100	-0.345	-0.334	-0.683**	-0.664**
	[0.292]	[0.291]	[0.331]	[0.335]
Iraq Lag In-County Deaths/100	-1.633***	-1.589***	-1.709***	-1.626***
	[0.406]	[0.389]	[0.451]	[0.433]
Afghanistan Lag In-County Deaths/100	2.416*	2.298*	2.926**	2.699**
	[1.370]	[1.394]	[1.205]	[1.184]
Out-of-County Deaths/100	-0.043	-0.026	-0.285***	-0.253***
	[0.065]	[0.065]	[0.071]	[0.072]
Lag Out-of-County Deaths/100	-0.489***		-0.485***	
	[0.056]		[0.064]	
Iraq Lag Out-of-County Deaths/100		-0.527***		-0.564***
		[0.058]		[0.066]
Afghanistan Lag Out-of-County Deaths/100		0.421*		1.115***
		[0.232]		[0.254]
State Unemployment	-0.075***	-0.075***	-0.073***	-0.073***
	[0.006]	[0.006]	[0.006]	[0.006]
County Unemployment	0.018***	0.018***	0.018***	0.019***
	[0.004]	[0.004]	[0.004]	[0.004]
Population Offset	YES	YES	YES	YES
County FE	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES
State Time Trends	YES	YES	YES	YES
Observations	178,169	178,169	178,112	178,112
Number of fips	3,127	3,127	3,126	3,126
Likelihood	-334757	-334742	-281121	-281091
Test Lag County Deaths	0.003	0.006	0.001	0.001
Test Lag State Deaths		0.000		0.000

Notes: Dependent variable is the number of active duty recruits in a county-month. Robust standard errors in brackets, clustered by county. Regressions include county young male population as an offset (using exposure option). County and monthly fixed effects are included, as well as state time trends. Beneath the regression are the p-values from tests of equality of the coefficients for the deaths of each of the two wars, separately by state and county.

*** p<0.01, ** p<0.05, * p<0.1

Table 1.14: Poisson Regressions of Recruits on Cumulative Lags of Deaths and Unemployment

Poisson Regressions of Recruits on Cumulative Lags of Deaths and Unemployment

VARIABLES	Applicants				Contracts			
	(1) Two Lags	(2) Four Lags	(3) Six Lags	(4) Twelve Lags	(5) Two Lags	(6) Four Lags	(7) Six Lags	(8) Twelve Lags
In-County Deaths/100	-0.490***	-0.423***	-0.377***	-0.163*	-0.368*	-0.362**	-0.403***	-0.373***
	[0.177]	[0.161]	[0.126]	[0.097]	[0.206]	[0.167]	[0.127]	[0.092]
Out-of-County Deaths/100	0.122***	0.089**	0.102***	0.087***	0.073*	0.068*	0.096***	0.100***
	[0.038]	[0.038]	[0.032]	[0.025]	[0.041]	[0.038]	[0.032]	[0.027]
State Unemployment	-0.007***	-0.002	0.001	-0.001	-0.007***	-0.005**	-0.002	-0.002**
	[0.002]	[0.002]	[0.001]	[0.001]	[0.002]	[0.002]	[0.001]	[0.001]
County Unemployment	0.004***	0.001	0	0.002*	0.005***	0.002*	0.001	0.002**
	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]	[0.001]
Population Offset	YES	YES	YES	YES	YES	YES	YES	YES
County FE	YES	YES	YES	YES	YES	YES	YES	YES
Monthly FE	YES	YES	YES	YES	YES	YES	YES	YES
State Time Trends	YES	YES	YES	YES	YES	YES	YES	YES
Observations	175,035	168,727	162,475	143,581	174,979	168,727	162,423	143,489
Number of Counties	3,127	3,126	3,126	3,123	3,126	3,126	3,125	3,121

Notes: Dependent variable is the number of recruits in a county month. Right hand side is the number of deaths (divided by 100) in the current month plus previous two, four, six, or twelve months. State and county unemployment are also cumulative in the same manner. Robust standard errors in brackets which allow for clustering by county. All regressions include an offset of county population (using the exposure option).
*** p<0.01, ** p<0.05, * p<0.1

Chapter 2

Baseball Awards and Subsequent Performance

2.1   Introduction

Performance of Major League Baseball players after they win awards can be used to shed light on a more general economic question of how rational agents perform after they have been rewarded for good behavior up to that point. There are many articles in the popular media concerning the "curses" of major awards or forms of recognition. Perhaps the most famous of these is the *Sports Illustrated* cover curse, which earned its own self-referential *Sports Illustrated* article with a cover photo of a black cat in 2002[Wolff, 2002]. There has also been discussion of sophomore slump curses for winners of the Rookie of the Year, curses from winning the Heisman trophy, appearing on a Wheaties cereal box, or appearing on the box-cover of EA Sports' Madden NFL video game. Any good sports fan can name several instances of these curses affecting their favorite teams, while the reaction of most statisticians is likely to write off the curses as mean reversion. However, recent research has shown that the companies of awarding-winning CEO's do in fact underperform after winning awards from popular business publications[Malmendier and Tate, 2008]. Thus it seems worthwhile to rigorously evaluate whether declines in athletic performance after awards actually exist, and if they do, whether they are evidence of anything other than mean reversion.

To do this, I have used statistics from Major League Baseball to analyze performance of players after winning the most important awards: Most Valuable Player and Rookie of the Year for offensive players and the Cy Young award for pitching. With the longest history, best-kept, and most easily available statistics, Major League Baseball is an obvious source of data with which to attempt to analyze this question.

The rest of the paper is as follows. Section 2.2 reviews the relevant literature, section 2.3 summarizes the data used in this project, section 2.4 shows the statistical estimation completed up to this point, and section 2.5 discusses the results and directions for future

research.

2.2   Literature Review

This work is designed to shed light on not just baseball but also more general theoretical economic questions, and thus the theoretical research to which this paper speaks is well developed. The phenomenon of baseball players and their performance after external awards (and the obvious next question of how to design contracts for these players) is related to ideas from optimal contract theory. Some of the earliest work to address the issues of optimal contracts with monitoring costs and the further complication of non-separable team production and "team spirit and loyalty" (their term not mine, that although obviously not intended for sports analysis, seems relevant) was Alchian and Demsetz [1972]. Although player performance is clearly well defined and observable, it is not completely clear that player effort is the same as player performance (although I assume they are strongly positively correlated), and managers may be unable to completely observe player effort. Bengt Holmstrom has done a great deal of work in the area of moral hazard and imperfect monitoring [1979], and moral hazard and team effort [1982]. In addition to being complicated by limited observability and team effort, baseball players are subject to repeat contract negotiation, which has been addressed in Rogerson [1985], and possible renegotiation of current contracts, addressed by Hermalin and Katz [1991] and Fudenberg and Tirole [1990]. Gibbons and Murphy [1992] addressed the issue of "career concerns—concerns about the effects of current performance on future compensation" which is relevant to baseball players whose current performance could possibly lock in a team's expectations about the player's future value.

Important theoretical and empirical work has been done on the existence of superstars in general. The influential initial theoretical explanation for the existence of superstars (where "relatively small numbers of people earn enormous amounts of money and dominate the activities in which they engage") is that of Rosen [1981]. Empirical papers have also shown that superstars in sports generate a positive revenue externality for teams other than their own (Michael Jordan increased other teams' revenue by $53 million in the 1991-92 season) [Hausman and Leonard, 1997] and that competing against superstars may lead good non-superstar players to decrease their level of effort (good golfers do worse when playing in the same tournament as Tiger Woods) [Brown, 2007]. Hausman and Leonard [1997] may be especially relevant in that the authors use their findings to analyze the efficiency of a player salary cap, something not present in Major League Baseball but often discussed, and Brown [2007] is especially relevant in that baseball players may suffer the same decrease in performance when playing against (or with?) particularly good players.

There are also a handful of empirical papers that have used the same sort of design that I have used for my analysis. The most similar is Malmendier and Tate [2008], which compares company performance of *Businessweek* magazine (and other publications) award-winning CEO's to that of predicted winners–firms that were to a large degree observably similar

to the winners at the time of the award. The advantage that baseball awards provides is clear knowledge of the runners-up. In addition to using statistical models to predict the runners-up, the Baseball Writers Association of America's voting (and thus the runners-up for baseball's major awards) is observable by the econometrician. Malmendier and Tate find that winning CEO's do in fact see larger declines in performance, perhaps because they spend more time on outside activities such as sitting on outside boards and writing books. Despite this, they observe an increase in pay. They also observe that the effects are stronger in firms with weak corporate governance.

Another empirical example of this "superstar" methodology is Greenstone et al. [2008], in which the authors use data from a corporate real-estate magazine called *Site Selection* to evaluate what happens to the economies of 82 counties that were selected as the home of a company's new factory, and compare them to the 129 counties that were finalists but ultimately not selected to have the new factory. They find that the economies of the winning towns improve relative to the losers, and that towns generally do not suffer from giving tax-breaks or other concessions designed to lure the new factories. Another similar paper is that of Jacob and Lefgren [2007], which compares the publication records of winners of NSF grants to those of non-winners. This is a more straightforward regression discontinuity design instead of matching, however I mention it because another superstar-matching type design would be similar—looking at the publication records of Clark Medal winning economists compared to those of the economists considered to be the runners-up.

Baseball has had number-crunching fans for its entire history, as anyone who has read Michael Lewis' best-selling book *Moneyball* can attest [2004]. Bill James and legions of "sabermetricians" have analyzed endless aspects of the game, and they have had a significant impact on the way the game is played–most noticeably causing an increase in the amount of attention being paid to a player's on-base percentage (and his number of walks). As important as their work has been to the game, I build on their work in that I include on-base percentage and slugging percentage (the ingredients in James' OPS–on-base plus slugging) in my evaluations of performance, but I make no claim that my work will be important in the course of game-play itself. Rather, it might be useful in the course of baseball contract negotiations and hopefully contract design in general. The most similar work by sabermetricians is likely Nate Silver (of recent FiveThirtyEight.com fame) and his PECOTA (Player Empirical Comparison and Optimization Test Algorithm, but also named after journeyman professional player Bill Pecota)[1] system, which uses a proprietary nearest neighbor matching algorithm to predict the performance of every major league baseball player each year.

The most similar baseball-related work I have found by academic econometricians are Krautmann and Solow [2009], which looks at player effort (measured by comparing player performance to average performance of players of equal age) over the course of a contract,

---

[1]Description retrieved from http://en.wikipedia.org/wiki/PECOTA May, 2009; Actual Pecota statistics available at http://baseballprospectus.com/pecota/

both when a player might be trying harder (the year before he goes on the market) and when a player might be slacking off (after he has signed a lucrative guaranteed contract). Using data from 550 or so contracts, they find that older players who do not expect to sign a new contract see significant decreases in performance, but this decline is offset for players who do expect to sign a new contract. Gathering this salary data, is one of the areas in which significant research progress could be made. Another short baseball paper [Smith and Schall, 2000] regarding regression towards the mean simply uses normalized player statistics and correlation coefficients to show that most players do in-fact regress towards the mean in any given year, and no special attention is paid towards award winners.

My empirical methods consist of ordinary least squares, a simple method of comparison similar to differences in differences, and matching estimation. The matching estimator used is described in Abadie et al. [2004].

2.3 Data

Data used in this paper is from Sean Lahman's Baseball Archive dataset, consisting of the complete set of statistics of all Major League Baseball players from 1871 through 2007[2]. We are interested in the largest awards: Most Valuable Player (MVP) and Rookie of the Year for position players and Cy Young for pitchers. The MVP award was first given from 1911 through 1914, then cancelled and given again in 1922 through 1929, and has been given by the Baseball Writer's Association of America (BWAA) in nearly the same form since 1931, with the exact same vote-tallying algorithm since 1938.

The Rookie of the Year award was given in a lesser form from 1940 through 1948 and in its present from from 1949 to the present. These awards are both voted on by a subset of the greater than 140 BWAA members and are given to one player from each league (American, National) each year. Ties have occurred only once for each award, both in 1979. From 1938 to the present, writers have voted for the their top 10 choices for the MVP (with the first-place vote worth 14 points and second through tenth place worth 9 points on down to 1 point.) Writers voted for just one player from each league for Rookie of the Year until 1980, when voters began allocating three votes worth 5, 3, and 1 points each.

Generally the MVP and Rookie of the Year awards are given to position players as opposed to pitchers, because pitchers play in fewer games and have several pitcher-specific awards all to themselves, but pitchers do occasionally still win these awards. Of the just over 100 MVP awards that have been given since the Cy Young award was created explicitly for pitchers in 1956, eight have gone to pitchers; the last time the award went to a pitcher was in 1992 to Oakland Athletics pitcher Dennis Eckersly. Roughly one quarter of the Rookies of the Year have been pitchers.

From 1956 through 1966, the Cy Young award was given to the best pitcher in the major leagues. Starting in 1967, the award has been given to the best pitcher in both the American

---

[2]Retrieved from http://www.baseball1.com September 16, 2008

and National leagues. From 1956 through 1969, each voter voted for just one pitcher, but since 1970 voters have had three votes, with point weighting the same as that for Rookie of the Year votes. The Cy Young award poses less of a difficulty in winner similarity than do the MVP and Rookie of the Year awards since all winners are pitchers. However, one could reasonably assert that there are important differences between starting and relief pitchers. Of the 94 Cy Young winners in the dataset, 9 have 21 or more saves, indicating relief pitching, while the remaining 85 have four or fewer saves, indicating likely starting pitchers. For the analysis thus far, I have chosen to treat starting and relief pitchers equally and leave both in the database.

Since it is difficult, if not impossible, to compare the performance of a pitcher to that of a position player, years in which a pitcher won the award are not included in many of my specifications. In section 2.4.3 when I match based on BWAA voting, runners-up who happen to be pitchers are replaced with the next runner-up who is a position player, and years in which pitchers win the award are dropped entirely.

2.4   Estimation

There are several ways to test if there is any effect of winning any of these awards. Obviously we cannot observe both a player winning the award and the counterfactual of the same player not-winning the award in the same period, and there is no conceivable way to effectively run a randomized controlled experiment with winning the award as treatment, especially given that randomly having the award bestowed upon you out of a large group of good players would be a fundamentally different experience than being told you were in fact considered the single best player in a given year. So cleanly identified causal estimates might be slightly beyond the grasp of this data, but there are several methods through which we can see consistent suggestive evidence in favor of the idea that there is no negative effect of winning an award.

First, I look at a given player's performance over time and compare his performance after the award to performance before the award. I compare the winning player to the runner-up using several different methods, differences in differences first. Then I match winners to their runners-up according to BWAA voting and compute estimates with Imbens' (2001) matching estimator. Unfortunately (for the sake of statistical comparison) winners and runners-up are not statistically identical according to several important observable measures of performance, thus one could safely assume they differ on unobservable and thus confounding measures as well, leading to biased results. However, we also run a logit model to predict the likelihood of winning an award, and find much closer matches based on the predicted value of these regressions. In addition to matching estimates, I use show estimates from differences in differences methods, and I also look for effects of awards on the performance of winners' teams and teammates.

2.4.1 Player Performance Over Time

We can investigate the effect on performance of winning an award, if any, by looking at a single player's performance over time. We clearly should not interpret estimates of a single player's performance before he won an award as a perfect counterfactual for his not having won the award, but Figure 2.1 seems to suggest graphically that for MVP award winners, although performance does decrease after a player wins an award, excluding the award-winning year, performance before and after awards are very nearly identical. At least graphically, it clearly seems as if players, both winners of the award and the runners-up, have an unusually good year during the year they won or were closely considered for the award, and then drop off to the same level of performance they had in the years prior to the award.

Figure 2.2 shows the same comparison for Rookie of the Year winners. Winners and runners-up show a different pattern: post-award performance is indeed lower than award-year performance, but post-award performance is far better than pre-award performance. This might be expected due to the word "Rookie" which has the following official definition according to Major League Baseball:

> A player shall be considered a rookie unless, during a previous season or seasons, he has (a) exceeded 130 at-bats or 50 innings pitched in the Major Leagues; or (b) accumulated more than 45 days on the active roster of a Major League club or clubs during the period of 25-player limit (excluding time in the military service and time on the disabled list).[3]

In other words, the sample size of performance one year before the award is small by definition, and we do not show performance from two and three years before the award because of even smaller (sometimes zero) sample sizes.

Figure 2.3 shows the analogous graphs for winners and runners-up for the Cy Young award. Earned run average (ERA), strike-outs, and wins and losses are used to analyze performance. The graphs seem to indicate that after improvement in performance the year of the award, winners and runners-up perform slightly worse (i.e., they exhibit continued declines in performance) two to three years after the award in terms of ERA and strike-outs, but not in terms of wins and losses.

To analyze pre and post-award behavior in a slightly more rigorous manner, we can run t-tests comparing the average performance of winners before and after winning. Table 2.1 shows the results comparing one year after to one year before, two years after to two years before, and three years after to three years before for MVP's and Cy Young winners and one year after to one year before for Rookies of the Year. The estimates are almost uniformly small (less than one percent of the average MVP's performance statistic) and insignificant for MVP's and Cy Young winners, with only one of the twenty seven estimates (MVP On-base percentage three years before and after) showing significance at standard levels.

---

[3]http://mlb.mlb.com/mlb/official_info/about_mlb/rules_regulations.jsp retrieved 3/13/09.

Estimates for Rookies of the Year are large and statistically significant. Again, only comparisons for one year before and after are shown. One facet of the Rookie of the Year numbers is that although all the p-values are small, the p-values for Home Runs and RBI's are many orders of magnitude smaller than those for batting average, on-base percentage, and slugging, which makes sense given that the former group are cumulative and the latter are percentages. For example, if a player was called up to the Majors near the end of the 2005 season and had only 20 total at-bats, but had done very well, his batting average would be indistinguishable from players of a similar caliber, but his season home runs total would be very low. Normalizing by dividing by number of at-bats and testing the difference in means by regressing home run per at-bat and RBI per at-bat gives p-values of .008 and .122, respectively, closer to those of average-based statistics.

2.4.2 Differences in Differences

In addition to the matching estimates shown below in sections 2.4.3 and 2.4.4, another way to compare winners and runners-up is to look at the changes from year to year by constructing a differences in differences estimator. Table 2.2 shows these estimates for the three major awards for both voted and predicted runners-up. The center column is the major analysis of interest, showing the change in performance for award winners from the award year to the next minus the same change for runners up. The estimates seem to indicate that award winners see smaller drops in performance than runners-up.

2.4.3 Matching by BWAA Voting

For the larger baseball awards (MVP, Rookie of the Year, Cy Young) we can directly observe the complete list of both winners and runners-up. If we wanted to view winning an award as an exogenous treatment, then the winner and the runner-up according to writer voting would be statistically similar using observable measures of performance, but in our case this is not so. In a given year, anywhere from 16 to 38 players have received MVP votes, and the average winner has received 84.93% of the maximum possible vote point score, while the runner-up has received 61.51%. The gap between the recipient of the first and second-most points is larger than the gap between the recipient of the second and third-most points, and the gap continues to decrease as the vote rank decreases.

Table 2.3 shows the differences more clearly, showing the average of numerous performance statistics for award winners, first runner-up, first and second runners-up, and all non-winning players, separately for MVP, Rookie of the Year, and Cy Young winners along with p-values for t-tests of equality between winners and each of the other groups. For both awards, roughly half of the performance statistics are significantly different for winners and first runners-up. Clearly, award winners outperform runners-up, "treatment" with the award is not randomly assigned, and regressions or methods of estimation that treat the award as if it were exogenous are likely to lead to biased results.

For the sake of completeness, these (biased) results are shown for MVP, Rookie of the Year, and Cy Young winners in Table 2.4. These estimates are the associated differences in performance when MVP and Rookie of the Year winners are matched with the runner-up from their specific award using Abadie and Imbens' matching command. As one might expect based on the simple differences between winners and runners-up shown in Table 2.3, there are significant differences between MVP's and runners-up in the year of the award for most of the performance measures shown. One would struggle to tell a convincing story about strong effects of awards based on these estimates; they are almost all positive, both before and after the award, and there is not strong significance. They appear slightly stronger for the pitchers, as Cy Young winners seem to significantly outperform runners-up after the award in terms of the major performance measures Earned Run Average, Wins, and Strike-outs, but it would be incorrect to interpret this as a causal effect of the award itself.

Table 2.5 is similar to Table 2.4, except that the estimates are calculated using OLS–the sample is limited to winners and voted runners-up. The estimates shown are the coefficients on a binary variable associated with having won the award. I include these estimates partly as a comparison against those in Table 2.4, and because OLS is *far* less processor intensive than the matching estimator, it's helpful to have this quicker method available. The OLS estimates, which are basically the numerical form of Figures 2.1, 2.2, and 2.3, are far larger and far more significant than the matching estimates, indicating that the average award winner performs significantly better than the average voted runner up.

2.4.4   Matching by Performance

In addition to matching by BWAA voting, which limits us to pairing each winner with the second best person from their same league and their same year, we can match across years based on a statistical measure of performance that we create. To do this, we estimate a logit model that predicts winning awards based on the statistics that one might a priori assume are factors in determining awards: games played, at-bats, batting average, runs, home-runs, runs batted in, on-base percentage, slugging percentage, stolen bases, switch-hitting status, team wins, age, American nationality (there is as yet no race variable in the data), and tenure in the major leagues for MVP and Rookies of the Year. For the Cy Young award for pitchers I match based on earned run average, wins, losses, games, strike outs, complete games, shutouts, homeruns, bases on balls, and saves. With 45,777 observations for position players and 35,202 for pitchers, these models explains nearly half the variation for the MVP and Rookie of the Year, and over half (58%) for the Cy Young.

Figures 2.4, 2.5, and 2.6 show the analogous graphs as those in Figures 2.1, 2.2, and 2.3, except this time done with the winners and the logit-predicted runners-up instead of the BWAA-voted runners-up. In general it appears that the predicted runners-up are closer matches in absolute terms than the voted runners-up, although the trends and kinks are not as closely aligned as those for winners and voted runners-up. Table 2.6 shows the means from the actual winners and the predicted runners-up, where predicted runner-up is

defined as the non-winning player-year with the nearest likelihood of winning for each winner. The same non-winner can be selected as the match for more than one winner. Along most measures of performace, the winners and predicted winners are statistically indistinguishable at standard confidence levels. This fact is unfortunately somewhat dependent on the exact variables included in the predictive logit regression, especially so for the MVP, but less so for the Rookie of the Year and Cy Young.

Using the same group of observable characteristics from the logit estimation, Table 2.7 shows estimates using Abadie and Imbens' bias-adjusted matching estimator. Estimates are almost all very small, positive, and statistically indistinguishable from zero. This seems to be evidence that, as far as our small sample can detect, there is no negative effect from winning major awards. Table 2.8 shows estimates similar to those in Table 2.7, except the estimates are calculated using OLS, limiting the sample to award winners and the logit-predicted runners-up. The OLS estimates are shown as a comparison for the matching estimates, and are perhaps more broad than the matching estimators since they compare the average winner to the average runner-up, whereas the matching estimator compares specific winners to their specific runner-up.

Since my research design suffers from small sample sizes and thus struggles for enough statistical power, I have attempted to group awards into larger categories. The first is basically any offensive award: MVP, Rookie of the Year, ALCS MVP, All-Star MVP, Babe Ruth award, Hank Aaron award, Major League Player of the Year, NLCS MVP, Silver Slugger, Triple Crown or World Series MVP. The second is winners of either the MVP or Rookie of the Year, and the third is the combination of either the Cy Young pitching award or *The Sporting News* Pitcher of the Year award. Since voting for many of these awards is not public (not to mention combinations of the awards), I match using the same logit estimation as before. Table 2.9 shows the baseline comparison of winners and the predicted runners-up. Unfortunately for the sake of good counterfactuals, the winners and runners-up are significantly different across several important characteristics. Table 2.10 shows the regression estimates analogous to those in Table 2.8 except for the grouped awards. We can clearly see that for the position player awards, winners and predicted winners are observably different before, during, and after the award, while the pitching winners and predicted winners seem to not have large differences in the year of the award, and then they also do not exhibit large differences before and after the award. The Abadie and Imbens matching estimators for these grouped awards are currently being calculated.

2.4.5 Team Performance

Baseball, although not quite as tightly intertwined a team sport as basketball, is a team sport nevertheless. Due to the team nature of the sport, baseball provides us with an opportunity to look at not only the performance of the award-winner himself, but the performance of the award-winners' team as well. Table 2.11 shows team statistics for winners and both voted and predicted runner-ups' teams before, during, and in the year after an award. (Comparisons

with voted runners-up are in the top half of the table, comparisons with predicted runners-up are in the bottom half.) The results show that winners' teams tend to have better performance in the year of the award and compared to the voted runners-up, both before and after the award as well.

2.4.6 Teammate Performance

Because the baseball data is team data, we have a unique opportunity to analyze the effect of not just awards on the winners, but on the teammates of the winners as well. There are a few reasons why we might think that award-winning players (or just good players in general) might have effects on their teammates. For certain statistics, it is obvious mathematically that having a good teammate could hurt your performance–most obviously in the case of Runs Batted In. If the clean-up hitter on a team hit a home run in every at bat, it would be impossible for the number five batter to drive in any runners other than himself. This could however, work in the opposite direction if the clean-up hitter always hit a single or reached first base on a walk he would likely be increasing the potential RBI's of the number 5 hitter. Since awards have tended to go to (base-clearing) power hitters of late, one might guess that the former factor outweighs the latter, but the exact empirical answer is unclear. A more subtle way in which superstars could affect their teammates is by being "pitched around." If a batter is good enough, he may receive an intentional walk–the pitcher gives up a base, often creating the opportunity for a force-out, but this is better than giving up a homerun, and the pitcher can focus his energy on the (assumed) weaker next batter. It is unclear which direction this would effect teammates, if at all.

Table 2.12 shows estimates of the differences in performance between the teammates of winners and the teammates of both voted and predicted runners-up. The results seem to indicate that teammates of actual winners are not significantly different than teammates of winners. In fact, of the 30 different estimates for effects on teammates' batting averages, the largest magnitude point estimate is only 0.004, with a very tight zero result. There is definitely no spillover effect in terms of award winners improving the individual performance of their teammates.

2.5 Discussion and Future Work

I think some of the most reliable evidence from this project comes from the simple pictures provided in Figures 2.1, 2.2, and 2.3. Winners and runners-up see significant positive shocks in performance in the year of the award, but before and after the award, winners perform no worse than they had before the award, and the same goes for runners-up. This is documented in more complicated statistical fashion in tables 2.1 and 2.2, but the story remains the same. My matching estimation (primarily tables 2.4 and 2.7) are a little disappointing—first of all the winners and runners-up are significantly different across important measures (tables 2.3

and 2.6), so it would be unwise to interpret the award as any sort of exogenous treatment, but even if I were to do so, according the matching estimation, the effect is not consistently statistically significantly different than zero. The team and teammate estimates shown in tables 2.11 and 2.12 show that team performance for winners is significantly better than runner-up team performance in the year of the award. It is unclear whether this is evidence of selection (the BWAA votes for players who happen to be on better teams) or evidence that MVP's improve team performance. There is little evidence that MVP's have any effect on the performance of their teammates.

Although many of these results are not flashy or counter-intuitive, I do believe they are of value. First, it seems that baseball is a world unlike that of CEO's. If a CEO becomes famous thanks to an award and then diverts effort away from firm-production and towards book-production, I may have evidence that baseball players do not face similar incentives. With so much popular media devoted to the non-sports related pursuits of many major sports stars (Shaquille O'Neal has released five rap CD's and appeared in six feature films, Ron Artest managed a girl band, Tony Romo started dating Jessica Simpson, went to Mexico with her on his bye week during the playoffs and lost the following week) there might be reason to believe that awards would in fact lead to newfound popularity and enable players to divert efforts to other pursuits. But it seems this is not generally the case (New York Yankees shortstop Derek Jeter seems to have dated an endless stream of super-models without it affecting his performance.) This could be for any number of reasons: perhaps baseball players are *already* famous and don't have to win awards to have adoring fans, or perhaps player effort and performance have a correlation of near one, and since performance is so clearly observable, baseball players must continue to exert maximum effort regardless of any level of fame.

This result of a very brief increase in performance in award years and no difference in performance before and after awards could be very useful in terms of contract negotiations. Managers should expect decreased performance in the year after the award, but they could reasonably expect performance to stay at that same slightly reduced level for at least a few years. Many contracts feature financial incentives for good performance; it seems obvious that after the fact increasing salary for winning an award is foolish because performance will almost certainly not stay that high for long.

Contracts are one of the major areas in which this line of research could be extended. The baseball data set that I have has a large amount of salary data, but as far as I can tell it is simply the amount each player was paid during the course of a year, and does not include when new contracts were signed or how much of the money was base salary and how much was incentive-dependent. However, I feel confident that researchers could use additional sources (the Internet, mostly) to gather news reports of many free-agent signings in baseball and use this data to see whether there is any relation between awards and salaries. For example, if I were able to model a player's salary as a function of their performance, it might be interesting to see if winning awards per se has some effect on salary. If winners and runners-up were observably similar across all measures of performance, I would have an ideal test for whether awards lead to increased salary. This might be worthwhile for managers if

award winners lead to increased attendance and thus increased ticket revenue, but not in terms of performance alone.

My research design also enables me to possibly shed light on fundamental issues related to matching. First of all, are there systematic differences between voted runners-up and predicted runners-up? If so, what does this say about research designs where the voted runners-up are not observed? Also, the superstar-matching design is perhaps in some ways less than ideal, because it is fundamentally a regression discontinuity-type design, only instead of looking at an arbitrary cutoff in the middle of the distribution where there is a lot of mass, since we are looking at superstars we are looking at the tails of the distribution. With less mass, the observations in the sample are likely to be farther apart. Observables and likely the unobservables are less similar. Even though in many years there is significant debate over who deserves to win the major awards, I have shown that MVP winners are in fact observably different from their runners-up. If we make certain assumptions about the distribution of player performance, perhaps standard errors need to be inflated by some function of distance from the mean of the distribution, or bounds could be placed around the estimates.

Another possible angle to pursue is the amount of regression towards the mean for second runners-up and third runners-up. If I assume that the runners-up feel no significant accomplishment from having been a runner-up, then the difference between the second and third players provides us just information on the likely declines in performance for good players, and to some degree eliminates the possibility of a selection problem involving the winners themselves.

In addition to possibly doing work on the superstar matching research design itself, there are a handful of other situations I have discovered that might prove interesting. The Emmy Awards, having begun in the 1950's, nearly the same time as Nielsen ratings data came into being, provide another setting where the nominees are known. It could be the case that after an actor or writer wins a major Emmy Award, they divert attention from TV-show production towards feature film production (largely viewed as more desirable). This could be measured using the ratings of shows before and after awards, as well as the percentage of a season's episodes in which an actor or writer is involved, as well as the number of feature films in which someone is involved. Perhaps Emmy-winning actors save their best effort for films and the TV show's ratings decline, or perhaps their fame helps attract new viewers to the show.

Figure 2.1: MVP Winner and Runner-up Performance over Time

Figure 2.2: Rookie of the Year Winner & Runner-up Performance over Time
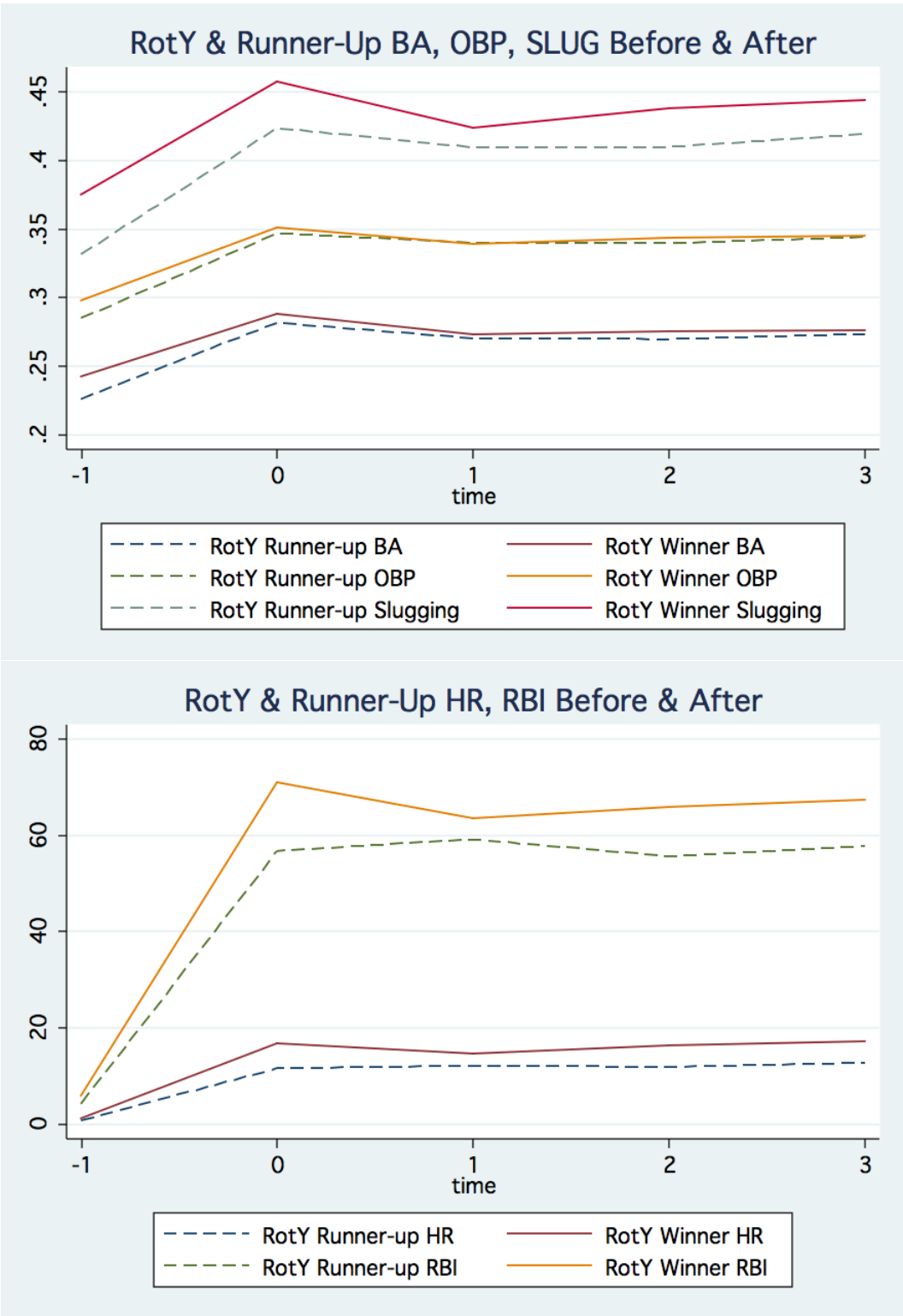
Figure 2.3: Cy Young Winner & Runner-Up Performance Over Time
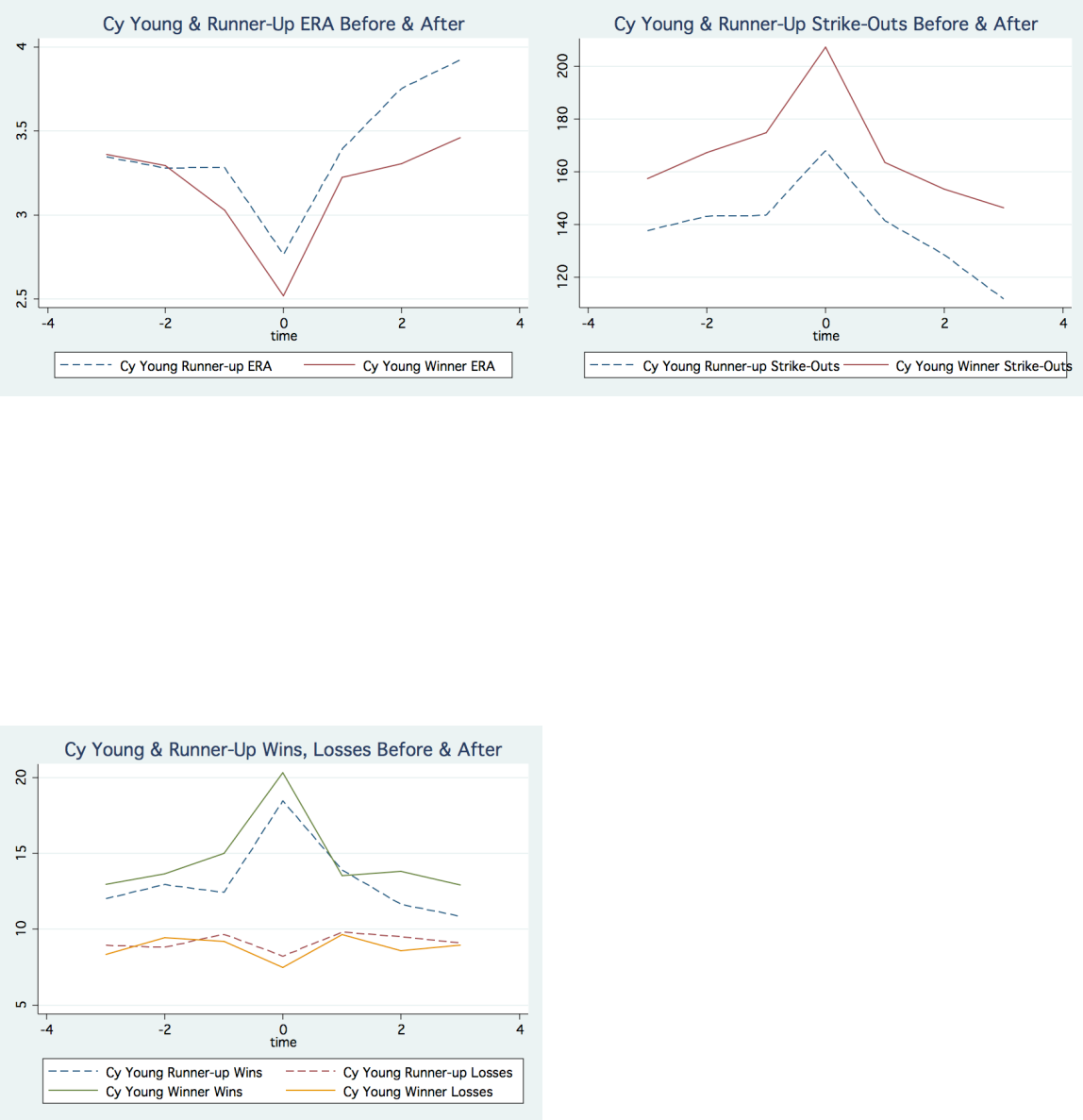
Table 2.1: Award Winner Differences in Performance Before & After Awards

Award		One Year	p-value	Two Years	p-value	Three Years	p-value
MVP	Batting Average	-0.001	0.803	-0.003	0.499	0.001	0.889
	Home Runs	0.884	0.596	-0.104	0.949	0.651	0.700
	RBI	-1.317	0.715	-1.368	0.718	-0.568	0.895
	OBP	0.008	0.151	0.008	0.212	0.013	0.063 *
	Slugging	0.004	0.718	-0.004	0.717	0.007	0.583
	(N)		290		292		282
RotY	Batting Average	0.031	0.014 **				
	Home Runs	13.540	0.000 ***				
	RBI	57.705	0.000 ***				
	OBP	0.041	0.002 ***				
	Slugging	0.049	0.022 **				
	(N)		140				
Cy Young	Earned Run Average	0.126	0.310	-0.008	0.958	0.087	0.640
	Strike Outs	-4.059	0.719	-10.384	0.400	-5.830	0.650
	Wins	-1.255	0.162	0.423	0.653	0.250	0.805
	Losses	0.293	0.624	-0.890	0.170	0.734	0.275
	(N)		173		171		163

Notes: Samples are limited to award winners' performance X years before and after award. Coefficients shown are OLS coefficients on binary variable indicating X years after award. P-values are for t-tests of differences in means of pre and post-award performance. *=90% confidence, **=95% confidence, ***=99% confidence

Table 2.2: Crude Test for Difference in Average Differences

			(Of-Before)-(Of-Before)		(After-Of)-(After-Of)		(2 years After-After)		(After-Before)-(After-Before)	
			Year Before	p-value	**Award Year**	p-value	**Year After**	p-value	**After-Before**	p-value
Voting Matches	MVP	Batting Average	0.007	0.045 **	-0.006	0.090 *	0.001	0.762	-0.002	0.734
		Home Runs	1.910	0.038 **	-2.208	0.014 **	0.068	0.941	-0.336	0.731
		OBP	0.007	0.091 *	-0.006	0.196	0.001	0.829	-0.001	0.843
		RBI	3.539	0.151	-7.338	0.004 ***	1.505	0.591	-3.706	0.209
		Slugging	0.016	0.036 **	-0.016	0.031 **	-0.002	0.801	-0.002	0.779
		(N)		437		441		434		436
	RotY	Batting Average	0.002	0.934	-0.005	0.539	-0.002	0.826	-0.002	0.929
		Home Runs	4.959	0.001 ***	-1.869	0.010 **	1.202	0.095 *	3.146	0.062 *
		OBP	-0.004	0.851	0.001	0.935	-0.004	0.668	-0.003	0.884
		RBI	13.613	0.004 ***	-7.341	0.001 ***	3.556	0.121	6.915	0.199
		Slugging	-0.003	0.921	-0.025	0.131	0.002	0.832	-0.023	0.483
		(N)		178		333		318		174
	Cy Young	Earned Run Average	0.005	0.967	0.069	0.633	-0.282	0.147	0.047	0.782
		Wins	-0.815	0.307	-2.190	0.015 **	2.713	0.006 ***	-3.045	0.004 ***
		Losses	-0.273	0.624	0.605	0.267	-0.676	0.347	0.251	0.722
		Strike Outs	7.986	0.324	-16.246	0.033 **	5.855	0.534	-7.251	0.473
		(N)		180		177		171		175
Predicted Matches	MVP	Batting Average	0.011	0.133	-0.010	0.053 *	0.004	0.528	0.001	0.873
		Home Runs	3.636	0.001 ***	-2.779	0.011 ***	0.694	0.558	0.769	0.550
		OBP	0.014	0.046 **	-0.010	0.076 *	0.002	0.803	0.004	0.515
		RBI	8.014	0.006 ***	-3.785	0.251	2.474	0.514	3.928	0.299
		Slugging	0.038	0.001 **	-0.025	0.008 ***	-0.002	0.830	0.009	0.437
		(N)		311		307		298		305
	RotY	Batting Average	0.027	0.163	-0.009	0.290	-0.001	0.893	0.006	0.758
		Home Runs	8.027	0.000 ***	-0.458	0.636	0.974	0.321	7.585	0.000 ***
		OBP	0.022	0.277	-0.010	0.300	-0.003	0.792	0.006	0.774
		RBI	32.462	0.000 ***	-0.789	0.784	0.227	0.938	30.385	0.000 ***
		Slugging	0.048	0.144	-0.025	0.073 **	0.006	0.675	0.004	0.903
		(N)		123		213		200		119
	Cy Young	Earned Run Average	0.050	0.746	0.212	0.110	-0.172	0.283	0.266	0.136
		Wins	-1.048	0.320	-1.297	0.228	2.455	0.025 **	-2.347	0.063 *
		Losses	-0.877	0.208	0.958	0.110	-0.660	0.414	0.146	0.854
		Strike Outs	-12.581	0.196	1.478	0.872	8.948	0.389	-9.989	0.391
		(N)		164		161		153		161

Note: Estimates shown are a crude difference in difference estimate for award winners compared separately to both BWAA-voted runners-up and statistically-predicated runners-up. That is, a test for statistical difference between the average of decline in performance for winners compared the average of decline for performance for runners-up.

Table 2.3: Award-Year Performance of Winners and BWAA Runners-Up

	MVP	Runner-Up 1	Runner-up 1&2	All non-MVP	Differences in Means		
	(1)	(2)	(3)	(4)	p(1-2)	p(1-3)	p(1-4)
Games	148.914	149.728	149.386	76.212	0.524	0.668	0.000 ***
At-Bats	558.450	566.639	566.843	248.118	0.249	0.170	0.000 ***
Batting Average	0.326	0.320	0.318	0.242	0.085 *	0.014 **	0.000 ***
On-base Percentage	0.411	0.399	0.397	0.308	0.027 **	0.001 ***	0.000 ***
Home Runs	30.967	27.154	26.263	4.901	0.026 **	0.001 ***	0.000 ***
RBI	114.053	106.621	104.705	30.431	0.014 **	0.000 ***	0.000 ***
Slugging	0.576	0.545	0.538	0.344	0.002 ***	0.000 ***	0.000 ***
Stolen Bases	14.219	12.071	11.956	5.569	0.261	0.158	0.000 ***
Strike Outs	68.960	65.048	67.459	32.336	0.321	0.664	0.000 ***
Age	29.060	28.834	28.881	28.115	0.600	0.631	0.008 ***
Tenure	7.530	7.379	7.223	4.791	0.727	0.415	0.000 ***
Team Rank	1.642	2.118	2.207	4.249	0.001 ***	0.000 ***	0.000 ***
Team Wins	93.397	89.746	89.665	74.895	0.001 ***	0.000 ***	0.000 ***
Fielding Errors	12.967	12.839	12.954	9.959	0.911	0.989	0.011 **
(N)	151	169	319	47907	320	470	48058

	RotY	Runner-Up 1	Runner-up 1&2	All non-RotY	Differences in Means		
	(1)	(2)	(3)	(4)	p(1-2)	p(1-3)	p(1-4)
Games	141.604	133.086	131.238	76.317	0.008 ***	0.000 ***	0.000 ***
At-Bats	527.044	486.591	473.994	248.565	0.007 ***	0.000 ***	0.000 ***
Batting Average	0.288	0.283	0.281	0.242	0.120	0.023 **	0.000 ***
On-base Percentage	0.352	0.346	0.345	0.308	0.232	0.136	0.000 ***
Home Runs	16.846	12.276	11.970	4.961	0.001 ***	0.000 ***	0.000 ***
RBI	70.945	59.838	58.250	30.618	0.001 ***	0.000 ***	0.000 ***
Slugging	0.458	0.426	0.425	0.344	0.001 ***	0.000 ***	0.000 ***
Stolen Bases	13.659	12.848	12.202	5.581	0.731	0.483	0.000 ***
Strike Outs	83.604	72.076	71.637	32.354	0.018 **	0.005 ***	0.000 ***
Age	23.659	23.819	23.935	28.126	0.574	0.288	0.000 ***
Tenure	0.769	0.867	0.851	4.807	0.458	0.496	0.000 ***
Team Rank	3.451	4.076	3.952	4.243	0.055 *	0.073 *	0.001 ***
Team Wins	81.912	78.019	78.375	74.940	0.042 **	0.037 **	0.000 ***
Fielding Errors	14.468	13.955	13.037	9.960	0.725	0.265	0.003 ***
(N)	91	105	168	47967	196	259	48058

	Cy Young	Runner-Up 1	Runner-Up 1&2	All non-MVP	Differences in Means		
	(1)	(2)	(3)	(4)	p(1-2)	p(1-3)	p(1-4)
Wins	20.319	18.466	17.872	5.322	0.017 **	0.000 ***	0.000 ***
Losses	7.468	8.239	8.246	5.356	0.110	0.066 *	0.000 ***
Games	39.138	38.227	39.078	25.088	0.612	0.970	0.000 ***
Complete Games	11.638	10.500	9.810	3.941	0.358	0.077 *	0.000 ***
Shutouts	3.553	2.636	2.453	0.546	0.016 **	0.000 ***	0.000 ***
Saves	3.713	4.727	4.827	1.586	0.586	0.488	0.009 ***
ERA	2.517	2.765	2.775	4.988	0.003 ***	0.000 ***	0.000 ***
Base on Balls	64.511	61.773	62.670	33.989	0.450	0.593	0.000 ***
Strike Outs	207.181	167.989	166.486	49.446	0.000 ***	0.000 ***	0.000 ***
Age	29.840	30.523	30.073	27.902	0.312	0.673	0.000 ***
Tenure	7.968	7.920	7.648	4.241	0.941	0.553	0.000 ***
Team Rank	1.872	2.239	2.324	4.172	0.080 *	0.016 **	0.000 ***
Team Wins	90.574	88.307	88.000	75.906	0.139	0.062 *	0.000 ***
(N)	94	88	179	35373	182	273	35467

Notes: Table shows means for MVP's, Cy Young winners, and Rookie's of the Year, runners-up and non-winners, as well as p-values for t-tests of differences in the means of the different groups. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.4: Matching Esimation by BWAA Voting

	Year	MVP	p-value	RotY	p-value		Year	Cy Young	p-value
Batting Average	-3	-0.007	0.176			ERA	-3	0.067	0.689
	-2	0.000	0.918				-2	0.052	0.734
	-1	0.001	0.839	-0.006	0.796		-1	-0.241	0.051 *
	0	0.008	0.006 **	0.004	0.225		0	-0.209	0.014 **
	1	0.003	0.416	0.001	0.853		1	-0.139	0.398
	2	-0.002	0.649	0.000	0.960		2	-0.444	0.027 **
	3	0.004	0.396	-0.004	0.635		3	-0.452	0.017 **
	4	0.003	0.545	0.000	1.000		4	-0.340	0.090 *
Home Runs	-3	0.935	0.537			Wins	-3	0.707	0.460
	-2	2.482	0.089 *				-2	0.166	0.860
	-1	1.820	0.196	0.360	0.122		-1	2.368	0.003 ***
	0	2.616	0.068 *	3.641	0.009 ***		0	1.458	0.046 *
	1	1.476	0.277	2.091	0.255		1	-0.804	0.388
	2	1.137	0.410	3.317	0.088 *		2	2.101	0.020 *
	3	1.548	0.296	2.403	0.155		3	1.783	0.051 *
	4	0.658	0.651	0.985	0.616		4	2.158	0.038 **
Runs	-3	2.671	0.520			Losses	-3	-0.479	0.468
	-2	6.330	0.073 *				-2	0.701	0.236
	-1	5.002	0.096 *	-0.478	0.617		-1	-0.408	0.475
	0	4.654	0.040 **	13.284	0.000 ***		0	-0.623	0.152
	1	1.565	0.600	4.842	0.205		1	0.340	0.554
	2	1.163	0.742	9.821	0.041 *		2	-1.016	0.088 *
	3	3.662	0.337	6.561	0.122		3	-0.500	0.449
	4	2.191	0.566	7.125	0.202		4	0.302	0.608
RBI	-3	2.157	0.585			Strike Outs	-3	12.997	0.293
	-2	6.553	0.082 *				-2	16.201	0.182
	-1	3.919	0.209	0.747	0.412		-1	29.518	0.004 ***
	0	5.604	0.050 *	9.987	0.003 ***		0	33.602	0.001 ***
	1	-0.974	0.771	1.534	0.730		1	19.868	0.083 *
	2	0.194	0.958	6.383	0.195		2	21.500	0.061 *
	3	2.304	0.546	3.313	0.427		3	33.817	0.002 ***
	4	-0.851	0.833	-0.037	0.994		4	24.727	0.043 **
On-base Percentage	-3	-0.005	0.494				(N)	167	
	-2	0.002	0.746						
	-1	0.008	0.137	-0.004	0.862				
	0	0.013	0.008 ***	0.006	0.221				
	1	0.011	0.057 *	0.001	0.885				
	2	0.005	0.354	0.003	0.623				
	3	0.009	0.132	-0.002	0.830				
	4	0.009	0.197	0.009	0.252				
Slugging	-3	-0.010	0.441						
	-2	0.015	0.195						
	-1	0.015	0.160	0.026	0.465				
	0	0.027	0.002 **	0.024	0.010 **				
	1	0.019	0.091 *	0.006	0.633				
	2	0.005	0.612	0.020	0.157				
	3	0.013	0.246	0.007	0.599				
	4	0.017	0.190	0.007	0.684				
	(N)	292		153					

Notes: Table shows estimates of difference between MVP, Rookies of the Year, and Cy Young performances and their runners-up according to BWAA voting from three years before to four years after an award. Estimates are calculated using Abadie et al 2001 matching estimator. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.5: Regression Estimation by BWAA Voting

	Year	MVP	p-value		RotY	p-value		Year	Cy Young	p-value	
Batting Average	-3	-0.007	0.189				Earned Run Average	-3	-0.094	0.608	
	-2	0.003	0.610					-2	0.030	0.838	
	-1	-0.001	0.867		0.016	0.410		-1	-0.260	0.029	**
	0	0.006	0.053	*	0.006	0.033 **		0	-0.248	0.003	***
	1	0.001	0.813		0.003	0.514		1	-0.170	0.256	
	2	-0.001	0.725		0.006	0.264		2	-0.451	0.027	**
	3	0.001	0.760		0.003	0.604		3	-0.463	0.012	**
	4	0.002	0.761		0.005	0.401		4	-0.396	0.044	**
Home Runs	-3	3.692	0.016				Wins	-3	1.318	0.185	
	-2	4.621	0.002 ***					-2	0.639	0.476	
	-1	4.262	0.004 ***		0.437	0.059 *		-1	2.658	0.001	***
	0	6.357	0.000 **		5.245	0.000 ***		0	1.853	0.017	**
	1	3.851	0.011 **		2.570	0.075 *		1	-0.369	0.695	
	2	3.709	0.010 **		4.389	0.006 ***		2	2.173	0.024	**
	3	2.999	0.041 **		4.379	0.007 ***		3	2.070	0.038	**
	4	2.980	0.045 **		3.628	0.031 **		4	2.393	0.018	**
Runs	-3	3.910	0.325				Losses	-3	-0.453	0.466	
	-2	6.304	0.071 *					-2	0.696	0.249	
	-1	5.812	0.047 **		0.402	0.603		-1	-0.480	0.412	
	0	6.152	0.005 ***		15.292	0.000 ***		0	-0.771	0.110	
	1	1.900	0.529		4.725	0.168		1	-0.155	0.808	
	2	1.720	0.572		9.694	0.014 **		2	-0.926	0.151	
	3	2.533	0.450		8.443	0.040 **		3	-0.138	0.847	
	4	2.275	0.549		12.839	0.003 ***		4	0.439	0.515	
RBI	-3	6.054	0.142				Strike Outs	-3	23.059	0.068	*
	-2	9.546	0.013 **					-2	23.640	0.039	**
	-1	8.091	0.013 **		1.579	0.053 *		-1	31.637	0.002	***
	0	11.967	0.000 ***		14.103	0.000 ***		0	39.192	0.000	***
	1	3.388	0.309		4.421	0.254		1	21.963	0.058	*
	2	4.397	0.196		10.217	0.018 **		2	24.806	0.038	**
	3	4.425	0.215		9.708	0.018 **		3	34.478	0.004	***
	4	3.281	0.408		8.474	0.058 *		4	28.940	0.017	**
On-base Percentage	-3	-0.002	0.732				(N)		182		
	-2	0.004	0.507								
	-1	0.005	0.365		0.013	0.516					
	0	0.011	0.015 **		0.004	0.266					
	1	0.008	0.139		-0.001	0.917					
	2	0.004	0.429		0.004	0.423					
	3	0.005	0.362		0.001	0.913					
	4	0.005	0.468		0.010	0.098 *					
Slugging	-3	0.007	0.547								
	-2	0.032	0.006 ***								
	-1	0.026	0.008 ***		0.044	0.150					
	0	0.044	0.000 ***		0.034	0.000 ***					
	1	0.029	0.006 ***		0.014	0.176					
	2	0.020	0.050 *		0.028	0.021 **					
	3	0.020	0.065 *		0.025	0.049 **					
	4	0.027	0.033 **		0.022	0.097 *					
(N)		384			249						

Notes: Table shows regression estimates of differences between award winner performances and their runners-up according to BWAA voting from three years before to four years after an award. Estimates are calculated using OLS estimation when sample is limited to winners and runners up. ***=99% confidence, **=95% confidence, *=90% confidence.

Figure 2.4: MVP Winner and Predicted Runners-Up Performance over Time
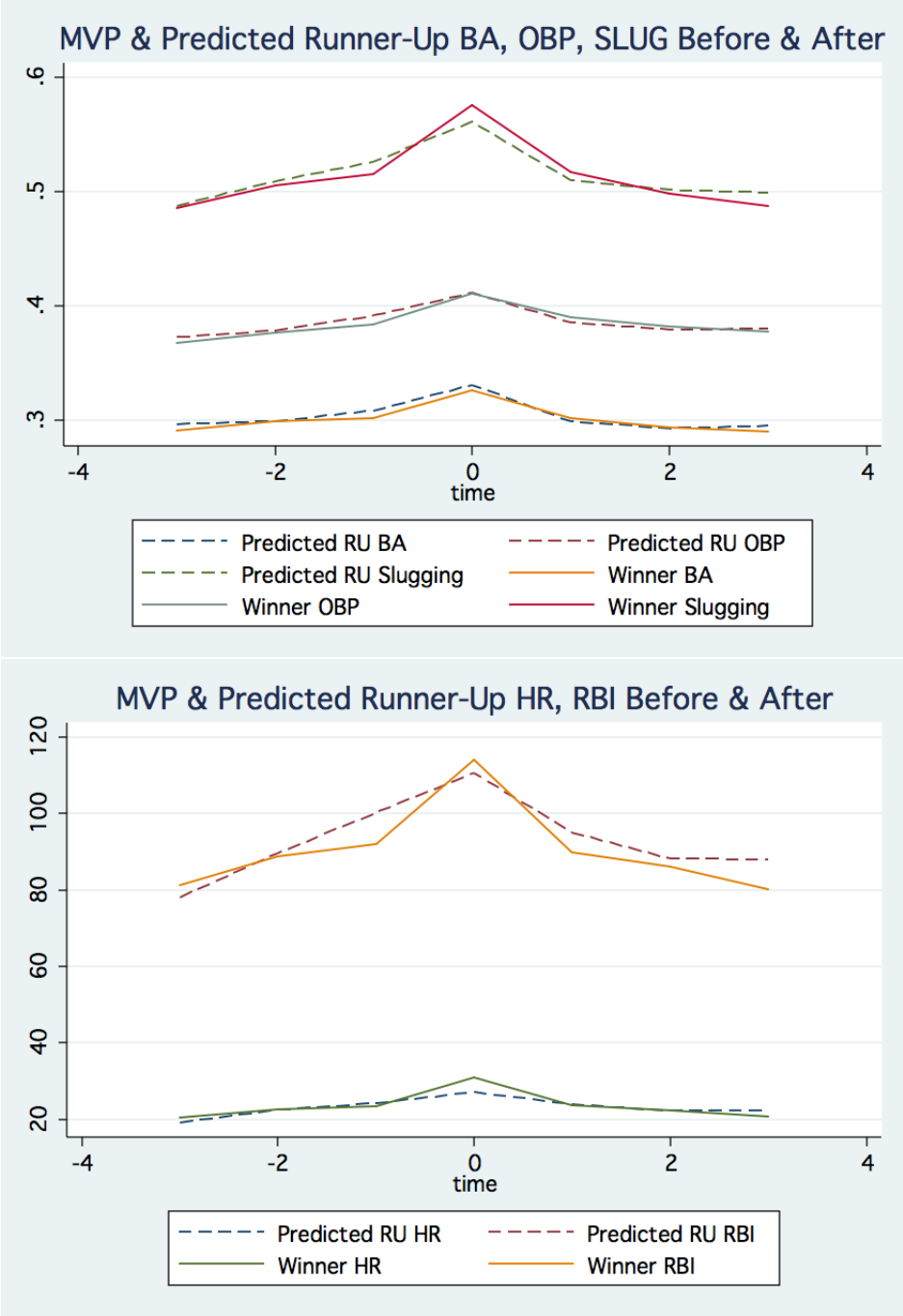
Figure 2.5: Rookie of the Year and Predicted Runners-Up Performance over Time
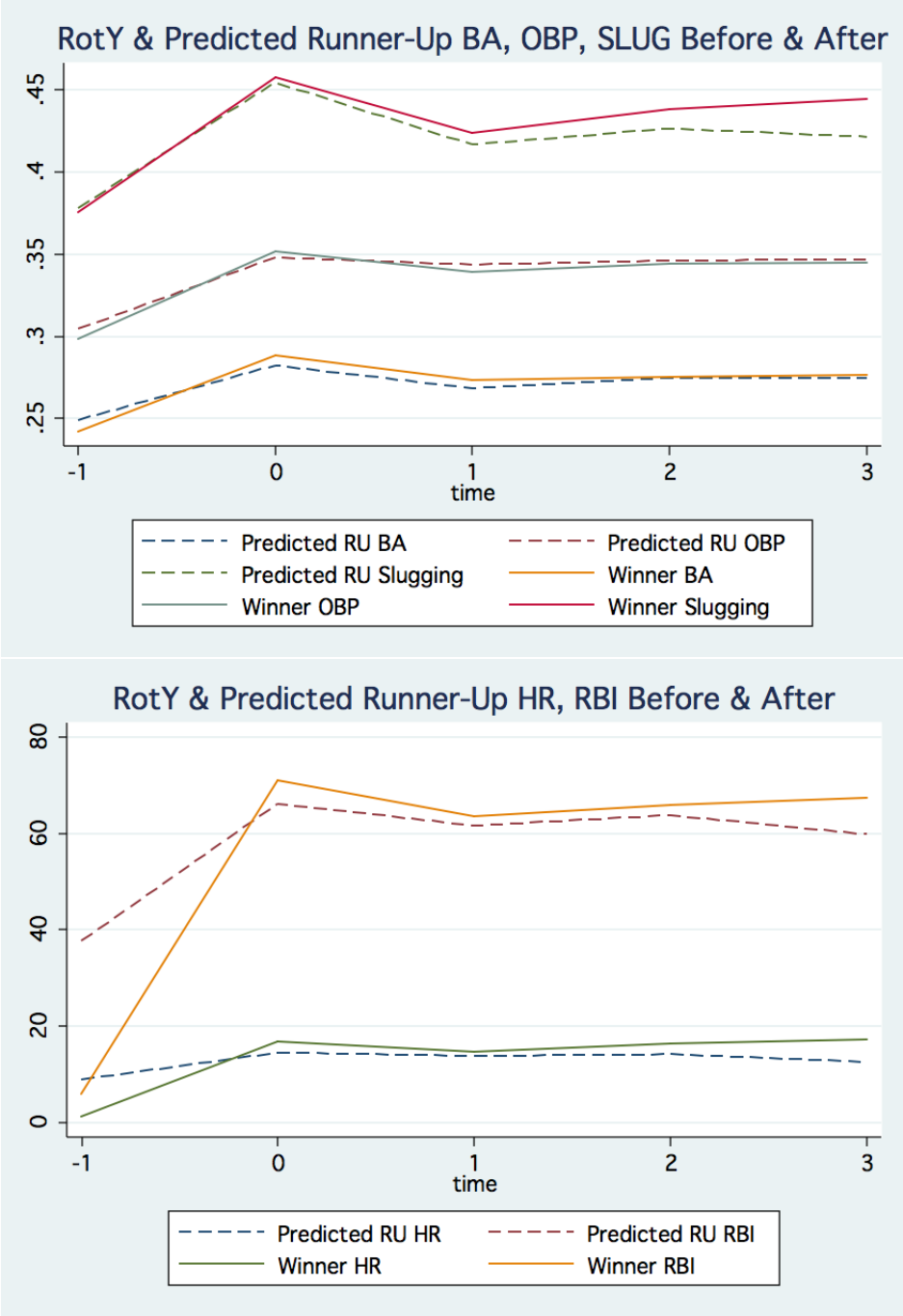
Figure 2.6: Cy Young and Predicted Runners-Up Performance over Time

Table 2.6: Average Performance of Winners and Logit-Predicted Winners

	MVP	Predicted	Difference	**RotY**	Predicted	Difference
Statistic	(1)	(2)	p(1-2)	(3)	(4)	p(3-4)
Games	148.914	145.443	0.151	115.623	116.533	0.887
At-bats	558.450	550.229	0.441	405.877	412.242	0.827
Batting Average	0.326	0.337	0.264	0.256	0.258	0.851
Runs	108.834	106.321	0.393	60.705	62.550	0.718
Home Runs	30.967	26.710	0.020 **	12.656	12.825	0.917
RBI	114.053	109.977	0.258	53.959	57.975	0.412
Stolen Bases	14.219	15.802	0.482	10.205	10.358	0.934
On-base Percentage	0.411	0.418	0.469	0.314	0.317	0.805
Strike Outs	68.960	74.960	0.167	66.631	56.045	0.045 **
Slugging	0.576	0.564	0.445	0.398	0.400	0.938
Team Wins	93.397	93.802	0.730	82.369	81.550	0.620
Team Rank	1.642	1.985	0.016 **	3.361	3.658	0.289
Age	29.060	28.916	0.739	23.754	28.605	0.000 ***
American (0/1)	0.894	0.893	0.980	0.836	0.850	0.767
Tenure	7.530	7.145	0.378	0.770	6.008	0.000 ***
Switch Hitter (0/1)	0.066	0.061	0.860	0.098	0.058	0.249
(N)	151	131	282	122	120	242

	Cy Young	Predicted	Difference
Statistic	(5)	(6)	p(5-6)
Wins	20.319	19.800	0.662
Losses	7.468	7.814	0.495
Games	39.138	43.543	0.036 **
Complete Games	11.638	11.529	0.955
Shutouts	3.553	3.229	0.524
Saves	3.713	3.457	0.871
Earned Run Average	2.517	2.670	0.123
Base on Balls	64.511	64.243	0.946
Strikeouts	207.181	196.557	0.384
Team Wins	90.574	88.771	0.268
Team Rank	1.872	2.443	0.012 **
Age	29.840	29.771	0.925
American (0/1)	0.883	0.857	0.627
Tenure	7.968	7.714	0.720
(N)	94	70	164

Notes: Table shows averages of MVP's, Rookies of the Year, and Cy Young winners performances and their runners-up according to logit estimation. P-values are from t-tests of differences in means. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.7: Matching Estimates for Winners and Logit-Predicted Winners

	Year	MVP	p-value	RotY	p-value		Year	Cy Young	p-value
Batting Average	-3	0.010	0.787			Earned Run Average	-3	-1.033	0.548
	-2	0.046	0.239				-2	0.006	0.998
	-1	0.052	0.208	-0.230	0.001 ***		-1	-2.067	0.356
	0	0.100	0.058 *	0.041	0.458		0	-2.226	0.486
	1	0.069	0.110	0.047	0.328		1	-0.224	0.932
	2	0.041	0.282	0.025	0.560		2	-1.576	0.501
	3	0.051	0.166	0.052	0.217		3	-0.876	0.697
	4	0.000	0.999	-0.003	0.946		4	-0.510	0.798
Home Runs	-3	2.606	0.603			Wins	-3	0.065	0.988
	-2	2.447	0.717				-2	1.252	0.761
	-1	1.460	0.835	-5.929	0.275		-1	12.569	0.007 ***
	0	0.132	0.979	6.017	0.247		0	15.712	0.002 ***
	1	2.784	0.631	4.994	0.423		1	-0.193	0.963
	2	-0.257	0.970	-2.471	0.692		2	5.372	0.225
	3	0.221	0.973	4.326	0.464		3	3.621	0.395
	4	-0.742	0.906	2.822	0.645		4	-1.426	0.723
Runs	-3	23.870	0.195			Losses	-3	-0.659	0.847
	-2	45.327	0.019 **				-2	13.635	0.002 ***
	-1	42.777	0.020 **	-38.299	0.073 *		-1	5.271	0.145
	0	43.225	0.012 **	10.719	0.593		0	7.840	0.043 **
	1	47.585	0.022 **	14.145	0.512		1	5.168	0.163
	2	23.137	0.210	1.946	0.925		2	6.575	0.074 *
	3	29.609	0.124	19.809	0.357		3	4.544	0.187
	4	17.217	0.356	10.391	0.606		4	5.474	0.112
RBI	-3	13.538	0.434			Strike Outs	-3	14.701	0.675
	-2	16.564	0.332				-2	61.785	0.067 *
	-1	34.584	0.048 **	-34.616	0.088 *		-1	46.182	0.161
	0	26.227	0.064 **	35.956	0.076 *		0	46.559	0.121
	1	14.498	0.377	47.264	0.048 **		1	-9.668	0.763
	2	14.802	0.418	12.258	0.556		2	10.406	0.750
	3	3.656	0.844	28.470	0.177		3	41.989	0.260
	4	0.109	0.995	25.583	0.210		4	-6.667	0.837
On-base Percentage	-3	0.015	0.699			(N)		35202	
	-2	0.048	0.249						
	-1	0.052	0.253	-0.290	0.000 ***				
	0	0.097	0.075 *	0.020	0.739				
	1	0.060	0.175	0.024	0.650				
	2	0.050	0.220	-0.008	0.871				
	3	0.047	0.220	0.029	0.525				
	4	0.025	0.485	-0.027	0.534				
Slugging	-3	0.024	0.708						
	-2	0.043	0.517						
	-1	0.069	0.345	-0.327	0.004 ***				
	0	0.095	0.248	0.074	0.416				
	1	0.078	0.281	0.060	0.468				
	2	0.029	0.666	0.006	0.932				
	3	0.042	0.507	0.068	0.348				
	4	-0.042	0.561	0.000	1.000				
	(N)	45761		45761					

Notes: Table shows matching estimates of differences between MVP, Rookie of the Year, and Cy Young performances and their runners-up according to logit model predictions from three years before to four years after an award. Estimates are calculated using Abadie et al 2001 matching estimator. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.8: Regression Estimates for Winners and Logit-Predicted Winners

	Year	MVP	p-value	RotY	p-value		Year	Cy Young	p-value
Batting Average	-3	-0.008	0.299			Earned Run Average	-3	-0.035	0.870
	-2	-0.006	0.399				-2	-0.153	0.448
	-1	-0.006	0.365	0.006	0.762		-1	-0.204	0.165
	0	-0.007	0.509	0.001	0.911		0	-0.153	0.123
	1	0.002	0.768	-0.004	0.669		1	0.060	0.677
	2	0.008	0.317	-0.013	0.190		2	-0.017	0.913
	3	-0.005	0.412	-0.004	0.704		3	-0.211	0.313
	4	-0.012	0.125	-0.015	0.150		4	-0.062	0.764
Home Runs	-3	0.587	0.735			Wins	-3	1.412	0.220
	-2	-0.388	0.822				-2	1.965	0.059 *
	-1	-0.978	0.566	-4.939	0.000 ***		-1	1.567	0.150
	0	4.032	0.033 **	1.933	0.205		0	0.519	0.662
	1	-0.486	0.783	0.276	0.863		1	-0.822	0.439
	2	-0.200	0.908	0.748	0.665		2	1.104	0.328
	3	-2.327	0.206	2.643	0.128		3	1.242	0.304
	4	-1.753	0.352	0.619	0.738		4	0.856	0.512
Runs	-3	-0.356	0.943			Losses	-3	0.571	0.408
	-2	-2.341	0.614				-2	1.557	0.019 **
	-1	-5.672	0.200	-24.833	0.000 ***		-1	0.531	0.429
	0	4.302	0.352	5.848	0.250		0	-0.346	0.495
	1	-4.265	0.338	-0.200	0.970		1	0.616	0.355
	2	-3.598	0.429	0.346	0.949		2	-0.156	0.841
	3	-7.386	0.116	3.005	0.596		3	0.590	0.462
	4	-7.378	0.142	0.034	0.995		4	0.787	0.298
RBI	-3	0.173	0.973			Strike Outs	-3	25.623	0.059 *
	-2	-3.464	0.476				-2	26.758	0.033 **
	-1	-8.493	0.070 *	-21.364	0.000 ***		-1	23.204	0.050 *
	0	5.971	0.242	5.307	0.275		0	10.624	0.384
	1	-5.637	0.232	-0.069	0.989		1	10.514	0.398
	2	-3.686	0.446	-1.673	0.761		2	14.077	0.293
	3	-11.274	0.026 **	2.051	0.713		3	30.029	0.028 **
	4	-8.844	0.091 *	-6.656	0.234		4	16.518	0.226
On-base Percentage	-3	-0.008	0.415			(N)		164	
	-2	-0.009	0.318						
	-1	-0.011	0.229	0.003	0.896				
	0	-0.009	0.432	-0.003	0.803				
	1	-0.004	0.727	-0.017	0.190				
	2	0.003	0.760	-0.011	0.327				
	3	-0.007	0.421	-0.010	0.432				
	4	-0.017	0.090 *	-0.020	0.112				
Slugging	-3	-0.007	0.669						
	-2	-0.016	0.337						
	-1	-0.011	0.449	0.028	0.369				
	0	0.014	0.455	-0.003	0.916				
	1	0.004	0.796	-0.011	0.560				
	2	0.004	0.807	-0.013	0.508				
	3	-0.013	0.408	0.010	0.612				
	4	-0.022	0.215	-0.019	0.367				
(N)		318		217					

Notes: Table shows regression estimates of differences between award winner performances and their runners-up according to logit model predictions from three years before to four years after an award. Estimates are calculated using OLS estimation when sample is limited to winners and runners up. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.9: Award Year Performance of Grouped Award Winners and Predicted Matches

Statistic	ANY	Predicted	Difference	MVP or RotY	Predicted	Difference
	(1)	(2)	p(1-2)	(3)	(4)	p(3-4)
Games	145.255	144.230	0.340	146.125	147.583	0.305
At-bats	545.218	541.110	0.404	546.113	553.543	0.311
Batting Average	0.306	0.301	0.164	0.311	0.302	0.038 *
Runs	94.589	90.663	0.005 ***	97.975	97.179	0.737
Home Runs	25.986	22.964	0.000 ***	25.750	25.184	0.682
RBI	95.968	90.987	0.003 ***	97.933	95.130	0.354
Stolen Bases	13.266	13.972	0.426	13.850	14.108	0.869
On-base Percentage	0.381	0.374	0.053 *	0.388	0.377	0.037 **
Strike Outs	80.601	80.161	0.819	74.538	82.060	0.020 **
Slugging	0.522	0.502	0.001 ***	0.532	0.516	0.178
Team Wins	86.308	85.379	0.156	88.942	88.139	0.445
Team Rank	2.440	2.971	0.000 ***	2.333	2.614	0.073 *
Age	28.815	28.716	0.649	27.042	27.067	0.940
American (0/1)	0.785	0.818	0.121	0.867	0.843	0.472
Tenure	6.784	6.611	0.432	5.025	4.933	0.794
Switch Hitter (0/1)	0.096	0.094	0.907	0.092	0.099	0.798
(N)	772	617	1389	240	223	463

	Any Pitching	Predicted	Difference
	(5)	(6)	p(5-6)
Wins	20.538	20.123	0.640
Losses	8.140	8.396	0.538
Games	38.769	38.821	0.972
Complete Games	13.133	13.142	0.995
Shutouts	3.538	3.358	0.626
Saves	2.874	1.425	0.146
Earned Run Average	2.659	2.660	0.990
Base on Balls	70.664	74.075	0.365
Strikeouts	192.909	175.340	0.054 *
Team Wins	89.650	88.547	0.405
Team Rank	2.112	2.538	0.035 **
Age	29.965	29.396	0.334
American (0/1)	0.923	0.915	0.820
Tenure	8.147	7.425	0.191
(N)	143	106	249

Notes: Table shows average performance in year of award of winners of any offensive award (MVP, RotY, ALCS MVP, All-Star MVP, Babe Ruth award, Hank Aaron award, Major League Player of the Year, NLCS MVP, Silver Slugger, Triple Crown, World Series MVP), MVP or RotY, and Cy Young or Sporting News' Pitcher of the Year winners performances and their runners-up according to logit estimation. P-values are from t-tests of differences in means. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.10: Regression Estimates for Grouped Winners and Predicted Winners

	Year	Any Off.	p-value		MVP or RotY	p-value			Year	Any Pitch	p-value	
Batting Average	-3	0.003	0.383		0.006	0.287		Earned Run Average	-3	-0.203	0.274	
	-2	0.004	0.106		0.004	0.523			-2	-0.407	0.007	***
	-1	0.003	0.181		-0.002	0.760			-1	-0.204	0.064	*
	0	0.004	0.164		0.009	0.038	**		0	-0.001	0.990	
	1	0.007	0.001	***	0.000	0.923			1	-0.158	0.287	
	2	0.006	0.008	***	0.001	0.826			2	-0.239	0.249	
	3	0.003	0.280		0.004	0.342			3	-0.351	0.086	*
	4	-0.002	0.466		-0.002	0.679			4	-0.478	0.091	*
Home Runs	-3	1.854	0.022	**	4.520	0.004	***	Wins	-3	0.825	0.389	
	-2	3.053	0.000	***	3.957	0.010	***		-2	0.797	0.370	
	-1	2.098	0.005	***	-1.799	0.215			-1	0.812	0.363	
	0	3.021	0.000	***	0.566	0.682			0	0.416	0.640	
	1	2.569	0.000	***	-1.006	0.440			1	-1.037	0.264	
	2	3.191	0.000	***	-0.569	0.655			2	0.169	0.852	
	3	2.290	0.003	***	-1.326	0.313			3	-0.251	0.780	
	4	2.563	0.001	***	-0.097	0.942			4	0.361	0.711	
Runs	-3	4.625	0.029	**	16.080	0.000	***	Losses	-3	-0.257	0.672	
	-2	5.592	0.004	***	13.083	0.001	***		-2	-0.143	0.808	
	-1	1.637	0.373		-9.032	0.017	**		-1	-0.436	0.432	
	0	3.926	0.005	***	0.796	0.737			0	-0.256	0.538	
	1	3.723	0.020	**	-1.251	0.642			1	-0.370	0.510	
	2	5.191	0.002	***	-1.667	0.559			2	-0.647	0.293	
	3	4.040	0.031	**	-2.619	0.393			3	-0.998	0.106	
	4	4.004	0.039	**	2.400	0.447			4	-0.814	0.176	
RBI	-3	3.275	0.150		15.960	0.000	***	Strike Outs	-3	29.114	0.006	
	-2	5.804	0.006	***	15.135	0.000	***		-2	22.400	0.017	**
	-1	1.994	0.316		-7.375	0.066	*		-1	15.912	0.084	
	0	4.981	0.003	***	2.803	0.354			0	17.569	0.054	*
	1	2.727	0.131		-4.935	0.108			1	12.204	0.193	
	2	6.166	0.001	***	-4.091	0.199			2	16.579	0.086	*
	3	4.646	0.021	**	-5.014	0.133			3	7.125	0.461	
	4	3.843	0.064	*	-2.360	0.489			4	2.600	0.795	
On-base Percentage	-3	0.001	0.706		0.013	0.071	*	(N)		249		
	-2	0.003	0.403		0.006	0.491						
	-1	0.001	0.644		0.000	0.964						
	0	0.006	0.053	*	0.011	0.037	**					
	1	0.008	0.003	***	0.003	0.467						
	2	0.008	0.003	***	0.004	0.402						
	3	0.005	0.148		0.005	0.323						
	4	0.001	0.706		0.002	0.653						
Slugging	-3	0.013	0.051	*	0.021	0.099	*					
	-2	0.018	0.003	***	0.015	0.272						
	-1	0.013	0.014	**	0.000	0.975						
	0	0.019	0.001	***	0.015	0.178						
	1	0.020	0.000	***	-0.002	0.869						
	2	0.020	0.000	***	-0.002	0.852						
	3	0.013	0.027	**	0.000	0.979						
	4	0.011	0.088	*	0.000	0.995						
(N)		1389			463							

Notes: Table shows regression estimates of differences between award winner performances and their runners-up according to logit model predictions from three years before to four years after an award. Estimates are calculated using OLS estimation when sample is limited to winners and runners up. ***=99% confidence, **=95% confidence, *=90% confidence.

Table 2.11: Performance of Team of Award Winners Before, During and After Award

Voted Runners-Up		Year Before	p-value		Year Of	p-value		Year After	p-value	
MVP	Team Wins	1.168	0.267		6.207	0.000	***	1.960	0.099	*
	Team Rank	-0.147	0.395		-0.985	0.000	***	-0.365	0.046	**
	League Win (0/1)	0.047	0.229		0.324	0.000	***	0.096	0.021	**
	ln(Attendance)	0.023	0.396		0.093	0.008	***	0.040	0.351	
	(N)		384			384			381	
RotY	Team Wins	0.257	0.836		1.592	0.241		2.462	0.094	*
	Team Rank	-0.173	0.366		-0.254	0.230		-0.139	0.528	
	League Win (0/1)	0.046	0.123		0.059	0.130		0.040	0.264	
	ln(Attendance)	-0.025	0.396		-0.044	0.242		-0.027	0.537	
	(N)		331			331			322	
Cy Young	Team Wins	3.510	0.036	**	2.336	0.126		0.501	0.778	
	Team Rank	-0.444	0.082	*	-0.393	0.059	*	0.022	0.935	
	League Win (0/1)	0.052	0.357		0.199	0.004	***	0.088	0.122	
	ln(Attendance)	0.012	0.721		0.031	0.461		0.031	0.552	
	(N)		176			175			170	
MVP or RotY	Team Wins	1.156	0.180		5.294	0.000	***	2.754	0.005	***
	Team Rank	-0.228	0.096	*	-0.893	0.000	***	-0.345	0.021	**
	League Win (0/1)	0.036	0.177		0.236	0.000	***	0.076	0.009	***
	ln(Attendance)	0.014	0.515		0.062	0.022	**	0.038	0.246	
	(N)		642			642			631	
Any 3	Team Wins	2.120	0.009	***	5.841	0.000	***	2.906	0.001	***
	Team Rank	-0.416	0.001	***	-0.939	0.000	***	-0.362	0.010	**
	League Win (0/1)	0.054	0.031	**	0.235	0.000	***	0.074	0.006	**
	ln(Attendance)	0.022	0.266		0.078	0.002	***	0.058	0.053	*
	(N)		719			718			705	
Predicted Runners-Up		Year Before	p-value		Year Of	p-value		Year After	p-value	
MVP	Team Wins	-0.447	0.711		4.067	0.001	***	-0.298	0.831	
	Team Rank	0.081	0.652		-1.096	0.000	***	-0.087	0.673	
	League Win (0/1)	0.002	0.959		0.307	0.000	***	0.048	0.351	
	ln(Attendance)	-0.039	0.191		0.053	0.156		0.011	0.823	
	(N)		295			295			291	
RotY	Team Wins	1.182	0.426		0.915	0.587		2.444	0.151	
	Team Rank	-0.110	0.621		-0.063	0.816		-0.305	0.256	
	League Win (0/1)	0.033	0.359		-0.001	0.978		0.039	0.364	
	ln(Attendance)	0.028	0.404		-0.006	0.877		-0.016	0.757	
	(N)		211			211			208	
Cy Young	Team Wins	1.829	0.304		2.802	0.088	*	-0.300	0.876	
	Team Rank	-0.284	0.310		-0.734	0.001	***	0.059	0.838	
	League Win (0/1)	0.086	0.141		0.175	0.021	**	0.055	0.392	
	ln(Attendance)	0.026	0.485		0.105	0.024	**	0.107	0.061	*
	(N)		154			153			151	
MVP or RotY	Team Wins	0.208	0.839		2.938	0.009	***	1.376	0.236	
	Team Rank	-0.042	0.781		-0.794	0.000	***	-0.247	0.160	
	League Win (0/1)	0.010	0.767		0.200	0.000	***	0.049	0.181	
	ln(Attendance)	-0.008	0.731		0.033	0.260		-0.001	0.971	
	(N)		468			468			461	
Any 3	Team Wins	0.851	0.374		3.578	0.000	***	1.079	0.310	
	Team Rank	-0.185	0.197		-0.845	0.000	***	-0.256	0.115	
	League Win (0/1)	0.027	0.380		0.190	0.000	***	0.058	0.081	*
	ln(Attendance)	-0.001	0.971		0.049	0.065	***	0.006	0.845	
	(N)		554			553			546	

Notes: Estimates shown are difference in average performance of team with an award winner as compared to team with a runner up (sample is limited to teams with winners and teams with runners-up), controlling for team's performance in given statistic 2 years before the award. The top half defines runners-up according to voting, while the bottom half defines runners-up according to the logit-predictions.

Table 2.12: Performance of Teammates of Award Winners Before, During, and After Award

		Year Before	p-value		Year Of	p-value		Year After	p-value	
Voted Runners-Up										
MVP	Batting Average	-0.001	0.604		-0.002	0.387		0.002	0.325	
	Home Runs	-0.010	0.919		-0.007	0.953		-0.026	0.849	
	OBP	-0.003	0.246		-0.002	0.389		0.001	0.618	
	RBI	0.281	0.476		0.456	0.322		0.200	0.715	
	Slugging	0.000	0.974		-0.004	0.270		0.001	0.690	
	(N)		7618			7446			6418	
RotY	Batting Average	0.004	0.091	*	0.000	0.958		0.000	0.888	
	Home Runs	-0.197	0.084	*	-0.317	0.013	**	0.062	0.681	
	OBP	0.004	0.115		0.000	0.954		0.001	0.672	
	RBI	-0.599	0.128		-0.334	0.473		0.195	0.722	
	Slugging	0.003	0.373		-0.002	0.670		-0.002	0.580	
	(N)		7032			6839			5722	
Cy Young	Batting Average	0.003	0.335		0.002	0.560		0.004	0.254	
	Home Runs	0.002	0.990		-0.029	0.872		-0.406	0.057	*
	OBP	0.005	0.187		0.002	0.548		0.002	0.537	
	RBI	-0.148	0.776		0.027	0.965		-0.350	0.631	
	Slugging	0.008	0.118		0.002	0.675		0.001	0.905	
	(N)		3587			3487			3007	
MVP or RotY	Batting Average	0.001	0.514		-0.001	0.667		0.000	0.964	
	Home Runs	-0.127	0.106		-0.198	0.023	**	-0.063	0.551	
	OBP	0.000	0.928		0.000	0.836		0.001	0.772	
	RBI	-0.308	0.287		-0.062	0.856		-0.009	0.983	
	Slugging	0.001	0.810		-0.003	0.253		-0.002	0.429	
	(N)		13123			12800			10857	
Any 3	Batting Average	0.003	0.092	*	0.000	0.990		0.001	0.703	
	Home Runs	0.010	0.889		-0.055	0.502		-0.037	0.712	
	OBP	0.002	0.242		0.000	0.888		0.001	0.521	
	RBI	-0.061	0.821		0.129	0.685		0.093	0.806	
	Slugging	0.005	0.051	*	-0.001	0.715		-0.001	0.660	
	(N)		14749			14388			12196	
Predicted Runners-Up										
MVP	Batting Average	-0.002	0.537		-0.002	0.572		-0.002	0.537	
	Home Runs	-0.148	0.203		-0.206	0.108		-0.091	0.554	
	OBP	-0.001	0.844		-0.001	0.759		-0.002	0.519	
	RBI	0.019	0.965		-0.297	0.570		-0.319	0.598	
	Slugging	-0.003	0.359		-0.004	0.354		-0.007	0.107	
	(N)		5920			5801			5002	
RotY	Batting Average	0.005	0.054	*	0.000	0.963		0.002	0.481	
	Home Runs	0.159	0.245		-0.226	0.150		0.073	0.694	
	OBP	0.008	0.017	**	0.004	0.250		0.005	0.202	
	RBI	-0.170	0.726		-0.367	0.534		0.132	0.848	
	Slugging	0.012	0.004	***	0.002	0.671		0.006	0.195	
	(N)		4362			4256			3553	
Cy Young	Batting Average	0.004	0.230		0.002	0.647		0.004	0.333	
	Home Runs	0.042	0.806		0.062	0.751		-0.178	0.443	
	OBP	0.008	0.045	**	0.003	0.481		0.003	0.569	
	RBI	-0.627	0.286		-0.619	0.369		-1.338	0.105	
	Slugging	0.007	0.192		0.004	0.435		0.000	0.980	
	(N)		3092			3019			2644	
MVP or RotY	Batting Average	0.000	0.879		-0.001	0.523		-0.002	0.484	
	Home Runs	-0.031	0.739		-0.268	0.009	***	-0.065	0.598	
	OBP	0.002	0.484		0.001	0.618		-0.001	0.785	
	RBI	-0.210	0.534		-0.560	0.169		-0.287	0.545	
	Slugging	0.002	0.463		-0.002	0.481		-0.004	0.230	
	(N)		9488			9281			7874	
Any 3	Batting Average	0.001	0.491		-0.001	0.658		-0.001	0.681	
	Home Runs	0.038	0.656		-0.120	0.216		-0.044	0.704	
	OBP	0.003	0.162		0.002	0.435		0.000	0.933	
	RBI	-0.159	0.610		-0.418	0.264		-0.406	0.355	
	Slugging	0.004	0.129		0.000	0.920		-0.004	0.241	
	(N)		11247			11002			9382	

Notes: Estimates shown are coefficient of being a team-mate of actual winner as opposed to being the team-mate of a runner-up for a given award (Sample is limited to only team-mates of winners and runners-up (and does not include winners and runners-up themselves)). All regressions also control for the given performance statistic 2 years before award year.

Chapter 3

Scholarships in Kenya

3.1   Introduction

Governments that cannot afford to provide free education for all their citizens may be interested in giving scholarships to the extent possible. Giving these limited scholarships to high achievers is attractive since it rewards effort and creates incentives for excellence, but this may create equity and sustainability concerns since a fully competitive program may tend to marginalize the poor and disadvantaged. To help determine the number of awards that might go to disadvantaged students, we use new tracking survey data from the Girls Scholarship Program (GSP) and the Kenyan Life Panel Survey (KLPS) to evaluate the effects of different targeting rules for a secondary-school scholarship using Kenyan Certificate of Primary Education (KCPE) exam scores. We also use preliminary data to evaluate the medium-run impacts of a randomized primary school merit scholarship program conducted in Western Kenya in 2001 and 2002.

We look at different methods of geographically dividing scholarships (i.e., giving them to the top scorers in each school as opposed to in each district) in order to investigate the possibility of clustering of people with similar socioeconomic status. If wealthier families all sent their children to the same school or lived in the same small area, we would expect to find that giving bursaries separately to each school instead of to the district as a whole would do much to target those who are less well-off. Instead we find that such methods do not do much to alleviate the under-representation of disadvantaged groups amongst the winners.

We also use data from an ongoing field tracking survey, the Girls Scholarship Program (GSP) to present evidence on the effects of scholarship programs. Preliminary evidence is inconclusive, but suggests there is a moderate positive effect on test scores from attending a school in which the scholarships were offered. The evidence indicates, however, that winning one of the scholarships had an insignificant effect on the winners.

Section 3.2 discusses different types of targeting rules, section 3.3 introduces the primary school scholarship program, section 3.4 presents preliminary findings on the effect of attend-

ing a school where merit scholarships were awarded, section 3.5 presents preliminary findings of the effect of winning a scholarship, and section 3.6 concludes.

3.2 Estimating Scholarship Winners under Different Targeting Rules

Current bursary schemes in Kenya have limitations in targeting those in most need. An easy way to determine scholarship winners would be to reward high scorers on the KCPE within regions, but it is uncertain how many of these awards would go to the underprivileged. We can help to answer this question using new tracking survey panel data from the Kenya Life Panel Survey (KLPS) and GSP surveys. Awards could be given to those with the highest KCPE scores in each district, or in each division, each zone, each school, or they could be based on other criteria as well. We examine what proportion of awards would be given to students with different characteristics depending on how the program is designed.

The KLPS has data from 5200 adolescents who were originally surveyed as part of a deworming project that started in 1998 (see Kremer and Miguel [2004] for details on the original deworming study). They were re-interviewed in 2004 and 2005 when they were between the ages of 14 and 25. Of the 5200 students surveyed in the KLPS data, 2015 of them have taken the KCPE at least once. We have scaled older 700-maximum KCPE scores to fit in line with the current 500-maximum. These 2015 students are from 75 different schools in the Budalangi and Funyula divisions of Busia district. Busia district is in Kenya's Western Province, the third poorest of Kenya's eight provinces, which the government has described as "uniformly and deeply poor" [Central Bureau of Statistics Kenya, 2003a]. 67% of households in the rural parts of Busia are below the poverty line, compared to 60% of households in the province [Central Bureau of Statistics Kenya, 2003b].

A rule of giving scholarships to the top 10% or top 20% of highest scorers on the KCPE in the district, or instead in each division, each zone, or each school would lead to under representation of orphans , girls, and to children neither of whose parents had any secondary schooling . Of the 2015 who took the test, 34.8% were orphans, 31.2% had no parental secondary education, and 44.67% were female. In the entire sample of 5200 students, these groups make up 34.4%, 34.0%, and 48.37% of the population, respectively. Table 3.1 shows the percentages of awards that would go to each of these populations along with average KCPE scores for the winners with those characteristics for different ways of dividing the bursaries.

All three of these populations are underrepresented in the awards compared to their population proportions (34.8% orphan, 31.2% with no parental secondary education, and 44.67% female). This is especially true for females, who would only receive 26 to 28% of the awards if the competition were not conducted separately for each gender. Orphan status does not seem to present as large of a disadvantage (1 to 2 percentage points underrepresented)

in the competition as does the socio-economic status measure of having parents without any secondary education (6 to 9 percentage points underrepresented). Although it is not true in every instance, it does seem that the finer we slice the competition groups, the more disadvantaged students would benefit, although this increase is not very large at all, and the average winning KCPE score becomes considerably lower.

Obviously, stratifying the award by a certain characteristic (gender, for example) would guarantee those with that characteristic would benefit. Table 3.2 shows the results from running the scholarship program separately for boys and girls. Since female KCPE scores are lower on average, this regime lowers the average score of the winners. This method also fails to benefit those that are not directly targeted (orphans and those without parental secondary education).

Giving awards separately by gender to the top percentiles ensures that girls receive awards nearly equal to their population proportion (45%), but it does not direct more awards to orphans or those with low parental education. We could also simply reserve fifty percent of awards for girls and see whether this would direct more awards to the underprivileged. Our simulations so far have resulted in between 201 to 273 scholarships using the top 10% and 304 to 467 using the top 20%. Table 3.3 shows the results if we chose to give 300 total scholarships. We divide the 300 scholarships evenly between boys and girls in the one district, then also between the two divisions, the 8 zones, and the 74 schools. Ties sometimes prevent the awards from being split exactly equally between boys and girls.

We can see that the trend is the same with this method of targeting—average KCPE scores decline and wider geographic dispersal does not effectively provide more scholarships to groups that are not directly targeted. We could more directly target disadvantaged groups and give awards separately to the top 10% of orphans, children whose parents have no secondary education, and girls, by school, zone, division, and the district as a whole. It is worth bearing in mind that unintended consequences may arrive from this type of direct targeting. For example, if a scholarship program used iron roof ownership as a measure of wealth that made children ineligible for scholarships, in the long run it is possible fewer people would buy iron roofs, and in the short run, reporting of iron roof ownership would likely decrease. Orphan status and parental secondary education are unlikely to be altered much in truth as a result of the scholarships, but the reported numbers may change to some degree since secondary school is such a large expense for most families. Table 3.4 shows the results of giving awards separately for each characteristic and by method of geographic breakdown.

This method of having separate competitions for orphans, those with no parental secondary education, and females definitely accomplishes the goal of giving more awards to these disadvantaged groups. As the sample size increases, the percentages of awards won by these groups should approach their population proportion, but with rounding and small samples, the disadvantaged groups are often overrepresented, the more so the finer the geographic stratification is done. Running the scholarships by school would ensure that as high as 45% of the winners were orphans and 45% of the winners' parents had no secondary schooling,

despite these groups making up only 34 and 31 percent of the population, respectively.

Appendix 1 repeats Tables 3.1, 3.2, and 3.4 using the data from both the GSP and the KLPS. The GSP data is similar to that from the KLPS; it is also a tracking survey of students from the Busia district of Western Kenya. More details on the GSP data are below in section 3.3. In this population, girls make up a much larger percentage of the population because they are 100% of the GSP sample. Using the combined sample, we have 2941 test-takers, of whom 31% are orphans, 32% have parents with no secondary education, and 62% are female. The larger trend of more benefit to disadvantaged groups through more specific geographic distribution remains similarly small, while direct targeting is again far more effective. Thus we can see that data from both the KLPS and the GSP indicate that running bursary programs separately by school as opposed to district does little to direct more awards to the underprivileged. In most cases only 1 or 2 percent more of the awards go to the underprivileged, while average winning KCPE scores decrease by as much as 20 points.

3.3   Primary School Scholarship Program

The next two sections look at the impacts of a specific primary school scholarship program. We examine two types of impacts, first, the impact of attending a school where the scholarship competition is offered, and second, the impact of winning one of the scholarships. Although the policy debate in Kenya now focuses on scholarships for secondary schooling, some light can be shed on the subject by analysis of data from a program at the primary level.

In 2001 and 2002, a merit scholarship program for girls was conducted in primary school standard 6 in 127 of the schools in the Busia and Teso districts of Western Kenya. The scholarship provided the cash equivalent of two years' worth of school fees and uniform costs, split between awards to the school in the girls' names and awards directly to the girls' families. Girls participating in the program took the annual mock exam in preparation for taking the KCPE after standard 8. The scholarship was awarded in each of the two years by International Child Support—Africa (ICS) to all standard 6 girls who scored in the top 15% of girls in the program schools in their district. Two hundred scholarships were awarded each year. Half of the 127 schools in the districts were randomly selected to have the scholarship competition, enabling reliable statistical comparison of those eligible for scholarships with those who were ineligible. Boys were not eligible for the scholarship in any school.

Girls attending schools eligible for the scholarship saw significant test score gains, and there is suggestive evidence that girls from program schools with low scores on pre-tests (that were thus unlikely to win the award) saw gains. And despite being completely ineligible, there was suggestive evidence that boys saw positive benefits from attending scholarship schools in the short run in the larger of the two program districts, possibly due to decreased teacher absenteeism. Analysis of the effects of attending program schools in the smaller

district was complicated by attrition perhaps due in part to a lightning strike at one of the schools. In Busia district, however, test scores improved by 0.2 to 0.3 standard deviations for girls in schools with the scholarship program[Kremer et al., 2009]. These effects were significantly larger than those found in Angrist and Lavy [2002], analysis of a somewhat similar merit scholarship program in Israel.

We are now interested in examining the medium-run effects of the program, both for girls who simply attended a scholarship school as well as for those who won the scholarship itself. Starting in late 2005, we began attempting to locate and survey all girls who were in standard 6 in 2001 and 2002 from the 68 program and comparison schools in Busia district. This gave us a sample of 3237 girls. Tracking concluded in February 2007 after successfully surveying 1,757 girls. Including deaths and refusals, the tracking team managed to locate a total of 1793 of the girls. Midway through tracking, one-fifth of those remaining were randomly selected for intensive tracking, of which 333 were successfully surveyed. Altogether this gave us an effective tracking rate of approximately 80%. The girls who were located were administered a survey containing questions on education, health, and labor market outcomes as well as cognitive tests and information on the family, crime victimization, social capital, and religion, among other things. This survey is similar in many ways to the Kenyan Life Panel Survey (KLPS), which was also recently conducted in two divisions of the Busia district.

3.4   Effect of Attending a Scholarship Program School

In the short-run, there was at least suggestive evidence that girls of all performance levels who attended a school in Busia district where the scholarships were offered saw test-score gains. We are now interested in seeing if benefits from attending a bursary program school persist in the long term. Since ICS randomly selected 68 of the 127 to be given the scholarships, we can use simple techniques from randomized evaluations or experiments to analyze the effect of attending a scholarship school without worry of contamination from unobservable characteristics. For this to be true, the baseline characteristics of the treatment and control groups should be balanced, which is in fact the case. For more information on the sample balance, please refer to Table 3 of Kremer et al. [2009].

Another possible concern is that differential sample attrition will affect our estimates. Due to the high mobility of our sample (girls have married and moved away to Nairobi or Uganda and even the USA without leaving any contact information, for example) there will inevitably be some attrition. With the 1757 surveyed girls who have been incorporated into the data, we have an effective tracking rate of 80.6% amongst treatment school girls, and 81.5% amongst control school girls. The high rate of successful tracking and the insignificant difference between control and treatment tracking rates should allay fears of differential attrition biasing our estimates and removes the need for more complicated methods of analysis such as the non-parametric trimming method in Lee [2002].

Analysis with the tracking data yields interesting results regarding the efficacy of the scholarship program. Measures of highest standard completed, whether or not the girl is still in school, KCPE score, whether or not the girl is idle, self-reported happiness, mental health, suffering a major health shock, having been married, having been pregnant, height, and weight all show no statistically significant effect of treatment (that is, attending a school in which the scholarship competition was held). Regressing the dependent variable on treatment alone sometimes yields significant results, but this is not robust to controlling for variables that one might expect to be strongly correlated with outcomes (age, average school scores from tests prior to the scholarship program implementation, year of survey). There is, however, some suggestive evidence that girls from treatment schools perform better on the arithmetic test we administered as well as the Raven's Progressive Matrices spatial pattern recognition test.

We show below in Table 3.5 three specifications modeling the effect of attending a program school on highest school standard completed. It should be noted that balanced control and treatment samples should mean that the coefficient on the treatment effect would remain fairly stable when additional controls are added. This is indeed generally the case, as the sample is well-balanced as can be seen in Table 3 of Kremer et al. [2009], except that the background characteristics included as extra controls are missing for some of the sample. The first specification is a simple analysis controlling for only treatment status, the second includes controls for baseline school test score, age, as well as cohort and year of survey controls. Program treatment appears to have a moderate positive but insignificant effect on education attained in our specifications—almost a tenth of a year of schooling—but again the coefficient is not significant. When we include interaction terms in the third specification, we also see that the coefficient for mean school test score interacted with treatment is positive and significant, suggesting that students in better schools benefited more from the program than otherwise equal students in schools with lower average test scores. These estimates give evidence that medium-run impacts of the bursary program on later educational attainment are small to non-existent. One possible explanation for this is that since the girls are still young and a high proportion (62%) of the girls are still in school, those who benefited may not yet have had time to distinguish themselves from their peers.

Table 3.6 shows the same three specifications, only with a student's normalized Raven's Progressive Matrices score as the dependent variable. This series of regressions indicates that having participated in the scholarship competition increased a girl's score on this test by 0.15 standard deviations. In regressions not shown, the increase in score on the arithmetic exam is found to be very similar. Theses increases are roughly one half of the increase in test scores found after the initial intervention. So it seems that there is limited evidence of smaller but persistent test-score gains from attending a school which participated in the scholarship competition.

The regressions also show that age at time of treatment is strongly negatively correlated and test scores before the program are strongly positively correlated. This is as one might expect since older girls have generally repeated a standard due to poor grades or poor finances

and younger girls are generally more on-track in their education and could be expected to attain more schooling.

3.5 Effect of Winning a Bursary/Scholarship

Another method through which the ICS scholarships could have had an impact is by directly benefiting the winners themselves. First we will briefly examine what types of girls won the awards. Table 3.7 shows the average of several demographic and socio-economic characteristics of winners compared to non-winners, most of which it could be argued were entirely pre-determined before the contest, or, failing that, not significantly affected by the program.

As one might have guessed, ICS scholarships went to girls who were better off on average. Their parents were more educated and more likely to still be alive, they were more likely to have electricity and a latrine in the home, and the girls themselves were likely to be a little younger. As indicated by the large t-statistics in the rightmost column, all of the differences are statistically significant. Of the winners, 18% are orphans, compared to 24% of the non-winners. Of the 165 winners surveyed in the data, 30% of their parents have no secondary education, 10% have electricity, 72% live in a home with an iron roof, and 99% live in a home with a latrine.

On average the winners and non-winners differed along several observable characteristics, but we can use regression discontinuity analysis to compare those that barely won the award to those that almost won. The main idea behind this type of analysis is to compare those just below the top 15% scholarship cutoff to those just above the 15% scholarship cutoff. We would argue that since girls did not know where the 15% cutoff would be when they took the test, there should be no endogenous sorting around the winning cutoff. Thus within a small bandwidth around the cutoff, winning was essentially random, and the unobservable characteristics of barely-winners and barely-non-winners should be the same in expected value, enabling easy comparison. By controlling for a girl's score on the competition test linearly, squared, and cubed, all that remains should be due solely to winning the scholarship. For a more detailed explanation of regression discontinuity analysis, see DiNardo and Lee [2004] or Lee [2005].

This analysis did not yield strongly significant results. Table 3.8 shows the regression discontinuity models for the effect of winning a scholarship on years of educational attainment. In naive specifications that merely regress outcomes on winning, such as in column 1, winning is strongly positively correlated with higher outcomes. However, this is as one should expect, and should not be given a causal interpretation. Girls who won the scholarship were (as shown above in Table 3.7) from wealthier homes and could thus be reasonably expected to have higher outcomes regardless of the scholarship. The second and third columns show better specifications that control for the score on the scholarship-determining test. The second controls for a cubic function of the test score, while the third controls for different cubic functions of the test score on either side of the scholarship cut-off. Here, the coefficient on

winning is insignificant, and not even consistently positive. Even if we simply control for a linear function of the test score instead of a cubic, the coefficient on winning the scholarship immediately becomes insignificant, indicating that winning the scholarship by itself did not lead to a significant increase in schooling.

One explanation for this is that although the $38 from winning is a sizable amount to the average Kenyan family, perhaps the winners were from wealthier families and were already able to afford primary schooling, or perhaps as above, since the majority of all girls are still in school with little variation in the sample, winners have not yet been able to distinguish themselves. The sample consists of girls that are presently between the ages of 13 to 23, with most between the ages of 16 to 20. Not a single one of these girls had completed secondary school by the time of the survey. So it could be that differences will appear between the treated and comparison groups eventually, but have yet to do so since the girls have not had time to differentiate themselves significantly in their accomplishments. For example, the average highest standard completed amongst girls not in school is 7.43 while amongst girls still in school it is only 8.15. Over 62% of the sample was still in school at the time of the survey, and it is likely that for some, being out of school is a temporary condition, and once health or finances improve the students will return to school.

Similar analysis has been conducted for the same set of health, fertility, education, and testing variables as in the previous section, and they mostly exhibit the same pattern—girls who won scholarships have higher measured outcomes, but this cannot be given a causal interpretation. Once their score on the competition exam is included, the effect of winning the scholarship by itself becomes insignificant. Also it should be noted that we were also able to effectively survey 80.1% of the girls who did not win individual scholarships, but 94.1% of the girls who did win scholarships. While we did not observe differential attrition with regards to treatment and control, we were in fact able to survey a significantly higher portion of the girls who won scholarships, perhaps because they appreciated the work of the NGO and made themselves more available to our field staff. Although it does not seem there are statistically significant gains from the scholarship by itself, were we to have found any, we would need to carefully consider potential biases from this differential attrition.

3.6 Conclusion

Merit scholarships are appealing to many, because, by definition, they go to those who earn them. They may not, however, go to very many underprivileged children, who may have had fewer opportunities to excel, creating an equity concern. If wealthier children were geographically clustered, disadvantaged children could benefit from running bursary competitions separately at each school rather than in the district as a whole. Our research indicates that within the geographic area we examine, dispersing merit scholarships more broadly geographically would not help much to steer awards to underprivileged students, and doing so would greatly reduce the average winning KCPE score. The increase in equity is not

nearly as great as what could be achieved by direct targeting rules. Also, using tracking data from the GSP follow-up survey, there is some suggestive evidence that a school's participation in a primary school scholarship program leads to moderate gains in test score performance for female pupils (0.15 standard deviations on arithmetic and Raven's Matrices tests), but there is not much evidence of higher level of school attainment, either by attending more schooling or having attended any secondary schooling. There is little evidence that the scholarship winners benefited from the scholarship itself as measured along similar dimensions. Caution should be taken when interpreting these results, as a high proportion of the sample is still in school, or still of the age where they may attend more schooling, so it is entirely possible that over a longer period of time, those who benefited from the competition, or the scholarship itself, may begin to distinguish themselves.

Table 3.1: Targeting Impact of Different Rules for Allocating Scholarships

Targeting Impact of Different Rules for Allocating Scholarships

Scholarships Awarded to Top 10% of KCPE Scores in Each	Percent of Awards Won by Students with Given Characteristics		
	Orphans	No Parental Secondary Education	Girls
School	32.05	23.93	28.21
Zone	28.37	23.08	27.88
Division	30.43	22.22	26.09
District	30.85	22.89	26.87

Top 10% in Each	Average KCPE Score for Winners with Given Characteristics			
	Orphans	No Parental Secondary education	Girls	Overall
School	338.73	338.08	336.72	342.24
Zone	350.57	350.68	348.38	352.13
Division	350.15	352.73	351.47	353.23
District	350.73	352.73	351.61	354.15

Top 20% in Each	Percent of Awards Won by Students with Given Characteristics		
	Orphans	No Parental Secondary Education	Girls
School	34.6	26.8	30
Zone	35.4	23.2	28.5
Division	34.3	23.2	28.6
District	32.6	22.2	30

Top 20% in Each	Average KCPE Score for Winners with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	Overall
School	324.8	321.1	321.8	327.1
Zone	328.7	331.8	331.4	333.4
Division	330.4	332.6	332.3	334.3
District	339.1	342.1	338.7	342.6

Notes: The far left-hand column indicates the geographic breakdown of the awards (i.e., awards go to the top 10% in each school, or the top 10% in each zone, or division, etc.). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.

Table 3.2: Targeting Impacts When Awards Done Separately By Gender

Targeting Impacts When Awards Done Separately By Gender

Scholarships Awarded to Top 10% of KCPE Scores in Each	Percent of Awards Won by Students with Given Characteristics		
	Orphans	No Parental Secondary Education	Girls
School and Gender	32.6	23.8	46.2
Zone and Gender	32.1	23.6	44.3
Division and Gender	29.6	21.4	44.2
District and Gender	31.4	21.6	45.1

Top 10% In Each	Average KCPE Score for Winners with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	Overall
School and Gender	329.9	327.3	317.1	331.9
Zone and Gender	343.9	345.4	335.2	348
Division and Gender	348.9	351.7	338.4	351.2
District and Gender	347.6	351.5	338.7	351.5

Top 20% in Each	Percent of Awards Won by Students with Given Characteristics		
	Orphans	No Parental Secondary Education	Girls
School and Gender	32.1	25.3	45.4
Zone and Gender	31.3	23.8	44.7
Division and Gender	31.9	23.5	44.6
District and Gender	31.3	22	44.7

Top 20% in Each	Average KCPE Score for Winners with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	Overall
School and Gender	320.4	317.3	305.5	320.3
Zone and Gender	329.8	328.3	317.2	330.7
Division and Gender	330.1	330.1	319	332.1
District and Gender	338	339.7	327.7	340.8

Notes: The far left-hand column indicates the geographic breakdown of the awards (i.e., awards go to the top 10% of boys in each school and the top 10% of girls in each school, or the top 10% of boys in each zone and the top 10% of girls in each zone, etc.). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.

Table 3.3: Targeting Impacts When Half of Awards Reserved for Girls

Targeting Impacts When Half of Awards Reserved for Girls

Percent of Awards Won by Students with Given Characteristics

Awards Divided Evenly amongst Each	Orphans	No Parental Secondary Education	Girls
School	32.4	24.4	50.5
Zone	32.7	25.7	50.5
Division	34.9	23.9	49.8
District	31.7	23.3	50

Average KCPE Score for Winners with Given Characteristics

Awards Divided Evenly amongst Each	Orphans	No Parental Secondary Education	Girls	Overall
School	325	319.8	310.8	325.6
Zone	331.7	330.1	319	334.9
Division	331.6	335.4	323.2	337.1
District	337.4	337.2	324.9	339.7

Notes: In this table half of the awards are reserved for females, and the awards are evenly divided between different geographic regions (that is, two to each gender in each school, 18 to each gender in each zone, 75 to each gender in each division, or 150 to each gender in each district). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.

Table 3.4: Targeting Impacts for Awards Done Separately for All Characteristics and by Geography

Targeting Impacts for Awards Done Separately for All Characteristics and by Geography

Scholarships Awarded to Top 10% of KCPE Scores from each Gender, Orphan, Secondary status and	Percent of Awards Won by Students with Given Characteristics		
	Orphans	No Parental Secondary Education	Girls
School	44.7	44.9	48.8
Zone	36.2	33.6	45.7
Division	35.6	32.2	44.1
Overall	35.3	31.9	44.4

Top 10% from each Gender, Orphan, Secondary status and	Average KCPE Score for Winners with Given Characteristics			
	Orphans	No Parental S	Girls	Overall
School	289.6	280.3	285.4	298.7
Zone	337.3	330	328.7	342.4
Division	343.9	338.7	336.4	348.8
Overall	345.8	339.7	337.1	349.8

Notes : In this table all awards are also done separately for each characteristic (the top 10% of orphans receive awards, the top 10% with no parental secondary education, and the top 10% of girls in each school, each zone, each division, or each district). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.

Table 3.5: Impact of Attending Scholarship Primary School (in 2001-2002) on Years of Educational Attainment (in 2005-2006)

Impact of Attending Scholarship Primary School (in 2001–2002) on Years of Educational Attainment (in 2005–2006)

	(1)	(2)	(3)
	Basic	Controls	Interactions
Scholarship Program School	0.088	0.087	0.041
	[0.121]	[0.083]	[0.093]
Age at time of treatment (demeaned)		-0.428***	-0.427***
		[0.044]	[0.068]
Mean School Test Score 2000		0.305***	0.163*
		[0.068]	[0.083]
Age at treatment * Treatment			-0.006
			[0.090]
Mean School Score * Treatment			0.291***
			[0.105]
Cohort, Year of Survey Controls	No	Yes	Yes
Observations	1756	1689	1689
R-squared	0.001	0.256	0.262

Notes : * significant at 10%; ** significant at 5%; *** significant at 1%. Heteroskedasticity robust standard errors clustered by program school shown in brackets. Dependent variable is the student's years of educational attainment bythe time of follow-up surveying. Cohort controls differentiate between girls eligible for the scholarship in 2001 or 2002. Controls also differentiate between year of followup survey, 2005, 2006, or 2007.

Table 3.6: Impact of Attending Scholarship Primary School (in 2001-2002) on Normalized Raven's Progressive Matrices Score (in 2005-2006)

Impact of Attending Scholarship Primary School (in 2001–2002) on Normalized Raven's Matrices Score (in 2005–2006)

	(1)	(2)	(3)
	Basic	Controls	Interactions
Scholarship Program School	0.145*	0.148**	0.155**
	[0.078]	[0.059]	[0.065]
Age at time of treatment (demeaned)		-0.350***	-0.389***
		[0.034]	[0.055]
Mean School Test Score 2000		0.205***	0.147*
		[0.043]	[0.075]
Age at treatment * Treatment			0.075
			[0.069]
Mean School Score * Treatment			0.117
			[0.092]
Cohort, Year of Survey Controls	No	Yes	Yes
Observations	1756	1689	1689
R-squared	0.005	0.153	0.156

Notes : * significant at 10%; ** significant at 5%; *** significant at 1%. Heteroskedasticity robust standard errors clustered by program school shown in brackets. Dependent variable is the student's normalized score on the Raven's Progressive Matrices test. Cohort controls differentiate between girls eligible for the scholarship in 2001 or 2002. Controls also differentiate between year of followup survey, 2005, 2006, or 2007.

Table 3.7: Average Characteristics for Winner and Non-Winners

Average Characteristics for Winner and Non-Winners

	Population	Non-Winners	Winners	H_0: diff=0 t statistic
Age at time of treatment	13.78	13.81	13.48	2.817
Mother's Years of Education	6.77	6.63	8.09	-4.127
Father's Years of Education	9.01	8.88	10.18	-3.443
Latrine Ownership	96.20%	95.90%	99.40%	-2.154
House Has Electricity	6.20%	5.80%	10.30%	-2.222
House Has Iron Roof	56.80%	55.00%	72.30%	-4.129
Ownership of House	92.10%	92.60%	87.70%	2.149
Either Parent Deceased	23.30%	23.90%	18.10%	1.63

Table 3.8: Effect of Winning Primary School Scholarship (in 2001-2002) on Years of Educational Attainment (in 2005-2006)

Effect of Winning Primary School Scholarship (in 2001-2002) on Years of Educational Attainment (in 2005-2006)

	(1) Basic	(2) Controls	(3) Interactions
Scholarship Program School	0.026	-0.133	-0.134
	[0.090]	[0.096]	[0.097]
Age at time of treatment	-0.384***	-0.241***	-0.241***
	[0.041]	[0.045]	[0.045]
Mean School Test Score 2000	0.129*	-0.064	-0.066
	[0.072]	[0.095]	[0.095]
Won Scholarship	0.896***	-0.072	0.092
	[0.110]	[0.161]	[0.491]
Competition Test Score (CTS)		0.685***	0.690***
		[0.107]	[0.114]
CTS^2		0.028	0.012
		[0.044]	[0.050]
CTS^3		-0.042**	-0.054*
		[0.020]	[0.030]
CTS*Won Scholarship			-0.416
			[0.771]
CTS^2*Won Scholarship			0.285
			[0.386]
CTS^3*Won Scholarship			-0.036
			[0.068]
Cohort, Year of Survey Controls	Yes	Yes	Yes
Observations	1689	1214	1214
R-squared	0.312	0.354	0.354

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%. Heteroskedasticity robust standard errors clustered by program school shown in brackets. All specification include cohort controls (whether they were in the 2001 or 2002 program), controls for year of survey (2005, 2006 or 2007) and for those in control schools who would have won an award had they attended a treatment school.

Bibliography

Alberto Abadie, David Drukker, Jane Leber Herr, and Guido W. Imbens. Implementing matching estimators for average treatment effects in stata. *The Stata Journal*, 4(3):290–311, 2004.

Armen A. Alchian and Harold Demsetz. Production, information costs, and economic organization. *The American Economic Review*, 62(5):777–795, December 1972.

Stuart H. Altman. Earnings, unemployment, and the supply of enlisted, volunteers. *The Journal of Human Resources*, 4(1):38–59, Winter 1969.

Stuart H. Altman and Alan E. Fechter. The supply of military personnel in the absence of a draft. *The American Economic Review*, 57(2):19–31, May 1967.

Joshua Angrist and Victor Lavy. The effect of high school matriculation awards: Evidence from randomized trials. *NBER Working Paper 9839*, 2002.

Beth Asch and Paul Heaton. Monopsony and labor supply in the army and navy. Working Paper Number 537, Princeton University Industrial Relations Section, October 2008.

Beth J. Asch. Do incentives matter? the case of navy recruiters. *Industrial and Labor Relations Review*, 43:89S–106S, February 1990.

Beth J. Asch, Paul Heaton, James Hosek, Francisco Martorell, Curtis Simon, and John T. Warner. Cash incentives and military enlistments, attrition and reenlistment. Monograph, June 2010.

Colin Ash, Bernard Udis, and Robert F. McNown. Enlistments in the all-volunteer force: A military personnel supply model and its forecasts. *The American Economic Review*, 73 (1):145–155, March 1983.

Charles Brown. Military enlistments: What can we learn from geographic variation. *The American Economic Review*, 75(1):228–234, March 1985.

Jennifer Brown. Quitters never win: The (adverse) incentive effects of competing with superstars. *UC Berkeley ARE Working Paper*, November 2007.

Central Bureau of Statistics Kenya. Kenya poverty mapping book launch, October 2003a. URL `http://www.cbs.go.ke/pressrelease/other/kenyapovertymap14102003.htm`.

Central Bureau of Statistics Kenya. Geographic dimensions of well-being kenya: Where are the poor?, 2003b. URL `http://www.worldbank.org`.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. 2007.

Raj Chetty, Adam Looney, and Kory Kroft. Salience and taxation: Theory and evidence. *The American Economic Review*, 99(4):1145–1177, 2009.

Luke N. Condra, Joseph H. Felter, Radha K. Iyengar, and Jacob N. Shapiro. The effect of civilian casualties in afghanistan and iraq. NBER Working Paper 16152, July 2010.

Charles Dale and Curtis Gilroy. Determinants of enlistments: A macroeconomic time-series view. *Armed Forces and Society*, 10(2):192–210, Winter 1984.

Charles Dale and Curtis Gilroy. Enlistments in the all-volunteer force: Note. *The American Economic Review*, 75(3):547–551, June 1985.

Stefano DellaVigna. Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372, 2009.

James N. Dertouzos. Recruiter incentives and enlistment supply. Rand Corporation, May 1985.

James N. Dertouzos and J. Michael Polich. Recruiting effects of army advertising. RAND National Defense Research Institute, January 1989.

John DiNardo and David S. Lee. Economic impacts of new unionization on u.s. private sector employers: 1984-2001. *Quarterly Journal of Economics*, 199(4):1383–1442, 2004.

Anthony C. Fisher. The cost of the draft and the cost of ending the draft. *The American Economic Review*, 59(3):239–254, June 1969.

Drew Fudenberg and Jean Tirole. Moral hazard and renegotiation in agency contracts. *Econometrica*, 58(6):1279–1319, November 1990.

Scott Sigmund Gartner and Gary M. Segura. War, casualties, and public opinion. *The Journal of Conflict Resolution*, 42(3):278–300, June 1998.

Scott Sigmund Gartner, Gary M. Segura, and Michael Wilkening. All politics are local: Local losses and individual attitudes toward the vietnam war. *The Journal of Conflict Resolution*, 41(5):669–694, October 1997.

Alexander M. Gelber. The supply of military enlistments. January 2007.

Robert Gibbons and Kevin J. Murphy. Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, 100(3):468–505, 1992.

Edward L. Glaeser, David Laibson, and Bruce Sacerdote. An economic approach to social capital. *The Economic Journal*, 112:F437–F458, November 2002.

Michael Greenstone, Richard Hornbeck, and Enrico Moretti. Identifying agglomeration spillovers: Evidence from million dollar plants. *NBER Working Paper 13833*, March 2008.

Dominique M. Hanssens and Henry A. Levien. An econometric study of recruitment marketing in the u.s. navy. *Management Science*, 29(10):1167–1184, October 1983.

Jerry A. Hausman and Gregory K. Leonard. Superstars in the national basketball association: Economic value and policy. *Journal of Labor Economics*, 15(4):586–624, 1997.

Benjamin E. Hermalin and Michael L. Katz. Moral hazard and verifiability: The effects of renegotiation in agency. *Econometrica*, 59(6):1735–1753, November 1991.

Bengt Holmstrom. Moral hazard and observability. *The Bell Journal of Economics*, 10(1): 74–91, 1979.

Bengt Holmstrom. Moral hazard in teams. *The Bell Journal of Economics*, 13(2):324–340, 1982.

Tanjim Hossain and John Morgan. ...plus shipping and handling: Revenue (non) equivalence in field experiments on ebay. *Advances in Economic Analysis and Policy*, 6(2), 2006.

Brian Jacob and Lars Lefgren. The impact of research grant funding on scientific productivity. *NBER Working Paper 13519*, October 2007.

David Karol and Edward Miguel. The electoral cost of war: Iraq casualties and the 2004 u.s. presidential election. *The Journal of Politics*, 69(3):633–648, August 2007.

Anthony C. Krautmann and John L. Solow. The dynamics of performance over the duration of major league baseball long-term contracts. *Journal of Sports Economics*, 10(1):6–22, February 2009.

Michael Kremer and Edward Miguel. Worms: Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217, 2004.

Michael Kremer, Edward Miguel, and Rebecca Thornton. Incentives to learn. *Review of Economics and Statistics*, 91(3):437–456, 2009.

David S. Lee. Trimming the bounds on treatment effects with missing outcomes. *NBER Working Paper T277*, 2002.

David S. Lee. Randomized experiments from non-random selection in u.s. house elections. *Institute of Governmental Studies WP 2005-13*, 2005.

Michael Lewis. *Moneyball: The Art of Winning an Unfair Game*. W. W. Norton Co., 2004.

Ulrike Malmendier and Geoffrey Tate. Superstar ceos. *NBER Working Paper 14140*, 2008.

Robert F. McNown, Bernard Udis, and Colin Ash. Economic analysis of the all-volunteer force. *Armed Forces and Society*, 7(1):113–132, Fall 1980.

Carole Oken and Beth J. Asch. Encouraging recruiter achievement: A recent history of military recruiter incentive programs. RAND National Defense Research Institute, 1997.

William P. Rogerson. Repeated moral hazard. *Econometrica*, 53(1):69–76, January 1985.

Sherwin Rosen. The economics of superstars. *The American Economic Review*, 71(5):845–858, December 1981.

Sherwin Rosen. The theory of equalizing differences. In O. Ashenfelter and R. Layard, editors, *Hanbook of Labor Economics*, volume 1, chapter 12. Elsevier Science Publishers, 1986.

A. D. Roy. Some thoughts on the distribution of earnings. *Oxford Economic Papers*, 3(2): 135–146, June 1951.

Curtis J. Simon and John T. Warner. Managing the all-volunteer force in a time of war. *The Economics of Peace and Security Journal*, 2(1):20–29, 2007.

Gary Smith and Teddy Schall. Do baseball players regress toward the mean? *The American Statistician*, 54:231–235, November 2000.

John T. Warner and Beth J. Asch. *The Economics of Military Manpower*, volume 1 of *Handbook of Defense Economics*, chapter 13, pages 347–398. Elsevier, 1995.

Alexander Wolff. That old black magic. *Sports Illustrated*, January 21 2002.

Appendix A

Scholarships Appendix

Table A.1: Targeting Impact of Different Rules for Allocating Scholarships Using KLPS and GSP

Overall Population Statistics, GSP and KLPS

	Population Proportion	Average KCPE Score
Orphan	0.308	255.8
No Secondary	0.315	253.4
Female	0.621	255.7

Targeting Impacts for Different Geographic Dispersal of Awards, Using Both KLPS and GSP

Percent of Awards Going To Students with Given Characteristics

Scholarships Awarded to Top 10% of KCPE scores in Each	Orphans	No Parental Secondary Education	Girls
School	28.7	27.3	52.8
Zone	25.3	24.4	51.3
Division	26.9	24.3	49.8
Overall	27.7	23.4	46.4

Average KCPE for Winners with Given Characteristics

Top 10% in Each	Orphans	No Parental Secondary Education	Girls	Overall
School	334.1	328.1	328.3	335.9
Zone	347.1	342.6	340	346.6
Division	347.4	245.5	344	349
Overall	349.9	349.9	349.9	352.7

Percent of Awards Going To Students with Given Characteristics

Top 20% in Each	Orphans	No Parental Secondary Education	Girls
School	30.3	28.1	53
Zone	30.1	25.7	51.5
Division	30	24.6	51.4
Overall	28.4	23.9	51.1

Average KCPE for Winners with Given Characteristics

Top 20% in Each	Orphans	No Parental Secondary Education	Girls	Overall
School	321.6	315.4	316.3	322.4
Zone	327	325.3	324.7	329.3
Division	328.6	329.2	327.6	331.2
Overall	338.3	338.6	337.1	340.6

Notes: The far left-hand column indicates the geographic breakdown of the awards (i.e., awards go to the top 10% in each school, or the top 10% in each zone, or division, etc.). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.

Table A.2: Targeting Impacts When Awards Done Separately By Gender Using KLPS and GSP

Targeting Impacts When Awards Done Separately by Gender, Using Both KLPS and GSP

Scholarships Awarded to Top 10% of KCPE Scores in Each	Percent of Awards Won by Students with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	
School and Gender	29.4	26.8	62.8	
Zone and Gender	27.9	24.7	62.2	
Division and Gender	26.3	23.7	62.2	
Gender	26.4	24	66.8	

Top 10% in Each	Average KCPE for Winners with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	Overall
School and Gender	327.9	322.3	320.3	329.4
Zone and Gender	342.2	339.4	335.2	343.9
Division and Gender	346.4	344.7	339.2	347.6
Gender	344.1	342.6	337.2	345.4

Top 20% in Each	Percent of Awards Won by Students with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	
School and Gender	28.8	27	62.6	
Zone and Gender	27.4	26.1	62.4	
Division and Gender	28.3	24.8	62.1	
Gender	26.8	23.6	67.2	

Top 20% in Each	Average KCPE for Winners with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	Overall
School and Gender	318.2	313	309.2	318
Zone and Gender	327.7	323.3	319	327.5
Division and Gender	328.5	327.6	321.9	329.8
Gender	333.5	332.6	326.2	334.5

Notes: The far left-hand column indicates the geographic breakdown of the awards (i.e., awards go to the top 10% of boys in each school and the top 10% of girls in each school, or the top 10% of boys in each zone and the top 10% of girls in each zone, etc.). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.

Table A.3: Targeting Impacts for Awards Done Separately for All Characteristics and by Geography Using KLPS and GSP

Table 4A: Targeting Impacts for Awards Done Separately by Targeted Characteristics and by Geography, Using Both KLPS and GSP

Scholarships Awarded to Top 10% of KCPE Scores from Each Gender, Orphan, Secondary Status and	Percent of Awards Going To Students with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	
School	42.5	44.8	64.8	
Zone	33.7	34.6	64.3	
Division	31.6	32.6	61.9	
Overall	31	31.7	61.7	

Top 10% from Each Gender, Orphan, Secondary Status and	Average KCPE Score for Winners with Given Characteristics			
	Orphans	No Parental Secondary Education	Girls	Overall
School	285.5	279.5	288.1	296.3
Zone	329.4	324.7	327	336.6
Division	341.5	336.5	338	345.8
Overall	344.7	340.2	341.3	348.4

Notes : In this table all awards are also done separately for each characteristic (the top 10% of orphans receive awards, the top 10% with no parental secondary education, and the top 10% of girls in each school, each zone, each division, or each district). The percentages indicate, for each geographic method of breaking down the award, the proportion of awards that would be awarded to persons with the characteristic listed at the top. The test scores indicate the average KCPE score for winners with the listed characteristic, again depending on how we geographically distribute the awards. For reference, the population of test-takers is 34.8% orphans, 31.2% have no parental secondary education, and 44.67% are female.