Essays on the Law and Economics of Public Institutions

by

Jonathan Blake Petkun

B.A., Swarthmore College (2007)
J.D., Yale University (2019)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Economics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2020

Author . . . .

Department of Economics
May 15, 2020

Certified by . . . .

Jonathan Gruber
Ford Professor of Economics
Thesis Supervisor

Certified by . . . .

Heidi Williams
Charles R. Schwab Professor of Economics, Stanford University
Thesis Supervisor

Accepted by . . . .

Amy Finkelstein
John & Jennie S. MacDonald Professor of Economics
Chairman, Departmental Committee on Graduate Studies

THIS PAGE INTENTIONALLY LEFT BLANK

Essays on the Law and Economics of Public Institutions

by

Jonathan Blake Petkun

Submitted to the Department of Economics
on May 15, 2020, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Economics

Abstract

This thesis consists of three chapters on the legal and economic organization of large-scale public institutions. Each chapter uses the tools of applied microeconomics in order to explore how the legal and organizational structure of public institutions affects the behavior of actors inside and outside of the organization as well as the quality of public goods provided. In other words, my thesis asks the question: how well do our public institutions function, or for whom do they function? Chapter 1 focuses on the inner-workings of the United States civil justice system; chapters 2 and 3 focus their attention on the U.S. military.

In Chapter 1 I study a procedural reform in the U.S. federal trial courts. Recent court reform efforts in the U.S. and elsewhere have focused on speeding up what are perceived to be slow and burdensome civil justice systems. I study a Congressionally-enacted reform known as the "six-month list," which uses social pressure to incentivize federal judges to decide cases more quickly. I construct an original dataset of nearly 500,000 federal district court motions—representing the approximate universe of summary judgment motions in federal civil cases for the period 2004-2014—and I exploit quasi-random variation in exposure to the six-month list in order to assess the causal effects of the six-month list on both the speed and quality of adjudications. My results indicate that the six-month list does indeed improve speed, though the effect is heterogeneous across judges, with judges who are young, non-white, or female being among the most responsive. Meanwhile, I find only mixed evidence of effects on the quality of adjudications. I interpret these results as consistent with a model of judicial behavior that

combines elements of career concerns, procrastination, and multitasking.

Chapter 2 maintains a focus on public sector personnel policy, bit it shifts contexts from the U.S. courts to the U.S. military. In this chapter, which is the product of joint work with Christina Patterson and William Skimmyhorn, we study how the structure of common retention incentives affects employee quality in the U.S. military. This complements the existing literature on the determinants of public sector worker quality, which has primarily focused on levels of compensation rather than the structure of personnel policy and other non-wage incentives. We combine administrative data with quasi-random variation to find that low-ability soldiers are relatively more responsive to both lump-sum bonuses and early retirement benefits, and both effects are large enough to lower the organization's average ability level. We provide suggestive evidence that neither access to credit nor differences in personal discount rates explain these selection patterns.

In Chapter 3, joint with Paul Goldsmith-Pinkham, we assess the potential for pecuniary externalities relating to military housing allowances. In providing housing to its troops, the U.S. military chooses between direct in-kind provision (i.e. on-base barracks and family housing) and cash transfers (i.e. lump-sum housing allowances). In areas with a high military share of the overall population, military housing policies can have potentially significant impacts on the local civilian housing market. Anecdotally, some worry that military housing allowances drive up local housing prices, making it difficult for civilians to compete with their military neighbors for affordable housing. We combine panel data on the evolution of ZIP code-level military housing allowances and rental and house prices with plausibly exogenous changes to the military's housing allowance formula in order to identify pecuniary externalities. We find suggestive evidence that increases to local military housing allowance rates generate sizeable pecuniary effects, with a 1% increase in military housing allowances leading to a 0.25% increase in local house prices in areas with a nearby military base.

JEL Classification: H11, K41, J45

Thesis Supervisor: Jonathan Gruber
Title: Ford Professor of Economics

Thesis Supervisor: Heidi Williams
Title: Charles R. Schwab Professor of Economics, Stanford University

Acknowledgments

My dissertation marks the end of a long graduate school journey, but in a lot of ways, it feels like just the beginning. I have learned a tremendous amount in my seven years as a graduate student and law student. None of it would have been possible without the support of my community, both at MIT and beyond.

First, I am deeply indebted to my primary advisors, Jon Gruber and Heidi Williams. They cannot know how much I look up to them for their wisdom, patience, kindness, and insightfulness. I do not claim—but I do hope—that their fingerprints are evident throughout this document. I am also grateful for the mentorship and encouragement of a number of other faculty at MIT, especially Amy Finkelstein. I am also grateful to the fantastic staff of the MIT economics department, past and present—especially Emily Gallagher, Deb Jamiol, Gary King, Eva Konomi, Kara Nemergut, Kim Scantlebury, and Beata Shuster. In addition to their tireless work on behalf of the department and its faculty, they are also amazing advocates and friends to the bedraggled grad students.

I am grateful to a number of brilliant and gracious faculty mentors at Yale Law School, including Christine Jolls, Ian Ayres, Harold Koh, and Zach Liscow. I am especially grateful to Judith Resnik, whose ability to connect rigorous analysis with compassion for others is truly inspiring. I am similarly grateful for the mentorship of the clinical faculty at Yale, especially Mike Wishnie, Aaron Wenzloff, and Renee Burbank. They make me proud to be connected to the legal profession. I am also indebted to the one-of-a-kind clients who I was lucky enough to represent as part of the Yale Veterans Legal Services Clinic. Their relentless advocacy continues to inspire me.

I am grateful to teachers and mentors in economics outside of the walls of E52, especially Mark Kuperberg, Steve O'Connell, Philip Jefferson, Michael Greenstone, Mike Frakes, and Bill Skimmyhorn. I've benefited beyond measure from the expert advice of Jediphi Cabal, Eric Zwick, Dan Hammer, Isaac Sorkin, and Paul Goldsmith-Pinkham, whom I don't know whether to count as mentors or friends. I also want to thank the many other teachers who have impacted me along the way, especially Yi Suan Barrow and David Nieslanik of Beaverton High School. My teachers at Beaverton taught me to take a first derivative, and they also taught me to think critically and compassionately about real world issues—that is to say, they taught me the building blocks of economics.

I cannot begin to express my gratitude to my cohort-mates at MIT. Grad school was worth it if only to share a brief time together with people like Ray Kluender, Ben Marx, Christina Patterson, Otis Reid, Michael Stepner, Dan Waldinger, Ariel

Zucker, and Chishio Furukawa. They dragged me through coursework, they improved my ideas in third-year and beyond, and they have been caring friends throughout. Our entire cohort was truly special, and I am unspeakably proud of all that they will accomplish. I've also been fortunate to learn from students in the cohorts above and below me, especially Ryan Hill, Maddie McKelway, Jane Choi, Colin Gray, Claire Lazar, Brendan Price, Alex Bartik, Sally Hudson, Manisha Padi, Manasi Deshpande, and Kyle Greenberg, not to mention the several graduate macro TAs to whom I owe my continued existence as an economist. My life as a graduate student was all the more enriched by relationships with friends and neighbors in Somerville (Yvonne, Katie, MJ, Dan, Tanya, and Catherine) and Cambridge (John, Malvina, Dava, Gui, Kristen, and all of the students and GRTs of Baker House). It was difficult to leave my community at MIT for law school, but I was beyond fortunate to encounter another group of brilliant and kind friends and running partners at Yale: especially Joseph Schottenfeld, Ashraf Ahmed, Michael Morse, Meghan Brooks, Alyssa Peterson, Jonathan Cohen, Arjun Mody, Sam Frizell, Samantha Peltz, and Diane de Gramont. I am also grateful to the friends who have known me the longest, and who continue to anchor me: Susannah Bien-Gund, Alix Gould-Werth, Bree Bang-Jensen, Rachel Ackoff, Nathaniel Peters, Lake Bookman, Martin Williams, and many others.

I am proud to be joining the economics and legal professions, but it's hard to beat the pride that I take in my brief career as a U.S. Marine. No organization is perfect, but at its best, the Marine Corps is a collection of dedicated, compassionate servants. I am forever transformed by the Marines I served alongside, especially Robert Black, Mike Thompson, Chris Motz, Chris Goepper, Blake Rice, Colonel Kenneth Kassner, and all of the other Marines, sailors, and family members of 3d LAR ("Wolfpack") and MTT 0731 ("Lumberjack").

I am blessed by the unconditional love and support of my mom, Mary Alison Leatart, my dad, Robert Petkun, and my brothers and sisters and their partners: Matt, Jessica, Jordanna, Ali, Steve, Adam, and Wynn. As if being born into this family weren't enough to convince me of my good fortune, I've also hit the jackpot when it comes to an amazing set of in-laws, including my mother-in-law, Carrie, father-in-law, Jim, and brother- and sister-in-law, Josh and Carly.

Finally, I am most grateful for the friendship, love, and patience of my wife, Mara Redlich Revkin. She's not only an amazing partner, but she is also the best example of a scholar that I know. She inspires me with her curiosity, analytical rigor, and compassion for others. Law school and graduate school were both infinitely better for having her by my side, and I can't wait to see what's ahead for us, academically and otherwise. It was also her idea to adopt our cat, Nimr

Revkin-Petkun, who deserves (and daily receives) the utmost acknowledgment.

My studies and research at MIT were made possible through the generous support of the National Science Foundation, the Department of Veterans Affairs, and the MIT Economics Department. I am tremendously grateful for the public dollars that have contributed to my education. I am also grateful for individuals and agencies that have shared their data and expertise, especially Jonah Gelbach, and the faculty and staff of the Office of Economic and Manpower Analysis (OEMA) at West Point.

. . . .

This thesis is dedicated to Cara Nickolaus, whose memory is indeed a blessing to all who were lucky enough to call her a classmate and friend. It is also dedicated to the most vulnerable among us—working families; incarcerated people; civil and criminal litigants; service members and veterans; and all others—who deserve a government that works for and with them.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

Chapter 1

Can Judges Be Shamed? Evidence from the "Six-Month List"

1.1 Introduction

Recent court reform efforts in the United States have focused on speeding up what is perceived to be a slow and burdensome federal civil justice system. But how is speed best achieved, and at what cost to other goals of the civil justice system? This paper offers one of the first empirical analyses of a civil justice reform initiative, known colloquially as the "six-month list," that uses social sanctions to incentivize judges to prioritize faster adjudications. This paper poses two questions. First, can social sanctions be an effective means of promoting judicial efficiency? And second, when court reform efforts put speed first, how, if at all, does that affect the *quality* of adjudications?

Article III federal judges enjoy life tenure and protected salaries, but that does not mean they are above reproach. In March 2017 then Chief Judge Louis Guirola of the United States District Court for the District of Southern Mississippi took the extraordinary step of temporarily relieving a fellow district judge from taking on new civil cases. Citing a backlog of more than fifty motions pending six months or more and twenty-four cases pending three years or more, the Chief Judge ordered that all new cases initially slated for his lagging colleague be reassigned to one of the court's senior judges.[1] This came after repeated admonishments from the U.S. Court of Appeals for the Fifth Circuit.[2] If the federal district judge in Mississippi represents one extreme, then at the other extreme we might find one California superior court judge who, in a 2009 employment discrimination case, granted the defendant's 1,056-page motion for summary judgment without, apparently, reading it.[3]

The above examples demonstrate two intuitions that most of us share about our justice system, both of which are so obvious as to typically go unstated. First, justice should be speedy. And second, if the process is *too* speedy, we begin to worry whether justice has truly been delivered. These examples also serve as a reminder that judges—especially U.S. federal judges—are accustomed to their independence, and they are not easily incentivized.

Accepting that speedy adjudications are important, we are left with the question of how speed is best achieved. One method of achieving speedier resolu-

[1]Jimmie E. Gates, *Judge Wingate still barred from handling new cases*, THE CLARION-LEDGER, Oct. 17, 2017, available at https://www.clarionledger.com/story/news/2017/10/17/judge-wingate-still-barred-handling-new-cases/771901001/.

[2]R.L. Nave, *Justice Delayed?*, JACKSON FREE PRESS, July 17, 2013, available at http://www.jacksonfreepress.com/news/2013/jul/17/justice-delayed/.

[3]Nazir v. United Airlines, Inc., 178 Cal. App. 4th 243, 289 (Cal. Ct. App. 2009) ("[W]hat apparently happened," according to the majority opinion of a panel of California's First District Court of Appeal, "is that the trial court did not read all the papers." Despite reversal, the appellate court seemed to show a degree of sympathy for their lower-court colleague: "While not reading the papers cannot be condoned, it can perhaps be understood, as we hesitate to speculate how long it would take a trial court to meaningfully digest over 2,200 pages of separate statements.")

tions in civil and criminal litigation is through policies promoting active "judicial case management." Policies of this genre tend to place much of the onus for ensuring speedy resolutions on the judges themselves. In recent years, legislatures and judiciaries have enacted various regulatory measures to strengthen judicial oversight and incentivize active judicial management of civil and criminal dockets. A prominent example is the Civil Justice Reform Act ("CJRA") of 1990. In the 1980s, litigants and legal observers complained of long delays in the federal courts, especially in civil litigation. In 1990, Congress responded by passing the CJRA, which aimed to encourage faster processing of civil litigation in federal courts. Among its many provisions, the CJRA mandated the formation of "advisory groups" tasked with identifying and reporting sources of excess cost and delay in civil litigation. However, the law is perhaps best remembered for its imposition of new judicial reporting requirements on members of the federal bench. Since 1991, the CJRA has required federal courts to prepare semiannual reports of all motions pending for more than six months and all civil cases pending for more than three years total.[4] These reporting requirements are known colloquially as the "six-month list."

This paper presents an empirical analysis of two related questions. First, does the six-month list's scheme of social sanctions accomplish its ostensible goal of expediting civil adjudications? And second, does the six-month list have any effect on the *quality* of adjudications? In order to answer these questions, I combine an original large-N dataset of federal district court dockets with a novel identification strategy based on quasi-random variation in exposure to the six-month

[4]Under the law's own sunset provision, the CJRA ostensibly expired in 1997. However, just months before sunset, Congress indefinitely extended the law's hallmark reporting requirements—including the semi-annual "six-month lists." For a discussion of the CJRA's peculiar status post-sunset, see Tobias (2002).

list. I find that the six-month list does indeed improve speed; the summary judgment motions that are most exposed to the six-month list are resolved almost a full month (15%) faster than those that are least exposed, and overall case durations are similarly impacted. I also find considerable heterogeneity across judges, with judges who are young, non-white, or female being among the most responsive to the incentives created by the six-month list. Speedier adjudications notwithstanding, I find only mixed evidence of effects on the quality of adjudications. While I do find modest effects on motion- and case-level outcomes—summary judgment motions that are most exposed to the six-month list are slightly less likely to be granted, and conditional on being appealed, judgments following motions that are more exposed to the six-month list are slightly more likely to be reversed—these results are only marginally significant and not robust to all specifications. My results suggest that—at least among federal judges—social pressure can be an effective substitute for monetary incentives. I interpret these results as consistent with an original model of judicial behavior that combines elements of career concerns, procrastination, and multitasking.

This paper contributes to several distinct literatures across multiple disciplines. First, this paper contributes to a robust literature on judicial management and judicial efficiency. Judith Resnik developed the concept of "judicial management" to describe the ways in which judges intervene in pretrial phases of litigation (Resnik, 1982). Judicial management conveys two potential benefits. First, judicial management has supposedly improved the allocation of scarce judicial resources and accelerated the administration of justice, although this claim is empirically unproven and may not apply equally to appellate and districts courts. Second, it has necessitated the collection and dissemination of new data about the federal courts, with accompanying benefits for transparency and accountability. Resnik argues that these benefits may be outweighed by significant negative externalities. These include the erosion of traditional due process safeguards; a vast expansion in judicial discretion, and with it, the potential for abuse of power; and, the undermining of judicial impartiality in exchange for privacy and informality outside of courtrooms.

Resnik's work raises important empirical questions about the consequences of judicial management for procedural justice. What are the tradeoffs between efficiency and justice? Does speed compromise the fairness of outcomes? Several scholars have attempted to address these questions through historical case studies (Post, 1998) and limited descriptive statistics (Rubin, 1980)[5], but with the exception of Jonah Gelbach's study of summary judgment empirics (Gelbach, 2014), there has been little empirical analysis of judicial management. This paper is, to my knowledge, among the first efforts to identify the causal effects of judicial oversight schemes on the speed and quality of adjudications.

Setting aside potential unintended consequences for the quality of adjudication, speedier adjudications are generally presumed to benefit all parties. That criminal defendants benefit from swiftness of process is presumed by both the Speedy Trial Clause of the Sixth Amendment and the Speedy Trial Act of 1974. Whether civil litigants are entitled to the same degree of promptness is something of an open question. Stephen L. Wasby has argued that procedural delay in the courts can itself amount to a violation of due process (Wasby 1994; Wasby 1997). In fact, courts have frequently recognized promptness as an element of procedural due process in public benefit cases,[6] but courts have been reluctant to prescribe

[5]The CJRA itself spurred a small number of descriptive analyses. See Johnston (1994) and Dessem (1993).

[6]*Fusari v. Steinberg*, 419 U.S. 379, 389 (1975) (holding that "[i]n [the unemployment benefits] context, the possible length of wrongful deprivation of unemployment benefits is an important

rigid timelines.[7]

Whether or not litigants have a procedural right to swiftness, speed can have concrete benefits. It is well documented that the quality and bureaucratic efficiency of public institutions matters for economic growth and development, and courts are no exception. Faster courts reduce transactions costs associated with enforcing contracts and protecting personal and property rights (Acemoglu and Johnson 2005; Visaria 2009; Chemin 2012), all of which are key ingredients to economic development. Moreover, the benefits of speedy adjudications also redound to the litigants themselves. Lengthy administrative and judicial delays can have real and lasting consequences for litigants, who may have to put aspects of their lives on hold while they await resolution of a pending dispute (Connolly and Smith, 1983). In the public benefits context, for example, longer processing times for SSDI applications are associated with lower levels of employment and reduced earnings for multiple years *after* the initial application (Autor et al., 2015). Similarly, evidence suggests that corporate litigants are willing to pay for speedier judicial procedures (Kondylis and Stein, 2018).

This paper also relates to a growing empirical literature analyzing the economics of litigation. Much of this literature focuses on the ways in which court procedures affect the speed and outcomes of justice systems in a variety of jurisdictions. In the Czech Republic, for example, it has been shown that legal reforms

factor in assessing the impact of official action on the private interests," and "the rapidity of administrative review is a significant factor in assessing the sufficiency of the entire process").

[7] *Wright v. Califano*, 587 F.2d 345, 354 (7th Cir. 1978) (reversing a district court's ruling that the SSA must either set a hearing schedule or make interim payments while continuing to review unsuccessful applications for old-age and survivor benefits, and observing that, while "[d]elay in administrative review . . . is a significant factor in assessing the sufficiency of process . . . [it is] not the only factor." *Cf. Cockrum v. Califano*, 475 F.Supp. 1222 (D.D.C. 1979) (holding that delays experienced by some social security applicants were unreasonable and in violation of both the Social Security Act and the Administrative Procedure Act).

allowing judges to follow a simplified set of judicial procedures in adjudications for minor criminal offenses caused increases to both the speed of adjudications and the likelihood that defendants were charged and convicted (Dusek and Montag, 2017). Other papers focus on the allocation of judicial resources. Yang (2016) assesses the impact of judicial vacancies on criminal justice outcomes, finding that prosecutors dismiss more cases during vacancies, and that prosecuted defendants are more likely to plead guilty and less likely to be incarcerated during vacancies. Whereas Yang focuses on variation in judicial resources, other papers have looked instead at variation in judicial caseloads. Huang (2011) and Lavie (2016) exploit an exogenous influx of immigration appeals to show that heavy caseloads caused federal appeals courts to reverse fewer lower-court decisions.

Closely related is a growing literature—including both theoretical and empirical contributions—analyzing the individual behaviors of judges and the group norms and practices of judging. Judge Richard Posner, for example, has developed what he calls a "labor market" model of judicial behavior Posner (2010). However, much of the previous scholarship has focused on judges' political ideology, and the emphasis has traditionally been on appellate courts (e.g., 2013; 2011). My paper is more closely related to a handful of papers that consider how district court judges respond to reputation concerns (e.g., Levy 2005). I contribute what is, to my knowledge, among the first models of judicial behavior that combines elements of career concerns, multi-tasking, and procrastination.

This paper is principally concerned with how to enhance judicial efficiency, but the questions posed here have significance well beyond the rarefied world of the federal judiciary. In particular, this paper may offer answers to a question that has long vexed economists, political scientists, sociologists, and just about anyone who cares about effective government: namely, how to get the most out of bureaucrats. At least two features of public sector work are relatively distinctive. First, bureaucrats are often granted wide discretion to perform tasks that are only broadly defined. From border patrol agents to scientists at the FDA, bureaucrats enjoy a great deal of control over what they do and how they do it. Second, bureaucrats are frequently immune from many of the more traditional workplace incentives. Relative to private sector employers, public sector managers enjoy a more limited array of tools for incentivizing worker behavior. Public sector salaries and benefits are often fixed by lawmakers or regulators, and tenure rules may even inhibit the manager's ability to promote, fire, or reassign. Insofar as federal judges offer an extreme example of both these features, we might think of the judiciary as an ideal laboratory in which to learn more about how non-monetary incentives can be properly deployed in the public sector. This paper contributes to a small but growing body of evidence demonstrating that non-monetary social incentives can, at least under certain circumstances, be used as an effective replacement for more traditional workplace incentives (Gauri et al. 2019; Ashraf et al. 2014; Mathauer and Imhoff 2006).[8]

Long ignored by empirical researchers, the CJRA—and especially the six-month list—has recently become the subject of renewed attention. In addition to my paper, the six-month list is also the focus of a recent article by Miguel de Figueiredo, Alexandra Lahav, and Peter Siegelman Forthcoming. While their paper contributes much to our understanding of the six-month list, their key claims are inherently limited by the correlational—rather than causal—nature of much of their evidence. In particular, they compare cases adjudicated in the weeks immediately preceding the publication of the six-month list against cases adjudicated in other weeks of the year, from which they observe that plaintiff win rates decline in the

[8]See Ashraf and Bandiera (2018) for a review of the literature on social incentives in work.

weeks immediately preceding the six-month list. Importantly, they cannot rule out the possibility that cases decided in the weeks preceding the list are systematically (and *ex ante*) different from cases decided at other times of the year. Furthermore, many of their results are based on a public-use dataset of federal civil case terminations. Since the most important features of the six-month list relate to the adjudication of *motions*, and not cases, they are limited in their ability to make causal claims regarding the effects of the six-month list on motion adjudication. Where they do make causal claims, their results are based on a hand-coded dataset of 781 summary judgment motions filed over a 60-day period in September and August, 2011. Given this small sample size, the authors' causal identification strategy has vulnerabilities. For example, they cannot take into account the potentially confounding effects of seasonality, nor are they able to rule out that there is something aberrant about this particular sample of hand-coded motions. Perhaps most importantly, their relatively small sample size makes it difficult to rule out strategic filing by litigants and their attorneys, which is a crucial assumption for their identification strategy.

I contribute the first empirical analysis of the causal effects of the six-month list based on a large-N motion-level dataset of almost 500,000 motions drawn from almost 300,000 separate cases. Some of my results are, in fact, quite similar to those found by de Figueiredo *et al*. For example, we both find that the six-month list tends to expedite motion and case disposition. However, some of their conclusions are at odds with my own results. For example, they find no evidence of an effect on motion grant rates, denials, or partial denials. In contrast, based on my much larger dataset, I find that exposure to the six-month list slightly decreases the probability of an order granting summary judgment. At the case level, they find correlational evidence that cases decided immediately before the six-month list are more likely to later be remanded back to the district court by the Court of Appeals. When I test this result in a causal framework, I find no evidence of an effect on the remand rate, but I do find evidence of a modest effect on the rates at which lower-court judgments are either reversed or affirmed. Finally, an important difference between the two papers is that mine is the first to contribute evidence on heterogeneity between judges—along dimensions including age, race, and gender—in their responsiveness to the six-month list.

This paper proceeds as follows. Section 1.2 provides some background on the six-month list, including the history of the initiative and details on its design. Section 1.3 discusses a brief conceptual framework for considering the likely effects of the six-month list, with an emphasis placed on how the six-month list has shaped judicial incentives. Section 1.4 describes the original motion-level data that will form the basis of my empirical analysis. Section 1.5 outlines the empirical framework for my analysis, with an emphasis on how I will tease causal effects out of a real-world policy change. Section 1.6 presents preliminary results on the two primary research questions. First, does the six-month list accomplish its ostensible goal of promoting speedy adjudications, and second, what—if any—are its consequences for the quality of adjudication? Section 1.6 also offers evidence on how the effects of the six-month list vary across judges. Section 1.7 offers insights—drawn from my empirical analysis—for the future of civil justice reform. Section 1.8 concludes with a discussion of directions for future research.

1.2 Legal & Policy Background: Where the "Six-Month List" Came From and What It Does

The "six-month list" refers to what is now codified as 28 U.S.C. § 476 (2012), which states that "[t]he Director of the Administrative Office of the United States Courts shall prepare a semiannual report, available to the public, that discloses for each judicial officer the number of motions that have been pending for more than six months and the name of each case in which such motion has been pending."[9] The law was just one component of the so-called Civil Justice Reform Act of 1990 ("CJRA"). Congress was explicit about its intentions. "The purpose of [the CJRA] . . . [was] to facilitate reduction in the delays and expense of civil litigation."[10]

The drafting and passage of the CJRA was swift—from introduction to enactment, it occupied the Congress for less than twelve months (Peck, 1991).[11] However, appetite for civil justice reform had long been growing. In a speech to the American Law Institute on May 17, 1983, Chief Justice Warren Burger decried what he saw as a nation plagued "with an almost irrational focus—virtually a mania—on litigation as a way to solve all problems."[12] Similar sentiments were voiced in the popular media.[13]

Among those listening to the calls for reform was Senator Joseph Biden of Delaware. Beginning in 1988, Senator Biden (then chairperson of the Senate Committee on the Judiciary) commissioned a report from the Brookings Institution and

[9] Judicial Improvements Act of 1990 § 103, 28 U.S.C. § 476 (2012).

[10] H.R. Rep. No. 101-732, at 7 (1990).

[11] See Peck (1991) for a history of the political, economic, and social forces that combined to create the CJRA.

[12] Stuart Taylor, Jr., *Justice System Stifled by Its Costs and Its Complexity, Experts Warn*, N.Y. Times, June 1, 1983, at 1, A1.

[13] See, e.g., Olson (1992).

the Foundation for Change. The request to the Brookings Institution was itself prompted by the results of a survey of judges and attorneys conducted by private polling firm Louis Harris and Associates, Inc. The Harris survey, which sought to identify sources of excess cost and delay in civil litigation, laid particular blame at the feet of "over-discovery" in civil cases. The Brookings Task Force was convened not only to transform the Harris survey results into actionable recommendations for reform, but also to build consensus around those recommendations. Members of the Task Force included "leading litigators from the plaintiff and defense bars, civil and women's rights lawyers, attorneys representing consumer and environmental organizations, former trial and appellate court judges, representatives of the insurance industry, general counsel of major corporations, and law professors." Among the recommendations of the Brookings Task Force was a prototype of what would become the six-month list: "Accordingly, we recommend that the Administrative Office of the U.S. Courts be directed to computerize, in each district, the court's docket so that quarterly reports can be made to the public of at least all pending submitted motions before each judge that are unresolved for more than 30, 60, and 90 days . . . We believe that substantially expanding the availability of public information about caseloads by judge will encourage judges with significant backlogs in undecided motions and cases to resolve those matters and to move their cases along more quickly."[14]

The Brookings Task Force Report informed much of the conversation on Capitol Hill. In fact, an early House Resolution called for implementing a near facsimile of the Task Force recommendations.[15] The proposal was based on the findings that "delays in deciding fully briefed motions contribute to the costs of litigation

[14]See on Civil Justice Reform (1989).
[15]H.R. 3898, 101st Cong. (1990).

by preventing the narrowing of issues, encouraging the parties to conduct unnecessary discovery and requiring rediscovery," and "the reduction of such delays can be encouraged by substantially expanding the availability of public information about backlogs in undecided motions."[16] While the language of the CJRA and its legislative history invoke principles of procedural fairness, Congress appears to have been largely driven by economic motives. Members of Congress observed that "the cost and delays in civil litigation . . . are harmful to both the national economy and to the fairness of our legal system."[17] Finally, after several committee hearings, the CJRA passed both houses of Congress on October 27, 1990.

Under the CJRA, federal courts must prepare semiannual reports of all motions[18] pending in civil cases for more than six months[19] and all civil cases pending for more than three years. Also listed in the semiannual reports are bench trials that have been submitted for six months or more,[20] and, since 1998, bankruptcy and social security appeals pending six months or more.[21] CJRA semiannual reports are posted to a United States Courts website, where members of the public can access approximately eight years of prior reports.[22] Appendix Figure A-1 dis-

[16] *Id.*

[17] *Federal Courts Study Committee Implementation Act and Civil Justice Reform Act: Hearing Before the Subcommittee on Courts, Intellectual Property and the Administration of Justice of the H. Comm. on the Judiciary on H.R. 5381 and H.R. 3898*, 101st Cong. 83 (1990) (statement of Rep. Hamilton Fish, Jr.).

[18] An exception is motions filed in habeas corpus petitions, which are generally exempt from the CJRA's reporting requirements. See Falkoff (2012).

[19] Implementation guidelines give motions a thirty-day grace period before they are considered "pending" for the purposes of the six-month list. As a result, motions actually have at least seven months before they could potentially appear on a six-month list. Implementation guidelines are available at `http://www.uscourts.gov/sites/default/files/data_tables/cjra_na_0930.2017.pdf`.

[20] Judicial Improvements Act of 1990 § 103, 28 U.S.C. § 476 (2012).

[21] See Judicial Conference of the United States (1998) JUDICIAL CONFERENCE OF THE UNITED STATES, REPORT OF THE PROCEEDINGS OF THE JUDICIAL CONFERENCE OF THE UNITED STATES 63 (Sept. 15, 1998).

[22] Available at `http://www.uscourts.gov/data-table-report-names/civil-justice-reform-act-cjra`.

plays an excerpt from the September 30, 2016, CJRA six-month report.

One might be skeptical that the six-month list would actually have any effect on judicial behavior. The six-month list provides little more than a behavioral nudge,[23] and federal judges are hard to incentivize. Article III judges enjoy lifetime tenure and protected salaries, and more generally, they are likely accustomed to being treated with independence and deference.

And yet, the data reveal that the six-month list *does* matter for judicial behavior. Figure 1-1 presents counts of summary judgment motion dispositions by calendar day for the period 2004-2014. Dotted lines mark the two six-month list deadlines of March 31st and September 30th. The effects of the six-month list are immediately discernible: the pace of motion dispositions begins to increase in the months of the reporting deadlines, with a large mass of motion dispositions in the days immediately preceding the deadlines.

Descriptive evidence suggest not only that the six-month list affects *when* judges do their work, but also *how* they do it. Table 1.1 presents descriptive statistics from two samples of summary judgment motions: those ruled on "in the shadow" of the six-month list—that is, in the two weeks immediately preceding a CJRA six-month list—and those ruled on at any other time of the year.

The patterns are striking. Motions decided in the two weeks immediately preceding either of the six-month lists are substantially older (by an average of more than 2.6 months). They are more likely to have been filed in a lawsuit involving at least one *pro se* litigant, and they are more likely to have been filed in a lawsuit

[23]See Thaler and Sunstein (2009): "A nudge, as we will use the term, is any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives. To count as a mere nudge, the intervention must be easy and cheap to avoid." Later in the article I will argue that the six-month list does indeed alter judges' economic incentives, although not significantly and not with any near-term consequences. Moreover, it is both cheap and easy for a judge to avoid an appearance on the list.

where one of the litigants has sought a waiver of court fees (i.e. *in forma pauperis*). Perhaps most striking, the rulings themselves are different. Motions decided in the two weeks prior to the six-month list are less likely to be granted in full (by approximately 2.2 percentage points), they are more likely to be granted in part (by approximately 2.5 percentage points), and they are nearly 6.9 percentage points more likely to be followed by a subsequent appeal to the Court of Appeals. In fact, Table 1.1 corroborates some of the main findings from the de Figueiredo *et al.* article, including their observation that the plaintiff win rate decreases in weeks immediately preceding the six-month list. The patterns shown here, however, are decidedly non-causal. Since there may well be systematic differences between the types of motions decided in the weeks preceding the publication of the six-month list and those decided at other times, the latter group does not constitute an adequate "control" group. Nonetheless, these patterns do suggest that the six-month list has *some* effect on judicial behavior. In the remainder of this article the goal will be to investigate with greater scientific rigor just what the nature of that effect is. First, the following section provides a brief conceptual framework for considering how the six-month list shapes judicial incentives and what the consequences are likely to be for judicial behavior.

1.3 Conceptual Framework: A Model of Judicial Behavior Against the Backdrop of the Six-Month List

According to Judge Richard Posner, "[t]he economic theory of judicial behavior has to surmount two difficulties. One is neglect of psychological factors—of cognitive limitations and emotional forces that shape behavior along with rational calculation . . . [and t]he other . . . is that of identifying the incentives and constraints that shape the vocational behavior of workers whose work is so structured as to eliminate the common incentives and constraints of the workplace. Federal judges cannot be removed from office, short of gross misconduct, and cannot be docked pay, exiled to undesirable judicial venues, or paid bonuses"(Posner, 2010). In the following section, I develop a model of judicial behavior that attempts to address both challenges identified by Judge Posner. My conceptual framework combines elements from three categories of models—namely, career concerns-style models, models featuring procrastination or "present bias," and multitasking models—all of which are common in economics and other social sciences. I will rely on this conceptual framework as I consider which types of incentives are likely to be most effective at influencing judicial behavior, and as I consider the potential trade-offs between various goals of the civil justice system.

The standard career concerns model was in part an attempt to explain the absence of performance-based incentive contracts in many real-world settings.[24] The basic idea is that, even in the absence of performance pay, agents will exert positive effort so long as their "career concerns" (often modeled as future compensation, perhaps due to raises, promotions, or outside offers by competitor firms) so dictate. The model has its roots in Eugene Fama's observation that corporate managers will be influenced by reputational concerns (Fama, 1980). The theory was later formalized in models by Bengt Holmstrom and Milton Harris (Harris and Holmstrom 1982; Holmstrom 1999). Career concern-like forces have been validated in various empirical settings, including among mutual fund man-

[24]Since performance-based pay alleviates many of the moral hazard problems seen in traditional employment/agency contracts, economist have long thought it puzzling that performance-based pay is not more common than it is.

agers(Chevalier and Ellison, 1999) and public utility regulators (Besley and Coate, 2003).

Even for federal judges, whose tenure and salaries are protected, the career concerns model may have some explanatory power. Particularly relevant to judges is the possibility of promotion. US. District Court judges may be motivated by the prospect of elevation to the U.S. Courts of Appeals, and appellate judges may be motivated by the prospects—however remote—of elevation to the Supreme Court. Moreover, while many judges retire from the bench, many others will continue on to a second career in private practice, legal academia, or elsewhere. Judges may therefore be motivated to maintain a good reputation in the eyes of future employers. The career concerns model suggests that judges are likely to comply with the six-month list, lest their non-compliance adversely affect their future career prospects. The career concerns model also predicts a degree of heterogeneity among judges. In particular, the six-month list is likely to generate the largest response from young judges, for whom a promotion is both more likely (since Presidents like to appoint judges who are young enough to sit on the bench for several years to come) and more valuable (since they have more years left during which to enjoy the fruits of a promotion). Since young judges have less professional history, each instance of compliance or non-compliance with the six-month list may also contribute more to their colleagues' posterior beliefs about their competency. I test for and confirm the presence of judge heterogeneity in Section 1.6.3 of the paper.

While reputational concerns are natural, so too is the tendency for procrastination. Procrastination is a common feature of behavioral economics models. The canonical model is attributable to George Akerlof, who observed that "present benefits and costs may have undue salience relative to future costs and benefits"(Akerlof, 1991). Procrastination in the economics literature is typically modeled with time-inconsistent preferences, often with hyperbolic discount functions (Laibson, 1997).

The empirical evidence for procrastination spans a wide variety of real-world settings (see, e.g., Kaur et al. (2015)). Perhaps most relevant to federal judges is a recent working paper by Michael Frakes and Melissa Wasserman (2016), who show that procrastination is commonplace among at least one group of judge-like bureaucrats: namely, examiners for the U.S. Patent Office. They find that patent examiners routinely procrastinate until just before deadlines. They find additional evidence that stricter deadlines are associated with reductions in examiner scrutiny, resulting in higher grant rates for low-quality patent applications. Also relevant is a paper by Raj Chetty, Emmanuel Saez, and Laszlo Sandor (2014), who document procrastination among journal referees. In fact, the experimental intervention studied by Chetty *et al.* is remarkably similar to the six-month list itself. In their study, journal referees are told that their turnaround times will be posted on a publicly-available website. They find that these social sanctions are nearly as effective as cash incentives at reducing delays in peer review.

Insofar as judges respond to the six-month list, a procrastination-style model may explain why. Due to the career concerns described above, the six-month list increases the cost of delay, which is likely to result in faster adjudications on average. Moreover, since present effort is still more costly than future effort, procrastination-style models predict that judges will wait until immediately before the six-month list publication dates to dispose of their overdue motions. This has the potential to generate the patterns observed in Figure 1-1.

Finally, my conceptual framework incorporates additional insights from Bengt Holmstrom and Paul Milgrom's 1991 multitask principal-agent (or "multitask") model (Holmstrom and Milgrom, 1991). The multitask model has quickly become a canonical model in law and economics (and particularly in the field of contract theory); multitask models are especially useful for considering trade-offs between competing goals or priorities.

Relative to the traditional principal-agent problem—wherein an agent performs a single task or makes a single decision on behalf of a principal, often contrary to the principal's best interests—the multitask model is most appropriate for settings in which an agent is simultaneously responsible for multiple tasks or decisions, or in which the agent's single task consists of multiple dimensions. The basic intuition of the multitask model is easily understood in the context of classroom teaching. Consider a school teacher who is responsible for several aspects of his students' enrichment. He is tasked with teaching his students reading, writing, and arithmetic, but he is also responsible for cultivating certain "soft skills," like their ability to work in groups and empathize with others. However, the students are subject to annual standardized testing, and the standardized tests measure only reading and math skills. If the teacher's performance evaluations are tied to his students' test scores, then common sense dictates that the teacher will spend a disproportionate share of his time teaching his students reading and math, and he will give less attention to the so-called soft skills.

The "teaching to the test" problem faced by teachers mimics some of the same incentives imposed on federal district judges. Judges are expected to meet simultaneous goals of speed, accuracy, and fairness. Among these goals, speed is almost certainly the easiest to monitor. In any given case, speed-related metrics can be easily calculated from basic docket information. Judges can be compared in terms of average age of caseload, average time until disposition, average decisional time for various types of motions, and so on. Accuracy and fairness, on the other hand, are much more difficult to monitor, and observable statistics are likely to belie the truth. Two judges may have very different plaintiff win rates, but how can we determine whether either is more fair or accurate? Do the fair and accurate judges grant more summary judgments or fewer? We can look to appellate outcomes—i.e., how often is the judge reversed on appeal, and how often is she affirmed—but most matters are never appealed, and even when they are, appellate judges are no less human than their lower-court colleagues. Under these conditions, where some tasks are more easily monitored than others, high-powered incentives are likely to distort judges' behavior towards the more easily monitored task. The reasoning is straightforward. If time is scarce, and efforts at judicial economy are rewarded more directly than efforts at accuracy or fairness, then the rational judge should take actions that tend to favor speed over either accuracy or fairness.

With respect to the six-month list, the multitask model predicts that, insofar as the list promotes speed, it may also have adverse effects on adjudicative quality. Any evidence of effects on substantive motion outcomes (e.g. grant & denial rates, plaintiff or defendant win-rates, etc.) or appellate outcomes (e.g. appeals rates, reversal rates, etc.) will tend to confirm this hypothesis.[25]

Putting together these various pieces, my conceptual framework yields several predictions. First, my model predicts that exposure to the six-month list will yield faster adjudications on average, with judges deciding many of their motions in the days and weeks immediately preceding the six-month list deadlines. Second, the multitask model suggests that exposure to the six-month list *may* result in changes to substantive case outcomes, but this relationship is likely to depend on factors

[25]This section is dedicated to a summary description of my conceptual framework. A preliminary version of my formal model is presented in Section **??** of the Appendix.

including the degree of judges' present bias, the strength of the reward for judicial effort, and the substitutability between speed and effort. Third, I anticipate that judges will respond to the six-month list heterogeneously, with judges for whom career concerns are especially salient being among the most sensitive to the six-month list. The remainder of this paper proceeds to test these hypotheses. But first, the following section introduces the data behind my empirical analysis.

1.4 Data and Descriptive Statistics

This paper makes use of novel motion-level data from civil cases filed in the United States District Courts. I constructed my original dataset from Westlaw's database of U.S. District Court civil docket reports.[26] Commonly known as the "DCT" database, these data contain much of the same docket information contained in the government's own PACER database. The same DCT database formed the basis of Jonah Gelbach's (2014) study of summary judgment motion filings and judicial characteristics. The data were obtained as raw XML files[27] consisting of both case-level background information (including case filing date; case termination date, if applicable; judge name; detailed names of parties and their lawyers; and standardized codes for the nature of the suit) as well as the text of docket entries pertaining to activity in the case. I wrote computer code to scrape and parse the docket entries and to re-organize them as a motion-level dataset of all summary judgment motions filed between 2004 and 2014. More specifically,

[26] A docket is an administrative record of the proceedings of a particular court case. Each event that transpires in the case—for example, when a litigant files a motion or a brief, when the judge holds a hearing, or when the judge issues a ruling—is recorded as a docket entry.

[27] XML files look much like basic text files, but with additional metadata to indicate the structure of underlying information.

my code searched each docket for docket entries corresponding to original motions for summary judgment. It then matched these motions to docket entries corresponding to court orders disposing of the motion. The motion-level data include the date on which a motion was filed; the identity of the moving party (i.e. whether the motion was filed by the plaintiff or defendant); the date, if any, on which the motion was decided by the court; and the outcome, if any, of the motion (i.e. whether it was granted, denied, granted-in-part, or dismissed due to mootness).

In addition to the original motion-level data, this paper leverages three public-use datasets. First, individual motions are merged with public-use case-level data from the Administrative Office (AO) of the U.S. Federal Courts.[28] In particular, I make use of the Integrated Database (IDB) of civil cases filed, terminated, and pending in federal district courts since the 1970 statistical year.[29] The IDB is prepared by the Federal Judicial Center (FJC), which is a government research and education agency housed within the federal judiciary. I matched motions to cases on the basis of docket number, filing date, and the court in which the case was filed. Although these public-use data provide very little information that was not already available in the Westlaw DCT database, what these data do provide is the opportunity to validate certain aspects of my motion-level data against a commonly-used public-use dataset.

Second, I have similarly merged my motion-level data with a dataset of appeals filed before the U.S. Courts of Appeals. The appellate dataset is also obtained from the FJC's IDB. By merging district court cases with subsequent appeals, I can begin to explore whether exposure to the six-month list had any effect

[28] Available at: `https://www.fjc.gov/research/idb`.

[29] The federal courts utilize a statistical year beginning on October 1st.

on either appeal rates or appellate outcomes (e.g. whether the appellate court affirms, reverses, remands, etc.).

Last, I have merged my data with a database of judge characteristics, also available from the FJC.[30] The FJC's database of judges contains a wealth of demographic and biographical details relating to U.S. federal judges. I will use the data on judge characteristics in order to probe potential heterogeneity in how judges respond to the six-month list.

The result is a dataset consisting of 481,262 summary judgment motions arising from a total of 297,153 separate cases, reflecting an average of approximately 1.62 summary judgment motions per case.[31] Of these, I was able to identify an explicit disposition (including both the date and outcome of the disposition) for 206,513 separate motions (43% of the total). The relatively low rate at which I was able to match new motions to motion dispositions reflects three realities. First, although I restrict to motions filed at least one year prior to the end of my sample period, there are some motions and cases that had not been adjudicated by the end of my sample period.[32] Second, when a case is disposed of on other grounds—for example, when the parties negotiate a settlement—the docket will not always clearly reflect a specific disposition for each pending motion. Insofar as a disposition on other grounds is increasingly likely the longer a motion has been pending, this is likely to limit variation in the amount of time that motions spend pending, and it is therefore likely to bias my results towards zero. Third, given the difficulty

[30] Available at: `https://www.fjc.gov/history/judges/biographical-directory-article-iii-federal-judges-export`.

[31] Although summary judgment motions are most often filed by the defendant, plaintiffs will often file summary judgment motions of their own. Moreover, in cases with multiple defendants, separate defendants will often file separate motions for summary judgment.

[32] In alternate specifications, I will also restrict to motions that were decided within one year, in order to ensure that all motions within my sample were given an equal opportunity to be decided.

of parsing highly variable text entries, it is quite possible that my algorithm has simply failed to identify some dispositions. Insofar as these limitations introduce simple measurement error, they are likely to further attenuate my findings (i.e. bias them towards zero).

Table 1.2 summarizes my original motion-level dataset. Among the sample of summary judgment motions in which a disposition could be identified, approximately 64% were filed by the defendant, and approximately 29% were filed by the plaintiff, reflecting the pro-defendant bias of the summary judgment device. I was unable to identify a movant in the remaining 7% of cases, which may indicate that summary judgment was entered by the court *sua sponte*. The average summary judgment motion was decided in approximately 5.36 months (compared to an average overall case duration of slightly less than two years.[33] The remaining rows show that motions for summary judgment are frequently granted, with approximately 62% of my sample being either fully granted or granted in part.

The summary statistics presented above show mean motion duration, but we may learn more by examining the full distribution of motion durations. Figure 1-2 shows a histogram of total summary judgment motion duration (i.e. months pending before disposition) for my main sample of adjudicated motions. Although the modal duration is less than five months, a large share ($\sim 32\%$) of motions are pending for between six and thirteen months. Fewer than 4% of motions extend beyond thirteen months.

Each motion in my dataset is assigned a "Nature of Suit" code indicating the nature of the underlying suit. Being that my data are drawn from the entirety of

[33]Since summary judgment motions occur relatively late in the course of litigation, the average overall case duration in my dataset is likely to be higher than the average overall case duration across all civil filings in U.S. district courts.

district court civil filings, my main sample spans a wide variety of case types. Appendix Figure A-2 shows the approximate distribution. It is worth noting that the legal significance of a summary judgment motion, as well as the value to litigants of judicial efficiency, is likely to vary across these case types.[34] Among the most common case types are employment discrimination, personal injury, prisoners' rights,[35] and other civil rights suits.

1.5 Empirical Framework

The following section provides details on my empirical framework, the goal of which is to estimate the causal effects of exposure to the six-month list on both the speed and quality of district court adjudications. Under the CJRA, federal courts must prepare semiannual reports of all motions pending more than six months and all civil cases pending more than three years. Because the reports are published just twice a year—on March 31st and September 30th—cases and motions vary in their "reporting time," which is the term I will use to refer to the amount of time that a judge could *hypothetically* spend reviewing a motion before that motion must appear for the first time on a six-month list. In other words, cases and motions can be more or less exposed to the list. Under implementation guidelines established by the federal judiciary, "[a] motion becomes pending 30 days after the date it was filed or was referred to a magistrate judge, whichever

[34]As detailed in the Section 1.5, most specifications will include nature-of-suit fixed effects in order to control for systematic differences between case types.

[35]Note that the prisoners' rights cases in my dataset do not include petitions for habeas corpus, which are excluded from the CJRA's reporting requirements. Prior to excluding them from my sample, summary judgment motions filed in habeas petitions accounted for approximately 2% of my raw sample of summary judgment motions.

is later."[36] Accounting for this 30-day grace period, motions will vary between approximately seven and thirteen months of reporting time.

Figure 1-3 illustrates two extreme examples of motions' relative exposure to the six-month list. Consider first a motion filed on February 29, 2016,[37] depicted by the top panel of Figure 1-3. According to the implementation guidelines, the motion becomes pending 30 days later, which happens to fall on March 30th. On March 31st, when the next six-month list is published, the motion has only been pending for one day, so the motion is of course ineligible to appear on the list. However, fast-forwarding to September 30, 2016, the motion has been pending for exactly six months, and if the judge has not yet disposed of it, it must appear on the September 30th list. Counting the days between February 29th (when the motion was filed) and September 30th (when the motion becomes eligible for its first six-month list), the motion enjoys 214 days (or approximately seven months) of reporting time. Now, consider a motion filed just one day later, on March 1st, 2016, depicted in the bottom panel of Figure 1-3. The motion becomes pending 30 days later, on March 31st. On September 30, 2016, the motion has been pending for just short of six months, so the motion is ineligible to appear on the September 30th list. Instead, the motion does not become eligible until March 31, 2017, at which point the motion has already enjoyed 395 days (or approximately thirteen months) of reporting time. Between these two extremes, motions will vary between seven and thirteen months of reporting time. Figure 1-4 plots reporting time as a function of motion filing date.

Stated in the simplest terms, my empirical strategy consists of comparing the

[36]Available at: `http://www.uscourts.gov/sites/default/files/data_tables/cjra_na_0930.2017.pdf`.

[37]Note that 2016 was a leap year, although a motion filed on February 28th of any other year would have exactly the same reporting time.

outcomes of motions with relatively high reporting time to the outcomes of otherwise similar motions with relatively low reporting time. In the following section I consider the assumptions that must be met in order for my approach to yield credibly identified causal estimates of the effects of the six-month list.

1.5.1 Identifying Assumptions

Identification requires that, conditional on the available motion- and case-level controls, the date on which a motion is filed is effectively random. In other words, it would a problem for my identification if parties timed their motion filings strategically in order to take advantage of the six-month list. If litigants file their motions strategically—for example, seeking to either expedite or delay the adjudication of their motions by filing just before or after a reporting deadline, or seeking to take advantage of a judge's tendency to either grant or deny motions depending upon their relative exposure to the six-month list—then it could be the case that motions filed with high reporting time are systematically difference from those filed with low reporting time. It would be similarly problematic if judges manipulated motion filing dates—for example, by issuing a scheduling order—in order to take advantage of the six-month list.

My key identifying assumption can therefore be stated as follows: while judges may allow the six-month list to influence how they adjudicate a motion, they do not preemptively manipulate the timing of motion filings; and, moreover, litigants and lawyers are either unaware of the six-month list or they do not care enough about it to take it into account when they choose a motion filing date. To be sure, this assumption violates some common sense. As we are reminded by Jonah Gelbach, litigants are not "inanimate particles bouncing around and filing motions exogenously," but rather "live parties—who, together with their attorneys, make deliberate, strategic decisions" (Gelbach, 2014). However, there are several good reasons to believe that litigants do not file motions strategically with respect to the six-month list. First, and perhaps most importantly, litigation is complicated even in the absence of judicial reporting rules, and to predict the impact of motion filing date on a judge's behavior would only complicate things further. In other words, lawyers and litigants are "boundedly rational" ((Simon, 1955)). Moreover, motion filing dates are often dictated by pre-established filing deadlines, and many motions are dependent upon the occurrence of other events. For example, motions for summary judgment must be filed within 30 days of the completion of discovery,[38] and the completion of discovery is itself likely to be dictated by local court rules and case-specific scheduling orders. It seems unlikely that either judges or litigants are thinking about the intricacies of the six-month list when, several months in advance of a summary judgment motion, they are formulating their discovery plans under Rule 26(f). These factors will only be amplified by the many simultaneous cases between which attorneys and judges must typically divide their attention.

Of course, when possible, the best place to look for support of an identifying assumption is in the data itself. If, contrary to our identifying assumption, motion filings are timed strategically in response to the six-month list, then we might expect to see such a pattern in the data. In fact, no such pattern is discernible. Figure 1-5 shows a histogram of the empirical distribution of motion filings by calendar day. Calendar dates with unusually high filing counts (more than two standard deviations above the daily mean) are labeled from above. While a pattern does emerge, there is no obvious relationship to the six-month list reporting

[38] FED. R. CIV. P. 56(b)

deadlines. Instead, what we see are merely bi-weekly spikes at approximately the beginning, middle, and end of each month—*regardless* of month—and large dips on or around major federal holidays like January 1st, July 4th, and December 25th. The bi-weekly spikes may reflect law firm customs, where billable hours are often due on a bi-weekly or monthly basis, or it may simply reflect judges' and lawyers' natural tendency to schedule business for certain "anchoring" dates. Regardless, after taking into account these bi-weekly spikes, motion filings appear to be relatively uniform throughout the course of the calendar year.

This point is further illustrated by Figure 1-6a, which shows a kernel density plot of the raw empirical distribution. In comparison, Figure 1-6b plots the same empirical distribution after controlling for dummy variables indicating the first, fifteenth, and last day of each month.[39] Neither graph shows any discernible relationship between motion filings and six-month-list cutoff dates.

Stepping back from the formal identifying assumptions, it is worth stating the goal of these assumptions, which is to establish a "control" group of motions that were relatively unexposed to the six-month list against which we can compare the motions that were most exposed. We want to establish that, aside from their exposure to the six-month list, motions in the treatment and control groups are otherwise similar. Reassuringly, Table 1.3 shows that a variety of *ex ante* motion- and case-level controls are "balanced" across motions with high and low reporting time. Relative to motions with high reporting time, motions with low reporting time are no more likely to be filed by either the plaintiff or the defendant, they are no more or less likely to be filed in a case with at least one *pro se* litigant, they are no more or less likely to be likely to be filed in a case where at least one litigant has

[39]Specifically, Figure 1-6b plots the residuals from a linear regression of total motion filings (per calendar day) on dummies for the first, fifteenth, and last day of each month.

sought *in forma pauperis* status, and they share a similar distribution with respect to the nature of the suit. While balance across observable characteristics does not guarantee balance across unobservable characteristics, it does suggest that motions with high reporting time represent a reasonable control group against which to compare motions with relatively low reporting time.

As shown above, the data offer little support for the notion that either judges or litigants are strategically manipulating motion filing dates in order to take advantage of the six-month-list. Nevertheless, an instrumental variables strategies may obviate the need for this identifying assumption altogether. Specifically, an IV approach could exploit certain milestones in the course of litigation (for example, the date on which the case was filed, or the date on which discovery was initiated or completed) as instruments for the date on which a summary judgment motion was actually filed.

The following section translates my basic empirical framework into a series of estimating equations. In particular, I will estimate three common econometric models: ordinary least squares (OLS), regression discontinuity (RD), and proportional hazard. While the models vary with respect to technical implementation, they share the same basic function, which is that they can be used to compare the outcomes of motions with high and low reporting time.

1.5.2 Econometric Models & Estimating Equations

First, I address the effects of the six-month list on what is perhaps the most common measure of judicial efficiency: mean time until disposition. In particular, I would like to know whether exposure to the six-month list causes motions to be adjudicated more quickly. I begin with an Ordinary Least Squares (OLS) model of the following general form:

$$\text{Months Until Disposition}^{\overline{T}}_{ijt} = \alpha + \theta \text{Reporting Time}_{ijt} + \mathbf{X}'_{ijt}B + \rho t + \lambda_t + \mu_j + \epsilon_{ijt} \tag{1.5.1}$$

where Months Until Disposition$^{\overline{T}}_{ijt}$ represents the total number of months that motion i filed before judge j at time t has spent pending at the time of disposition ($t = \overline{T}$). In other words, what was the motion's total duration? I regress Months Until Disposition on Reporting Time$_{ijt}$, which represents the amount of time the judge has to review the motion before it first becomes eligible for reporting on a six-month list.[40] Included in the baseline regression are a vector of motion- and case-level controls, represented by $\mathbf{X}_{ijt}$, and filing date time trends and fixed effects,[41] represented by ρt and λ_t, respectively. These linear time trends and fixed effects allow me to control for any confounding "calendar effects" that are correlated with but unrelated to the 6-month list—for example, it is conceivable that judges simply wait until the end of a month to take action on pending motions, or

[40]Recall from Section 1.5 that "Reporting Time" is a function of motion filing date, and it is completely independent of whether the motion is ever actually reported on a 6-month list. For example, two motions filed on January 1st will both have the same amount of Reporting Time, even if one is terminated the very next day and the other is still pending months later.

[41]In particular, my preferred specification includes filing year and day-of-month fixed effects. While I can include *either* day-of-month or month-of-year fixed effects, I cannot include both, since my variation comes entirely from the day-of-month and month-of-year combination.

perhaps they structure their schedules around holidays.

Basic case- and motion-level controls will include dummies for whether the motion was filed in a case with at least one *pro se* litigant, whether the motion was filed by the plaintiff or defendant, and whether any other summary judgment motions were filed in the same case. My preferred specification includes judge fixed effects (μ_j) as well as nature-of-suit fixed effects, filing year fixed effects, and district court fixed effects.

The coefficient of interest is θ, which measures the effect of an additional month of reporting time on the total months until motion disposition. Conditional on the identifying assumptions stated above, θ represents a causal estimate of the effect of additional reporting time on total case duration.[42]

One concern related to our regression analysis is that it may suffer from so-called "survivorship bias."[43] In other words, since they are based on a dataset of *completed motions*, my estimates may be biased by my inability to observe motions that are still pending at the time of my data collection. I therefore choose to complement my OLS with a Cox proportional hazards model. A proportional hazards model will allow us to estimate the effect of motion reporting time on the *rate* at which motions are resolved. In addition to addressing concerns of survivorship

[42]Equation 1.5.1 assumes that the effect of reporting time is constant (i.e., that each additional month of reporting time has the same treatment effect), but this may not be the case. In order to test this assumption, I will also estimate a model with separate coefficients for each month of reporting time.

$$\text{Months Until Disposition}^{\overline{T}}_{ijt} = \alpha + \sum_{q=8}^{13} \beta_q \mathbb{1}[q < \text{Reporting Time}_{ijt} \leq q+1] + \mathbf{X}'_{ijt}B + \rho t + \lambda_t + \mu_j + \epsilon_{ijt} \quad (1.5.2)$$

[43]The term "survivorship bias" may be somewhat misleading in our context, since the survivors are those motions that have been fully adjudicated.

bias, this strategy will allow me to leverage current data on still-pending summary judgment motions, which will substantially increase my sample size.[44]

While the above models provide an obvious starting place for our analysis, they fail to take advantage of one of the most distinctive features of the six-month list, which is the "jump" in reporting time that occurs on both March 1st and August 30th. Recall from Figure 1-4 that, while motions filed in the final days of February and August enjoy little more than seven months of reporting time, motions filed on or immediately after March 1st and August 30th enjoy almost thirteen months of reporting time. This natural discontinuity in reporting time motivates the use of a Regression Discontinuity (RD) design. An RD-style model is frequently used to study the effects of some policy or intervention when the policy is applied on the basis of some "cutoff" or "threshold" score. Here, by comparing motions filed just prior to March 1st and August 30th with those filed on or just after the cutoff dates, I can obtain causal estimates of the effect of exposure to the six-month list on the speed of adjudication.

The RD procedure can be expressed in a slightly simplified form with the following equation:

$$\text{Months Until Disposition}_{ijt}^{\overline{T}} = \alpha + \beta \text{Non-Reportable}_{ijt} + f(t) + \epsilon_{ijt}, \qquad (1.5.4)$$

where $\text{Non-Reportable}_{ijt} = \mathbb{1}(f(t) \geq 0)$. The function $f(t)$ is a "running variable"

[44]My basic proportional hazard model takes the following form:

$$\lambda(t) = \lambda_0(t) \exp(\beta \text{Reporting Time}_{ijt} + \mathbf{X}_{ijt}'\Gamma), \qquad (1.5.3)$$

where $\lambda(t)$ represents the Cox hazard function, Y_{idt} denotes the amount of time before case i filed in district d on date t becomes eligible for publication on a 6-month list, and $\mathbf{X}_{idt}$ is a vector of case-specific controls. The coefficient of interest is β, which reflects the effect of additional review time (i.e. less exposure to the six-month rule) on the log of the hazard ratio.

that measures the distance between the motion's actual filing date and the two filing dates with maximum reporting time (i.e. March 1st and August 30th). The function is slightly negative for motions filed just before March 1st or August 30th and slightly positive for motions filed just after those dates. This can be seen graphically in Figure 1-7, below, which plots the running variable $f(t)$ as a function of filing date. Since the filing date cut-offs are semi-annual, no day of the year is more than approximately ninety days distant from the nearest cutoff, and the running variable therefore varies between -90 and 90.

Regression discontinuity designs are subject to a few specific identifying assumptions. In particular, the key assumption of an RD design is that the underlying conditional expectation functions $\mathbb{E}[Y_i(1)|X]$ and $\mathbb{E}[Y_i(0)|X]$ are continuous across the cutoff in the forcing variable X (Imbens and Lemieux, 2008). In my setting, this is equivalent to saying that unobservable factors are continuously related to the running variable $f(t)$, including at the cutoff dates. While there is no direct test for this "continuity assumption," it is likely to be met when the distribution of observed baseline covariates do not change discontinuously at the threshold (Lee and Lemieux, 2010). In fact, as shown in Appendix Figure A-3, several baseline covariates do appear to be distributed continuously at the threshold.[45]

[45]Recent research suggests that the regression discontinuity design is subject to several unique pitfalls when time is used as the running variable (Hausman and Rapson, 2018). In particular, the "regression discontinuity in time" (or "RDiT") approach is conceptually and practically distinct from the traditional cross-sectional regression discontinuity design because it typically relies on time-series variation for identification. As a result, the RDiT design often leverages observations far from the threshold and often ignores autoregression in the data generating process. Moreover, since time is uniformly distributed, McCrary tests are often irrelevant in an RDiT context. I argue that my context actually shares more in common with a conventional cross-sectional RD than it does with an RDiT. In particular, since hundreds or thousands or motions can be filed each day, I am able to leverage a great deal of cross-sectional variation close to the threshold. Moreover, since *motion filings* are not uniformly distributed across time, and because I argue that filing dates are locally random in the neighborhood of the threshold, manipulation tests continue to be highly relevant.

A related assumption of RD designs is that agents do not have precise control over the running variable. In other words, it must be that agents cannot "manipulate" their treatment status. Here, the running variable is a function of the motion filing date, which litigants obviously *can* manipulate. However, for the reasons stated above, I argue that litigants do not have *precise* control over their filing date; or, at the very least, they do not manipulate their filing date in order to take advantage of the timing of the six-month list. This proposition is supported by Figures 1-8a and 1-8b, which show the empirical distribution of summary judgment motion filings by filing date, where the filing date has been transformed into the RD running variable $f(t)$. Figure 1-8a shows the raw distribution of motion filing dates, while Figure 1-8b shows the adjusted distribution after controlling for dummies for the first, fifteenth, and last day of each month. If there were manipulation of the running variable, then we might expect to see bunching of summary judgment motions filed immediately before, on, or immediately after the cutoff. While both figures continue to show the same bi-weekly spikes that were observable in Figure 1-5, there does not appear to be any unusual bunching at or near the cutoff dates.[46] A more formal test of manipulation using the method outlined by McCrary (2008) similarly fails to reject the null hypothesis of no manipulation.[47]

It is worth noting one challenge related to RD designs, which stems from the fact that they rely on a narrow window of datapoints in close proximity to the

[46]That is, although there is some bunching directly at the cutoffs, the bunching appears to be approximately identical to the bunching that occurs throughout the year on an approximately bi-weekly basis.

[47]At least in theory, it could be that different types of litigants have different strategic incentives. For example, perhaps plaintiffs in employment discrimination suits like to draw out litigation in order to reach a settlement, in which case they file when reporting time is high, whereas defendants want a quick resolution, so they file when reporting time is low. If these two tendencies balance one another out, then in the aggregate, it might appear is if there is no manipulation. I can begin to account for this by running separate manipulation tests on different sub-samples of my data.

cutoff point. As a result, RD models frequently lack the statistical power to detect small effects.

The models described above allow me to explore the effect of the six-month list on the speed with which motions are adjudicated. But in addition to speed, I am also interested in the six-month list's effects on the *quality* of adjudication. Intuitively, if exposure to the six-month list causes a judge to adjudicate a motion more quickly, it may also affect *how* she disposes of the motion. Quality is an admittedly vague concept, and it can mean many things in the context of civil adjudication. From current and future litigants' perspective, quality may refer to the degree of substantive or procedural fairness accorded to the parties. From the court administrator's perspective, quality may refer to the efficient allocation of judicial resources. Neither notion of quality is easy to measure, nor are they entirely distinct. As preliminary evidence of quality effects, I will look for whether the six-month list had any effect on motion-level outcomes. In particular, I will ask whether motions that were more exposed to the six-month list were either more or less likely to be granted, denied, or granted in part, and whether they were more or less likely to result in a judgment favorable to either the plaintiff or the defendant. Intuitively, if the only effect of the six-month list was to expedite adjudications, then we would not expect to see any change in motion outcomes. While these indicators provide little in the way of a priori evidence for effects on quality—since it is impossible to say how these motions should have been decided in the first place, it is hard to say whether the result was higher or lower quality decisions—they are at least somewhat probative. In addition to the above outcomes, I will also ask whether motions that were more exposed to the six-month list were either more or less likely to result in an appeal, and whether there was an effect on the outcome of the appeal (e.g. whether the Court of Appeals affirmed, reversed, or remanded to the district court). These outcomes are slightly easier to interpret. While we cannot say whether a motion should or should have not been appealed, it is uncontroversial to say that a goal of the justice system is to reduce the need for appeals. Moreover, reversals and remands offer fairly direct evidence that the district court's initial judgment was either improper or inadequate.

Empirically, the goal will be to identify the causal effect of exposure to the six-month list on the likelihood of various motion-level and appellate outcomes. Specifically, I estimate a linear probability model identical in form to equation (1.5.1), except that the left-hand-side variable is replaced with a dummy variable for the outcome (e.g., whether or not the motion was granted). In order to test my linear specification, I will also conduct robustness checks with Logit and Probit models, which allow for a non-linear relationship between reporting time and the likelihood of a particular outcome. Finally, I will also use the regression discontinuity specification from equation (1.5.4) in order to look for evidence of an effect on motion-level outcomes in the vicinity of the reporting time discontinuities.

1.6 Results & Discussion

1.6.1 How Does the Six-Month List Affect the Speed of Adjudication?

I first present evidence of the effect of relative exposure to the six-month list on the speed of adjudication. Without even introducing the regression results, a single graph makes the key point: summary judgment motions that are most exposed to the six-month list are adjudicated much more quickly. Figure 1-9 shows kernel density plots of the empirical distributions of motion duration by relative reporting time. The blue curve corresponds to motions with low (fewer than eight months) reporting time, and the red curve corresponds to motions with high (greater than twelve months) reporting time. What stands out is that motions with low reporting time are considerably more likely to be adjudicated in fewer than ten months. While the modal motion duration for low-reporting-time motions is fewer than six months, there appears to be something like a bi-modal distribution, with the second peak at approximately eight months—that is, exactly when the motions are due for the six-month list. In fact, the high-reporting-time motions follow a similar distribution, except that the distribution appears to be stretched out over a larger interval. While the modal motion duration for high-reporting-time motions is fewer than six months, the second peak now occurs at approximately twelve months—again, exactly when the motions are due for the six-month list.

Next we consider the regression results, which allows us to quantify the effect observed in Figure 1-9. Table 1.4 presents OLS estimates of equation (1.5.1). Columns (1)-(4) correspond to various combinations of controls. We will focus on column (4), which includes various linear time trends (for day of year, day of quarter, and day of month), district*year fixed effects, and day-of-month fixed effects, but the results are robust across specifications. Column (4) tells us that, on average, each additional month of reporting time corresponds to ~ 0.13 additional months of total motion duration. Extrapolating linearly, since the least exposed motions enjoy an additional six months of reporting time relative to the most exposed motions, we can infer that the most exposed motions are adjudicated approximately 0.8 months sooner than those that are least exposed. Compared to the mean summary judgment motion duration of 5.36 months, this represents a nearly 15% effect.

Of course, from the perspective of both the litigants and the court administrators, one might suspect that what really matters is time until overall *case* disposition, and not merely time until motion disposition. In fact, here too we see substantial effects on the speed of justice. Table 1.5 presents OLS results where we replace the left-hand-side of equation (1.5.1) with months until overall case disposition. The variation on the right-hand-side still comes from the motion-level reporting time. The OLS results indicate that, on average, each additional month of summary judgment *motion* reporting time corresponds to ~ 0.08 additional months of total *case* duration. Once again multiplying this effect by six, it appears that, when a summary judgment motion is most exposed to the six-month list, the overall case of which it is a part lasts approximately half a month longer. Compared to the mean case disposition time of 23.37 months, represents more than a 2% effect.

One might question the assumption of linearity—that is, does each additional month of reporting time really have the same effect on the speed of adjudication? The answer is that, while the relationship between reporting time and speed of motion adjudication may not be quite linear, it is at least monotonically increasing. Appendix Figure A-4a plots the coefficients β_q from the non-parametric model in equation (1.5.2). Whereas motions with between eight and nine months of reporting time last only about 0.14 months longer than motions with less than eight months of reporting time, motions with between twelve and thirteen months of reporting time last more than 0.7 months longer.

The results in Table 1.4 and Figure A-4a are estimated from a sample of approximately 206,000 summary judgment motions. By construction, in order to know their final duration, the motions in this sample had to be fully adjudicated.

As discussed in Section 1.5, a proportional hazard model (like the one shown in equation (1.5.3)) allows us to leverage the full sample of nearly 500,000 motions, whether or not they have been fully adjudicated. The proportional hazards model therefore alleviates any concerns over survivorship bias. In fact, Appendix Table A.8 shows that the effect of the six-month list on motion duration are equally apparent in a proportional hazards model. In particular, the hazard rate of motion disposition decreases significantly with each additional month of reporting time. In other words, motions that are less exposed to the six-month list are disposed of at a slower rate.

Next we consider results from the regression discontinuity design. Recall from Figure 1-4 that motions experience a large, discontinuous jump in reporting time on March 1st and August 30th. Motions filed just one day prior enjoy only seven months of reporting time compared to thirteen months of reporting time for motions filed on or immediately after those dates. If reporting time is as influential for motion duration as I argue it is, then we would expect to see a similarly discontinuous jump in motion duration at the same filing date cutoffs. In fact, that is exactly what we see. Figure 1-10, which plots predicted values from a local linear regression against a scatter plot of actual average motion duration, indicates a substantial jump in average motion duration precisely at the cutoff dates. Table 1.6 quantifies this effect. While the estimates vary according to modeling assumptions and chosen bandwidths, the results are roughly consistent with the inferences we made from the OLS models. Namely, the most exposed motions are adjudicated up to 0.8 months faster than those that are least exposed to the six-month list.

1.6.2 How Does the Six-Month List Affect the Quality of Adjudication?

So far we have seen widespread evidence that the six-month list does indeed expedite the adjudication of summary judgment motions. This result is consistent with the notion that judges may believe their future career prospects partially depend on compliance with the six-month list. But what does exposure to the six-month list entail for the *quality* of adjudication? Recall from Section 1.3 that our predictions for adjudicative quality will likely depend upon the model that we have of judicial behavior. Judge's concern for their future career prospects is enough to predict an impact on the speed of adjudication, but it may not tell us much about the impact on the quality of adjudication. The model predicts that whether judges tend to compromise quality for speed is likely to turn on a number of factors, including: 1) the degree to which judges procrastinate, 2) the degree to which judges feel rewarded for the amount of care and effort they invest in motions, and 3) the substitutability of speed and quality.

In fact, I find only mixed evidence to suggest that exposure to the six-month list affects how judges dispose of the summary judgment motions before them. At most, the effects appear to small. Table 1.7 presents linear probability model estimates of the effect of additional six-month list reporting time on various motion-level and appellate outcomes.[48][49] Since the legal significance of these outcomes

[48]More detailed results, including robustness to various model specifications, are presented in Appendix Tables A.2 and **??**.

[49]It should be noted that columns (3)-(5), which report appellate outcomes, are conditioned on the outcome of the motion itself—that is, whether the district court granted, denied, granted-in-part, or otherwise disposed of the motion. Since appeals are more likely to be filed when a summary judgment motion is granted, and since the Court of Appeals is more likely to affirm when a summary judgment motion has been granted, the conditional effects reported in columns (3)-(5) tell us whether there is something else about motions with greater reporting time that make

is likely to depend upon which party filed the motion—a summary judgment filed by the defendant is more likely to be fully dispositive of the entire case, for example—I choose to restrict the sample to motions filed by the defendant, which are more common.

What is immediately apparent is that, in comparison to the effects on the speed of adjudication, the effects on on motion and appellate outcomes are small and relatively imprecisely estimated. We do observe what appear to be modest effects on summary judgment grant rates—for each additional month of reporting time, motions are approximately 0.19 percentage points likely to be granted—but the estimate is only marginally significant. This result is robust to various specifications of the OLS model, and it is also robust to the choice of Logit and Probit models. While the point estimate is small, when put in the proper context, it does appear to be somewhat meaningful. At first glance, this may seem like a small effect, but in context, it is meaningful. Given that the least exposed motions enjoy six months of additional reporting time compared to the most exposed motions, and given that on average 57% of motions are granted, this amounts to a 2% effect on the summary judgment grant rate.

The judicial multitasking model discussed in Section 1.3 predicts an effect on the grant rate and other motion-level outcomes, but it does not predict the sign (either positive or negative) or magnitude of these effects. Nonetheless, the observed effect on the summary judgment grant rate makes some intuitive sense. Summary judgments are dispositive motions. Whereas an order granting summary judgment often disposes of the case altogether, an order denying, granting in part, or otherwise dismissing a summary judgment typically allows the parties to live to fight another day. Judges may therefore view orders to deny, grant

them more or less likely to result in a particular appellate outcome.

in part, or moot as more conservative courses of action. Moreover, the decision to grant the motion may simply entail more work. While Rule 56 of the Federal Rules of Civil Procedure requires that judges must "state on the record the reasons for granting or denying the motion,"[50] judges typically only write lengthy decisions when they are granting the summary judgment (Gertner, 2012). As a result, judges who are under pressure to meet a deadline imposed by the six-month list may choose to deny or dismiss the motion in order to avoid the extra risk and extra work associated with an order to grant.[51]

The results on appellate outcomes are even less pronounced. It does appear that judgments in cases where the summary judgment motion was relatively unexposed to the six-month list (i.e. with greater reporting time) may be slightly more likely to be affirmed by an appeals court (conditional on appeal), but the effect is small and statistically insignificant.

In contrast to the results on the speed of adjudication, none of the results on motion-level an appellate outcomes are detectable using the regression discontinuity design. Appendix Figure A-5 presents regression discontinuity plots of selected outcome variables, and Appendix Table A.6 presents corresponding RD estimates; the plots show no discernible discontinuities at the reporting time cutoffs. This should give us some pause with respect to the OLS results presented above. Taken together with OLS results reported in Table 1.7, the RD results suggest that, insofar as the six-month list has any effect on motion outcomes, the effect

[50]Fed. R. Civ. P. 56(a).

[51]That the effects on orders to deny, grant in part, and moot are all small and statistically insignificant may reflect the fact that, whereas all three courses of action allow the case to proceed in one way or another, only the order to grant fully disposes of case. In other words, the opposite of an order to grant is not simply an order to deny, but rather any order *other than* an order to grant. If the effect is dispersed across all three courses of action, then any one of these effects will be smaller and more difficult to detect with statistical precision.

is very small.

As a final piece of evidence on the quality of adjudications, I consider how exposure to the six-month list affects the speed of overall case dispositions. We have already seen (in Table 1.4 and elsewhere) that exposure to the six-month list tends to expedite motion dispositions. Moreover, as shown by Table 1.5, faster motion processing does indeed translate into faster case processing. However, it is striking that the coefficients presented in Table 1.5 are quite a bit smaller than the coefficients presented in Table1.4. In other words, it appears that a month saved in the summary judgment phase *does not* translate into a full month of savings in overall case disposition time.

This observation motivates the following exercise, which attempts to dig more deeply into how the six-month list affects overall case processing. We can think of the six-month list as having two types of effects on overall case processing. First, there is the “direct” effect on motion processing. Ordinarily, the sooner a motion is decided, the sooner the overall case is decided. If all that mattered were the direct effect, then we would anticipate a one-for-one relationship between time until motion disposition and time until case disposition. However, the six-month list may also have “indirect” effects on case processing. The effects could go in either direction. For example, if exposure to the six-month list caused judges to resolve certain factual or legal questions in a way that narrows issues still in dispute, then that might tend to expedite the trial phase of the proceeding, even after the summary judgment phase has been decided. If that were the case, then a month saved in the summary judgment phase might actually translate to more than a month saved in overall case disposition time. Alternatively, if exposure to the six-month list causes the judge to “cut corners” during the summary judgment phase—for example, postponing certain factual or legal questions until later in the course of proceedings—then we might expect a month saved in the summary judgment phase may not translate into more than a full month of savings in overall case disposition time. In fact, if judges tend to reallocate work in an inefficient manner (e.g., postponing the resolution of some question until later in the proceedings when it is more time-consuming to resolve), then a month saved in the summary judgment phase may even translate into *less* than a month saved in overall case disposition time.

Column (1) of Table 1.8 reproduces the main result from column (1) of Table 1.5. Recall that the regression is based on equation (1.5.1), except that the left-hand-side variable is not months until motion disposition, but rather months until case disposition. Column (1) shows that, on average, each additional month of summary judgment *motion* reporting time corresponds to ~ 0.08 additional months of total *case* duration. But how of that effect is attributable to the "direct" effect on motion processing, and how much is attribute to "indirect" effects on other aspects of the case proceedings? Columns (2) and (3) attempt to decompose the overall effect into its constituent parts. Column (2) copies the specification from column (1), except that it controls for duration of the motion itself. This effectively controls for the direct effect, so that any remaining coefficient on reporting time must be attributable to the indirect effect. What we see is that, after controlling for the direct effect on motion disposition time, each additional month of reporting time reduces overall case duration by an average of 0.052 months. In other words, controlling for the direct effect on motion disposition time, cases that are most exposed to the six-month list actually last *longer* than cases that are least exposed. Column (3) shows that these indirect effects persist even after controlling for motion-level outcomes (i.e. whether the motion was granted, granted in part, etc., and whether an appeal was filed subsequent to motion disposition). Table 1.8 suggests that, although the six-month list is effective at expediting motion processing, the six-month list may also have the perverse effect of encouraging certain inefficient practices that tend to dampen the overall effect on case dispositions is somewhat. I interpret this as evidence that the six-month list may indeed cause judges to inefficiently "cut corners."

In future work I intend to investigate other proxies for judicial quality, including the frequency, content, and citation rates of written judicial opinions. I hope that these proxies will offer more insight into the how and why the six-month list affects adjudicative quality.

1.6.3 Do Judges Respond Heterogeneously?

Finally, I conclude this section by presenting evidence that judges exhibit a great deal of heterogeneity in their responsiveness to the six-month list. Table 1.9 presents results from OLS regressions that are similar to equation (1.5.2) except that they interact reporting time with selected judge traits, including whether the judge was under fifty-five years old at the time of the motion filing, whether the judge is non-white, whether the judge is a woman, whether the judge was serving as the Chief Judge of her district at the time of the motion filing, and whether the judge was appointed by a president of the same party as the current President at the time of the motion filing. All regressions include judge fixed effects, and where the trait in question varies with time, the uninteracted judge trait is also included.

I cautiously interpret these results as being broadly consistent with a model of career concerns, where judges are motivated to comply with the six-month list in order to enhance their opportunities for promotion. In fact, there are at least two explanations for why a career concerns-style model might lead to heterogeneity across dimensions including judges' age, race, and gender. The first story is slightly more uplifting, at least for those who care about diversity on the bench and equity in the workplace. Specifically, I argue that the observed heterogeneity may be driven by recent efforts to diversify the federal bench. Although the federal judiciary remains far more white and male than the American public overall (men represent 73% of Article III judges, and more than 80% of Article III judges are white/non-Hispanic, compared to the approximately 61% of Americans who are white/non-Hispanic)[52], the judiciary has grown more diverse in recent years, especially under President Obama. When the push to nominate a diverse pool of judges is combined with the current low baseline level of diversity in the judiciary, judges who are members of underrepresented minorities (namely, women and people of color) may perceive enhanced prospects for promotion. When prospects for promotion are more salient, judges are likely to be especially sensitive to the six-month list.[53]

However, a more pernicious story of workplace discrimination could also explain the pattern of observed heterogeneity. Specifically, it is possible that young, female, and racial/ethnic-minority judges simply need to do more and higher quality work in order to receive the same level of recognition as their white/male peers. If that is the case, then the returns to compliance with the six-month list are simply greater for judges who are members of these under-represented minorities. This, too, would explain greater sensitivity to the six-month list among

[52]Data on judge demographics available at `https://www.fjc.gov/history/exhibits/graphs-and-maps/demography-article-iii-judges-1789-2017-introduction`.

[53]This hypothesis depends upon whether we view under-represented minority status and compliance with the six-month list as either substitutes or complements with respect to the likelihood of promotion. I speculate that they are much more likely to be complements. That is, the probability of promotion is increasing in both under-represented minority status *and* compliance with administrative deadlines, and the presence of one quality does not diminish the returns to the other.

young, non-white, and female judges.

In future work I intend to exploit additional variation in the likelihood of promotion—including variation in judicial vacancies on the Courts of Appeals—in order to further investigate how career concerns interact with individual traits including race, gender, and age. I also hope to further evaluate the competing explanations for heterogeneous career concerns.

1.7 Discussion: What Can the Six-Month List Tell Us About Effective Civil Justice Reform?

The preceding empirical analysis reveals that social sanctions do indeed provide effective incentives, even among workers as elite and highly insulated as federal judges. However, my analysis also reveals that speedier adjudications may come at a cost. I find suggestive evidence that the six-month list may influence not only when judges do their work, but also *how* they do it, and it may cause judges to inefficiently cut corners. But what does this mean for optimal civil justice policy? In particular, what does the preceding analysis tell us about optimal judicial incentive schemes?

If nothing else, my analysis suggests that the six-month list would likely benefit from several minor tweaks. My analysis indicates that the six-month list suffers from two major deficiencies. First, even insofar as the six-month list is effective, motions and cases vary arbitrarily in their exposure to the list, and judges vary widely in their responsiveness to the list. The six-month list would benefit from reforms aimed at making its effects more uniform across motions, cases, and judges. Second, while the six-month list does indeed accomplish its ostensible goal of promoting speedy adjudications, it also appears to have unintended consequences for the quality of adjudication. An additional set of reforms should aim to reduce judges' incentives to cut corners.

1.7.1 Ensuring uniformity of judicial incentives

At present, motions vary enormously in their exposure to the six-month list. While judges have just seven months to review some motions before they appear on a six-month list, other motions enjoy nearly thirteen months of reporting time. While this variation is a boon to economists, who are always on the lookout for a good natural experiment, from the standpoint of judicial policy, this variation is sub-optimal. Variation in exposure to the six-month list creates unpredictability, and for especially savvy judges and attorneys, it does create opportunities for strategic behavior.[54]

One solution to the problem of non-uniformity would be to use a continuously-updating six-month list. In other words, motions and certain cases pending for six-months or longer would be added to a publicly available website at the end of each business day. Under this system, all cases would benefit equally from the judicial incentives for a speedy resolution, and judges would also have less opportunity to prioritize some cases while neglecting others. One potential pitfall of the continuously-updating list, however, is that it may become less salient to judges and other court observers. The current CJRA reporting system has the benefit of focusing attention on the two semi-annual reports. The semi-annual reporting

[54]This would, of course, violate my identifying assumption that litigants *do not* file strategically. While this assumption does appear to be met at present, as litigants and judges learn more about the six-month list, there is no guarantee that they would not learn to file or schedule motions strategically in the future.

dates help to coordinate behavior. Policymakers, members of Congress, and the especially-interested layperson know to check the report on or after these dates, and judges know there is a high likelihood that the report will be read. If, on the other hand, a new list is published each day, then the public may become inured, and judges may feel less social pressure as a result.

Another solution would be to maintain the current system of two reports per year, but to incorporate an element of randomness into the process. For example, if reports were published on 2-3 randomly selected dates per year, then judges might respond as if the reports are continuously updating.

Finally, my preferred solution to the non-uniformity problem would be to incorporate *aggregate* statistics into the current six-month lists. That is, in addition to (or even instead of) reporting *currently* overdue motions and cases, Congress[55] should consider calling on the Admistrative Office to also report semiannual judge-specific aggregate statistics, like how many motions were pending for six-months or longer at any point in the prior six months, average time-until-disposition for different types of motions, etc. This proposal is somewhat similar to proposals for "income averaging," which have gained favor among some tax scholars in recent years[56] This proposal has the advantage of not only reducing variation in exposure to the six-month list, but it also avoids penalizing judges who take on unusually complex cases. Even if a judge is slow to dispose of one or two particularly complex cases, her peers can nonetheless discern from her aggregate statistics that the

[55]From a practical point of view, whether a particular amendment to the reporting requirements necessitates Congressional action is likely to depend upon whether judges view the amendment as bolstering or eroding their judicial independence."[I]n a system where key participants have incentives to resist . . . reform, change is much more likely to occur through the force of law than through the nonbinding, hortatory proposals [of] the Judicial Conference" (Peck, 1991).

[56]See, e.g., Batchelder (2003), who argues that income averaging avoids for income tax purposes avoids penalizing the poor, who are particularly likely to experience large and frequent income fluctuations.

slowness is not part of an overall tendency for slowness.

1.7.2 Removing incentives to compromise on quality

My analysis reveals mixed evidence on the question of whether judges are sacrificing quality for speed in response to the six-month list. Nonetheless, one could imagine another set of reforms aimed at further preventing this possibility.

First, it is worth noting one feature of the six-month list that may already be mitigating some quality effects. Recall that, in addition to reporting motions that have been pending for six months or longer, the Administrative Office is also directed to publish reports on *cases* that have been pending for three years or longer.[57] Also recall, from my discussion regarding Table 1.8, that judges appear to be inefficiently deferring work until after the summary judgment phase of a given case. This is what I referred to as "cutting corners." Insofar as the three-year list focuses attention on overall case duration, the three year list may mitigate the incentives to inefficiently postpone work until a later phase of the case. In other words, the three-year list may have the effect of reducing judicial myopia. More research should be done on the effects of the three-year list, but it may offer a road map for future improvements to the six-month list. Insofar as the three-year list is effective at reducing myopia, it may be beneficial to reduce its horizon, perhaps even reporting on cases that have been pending two years or longer. Of course, since the three-year list may have its own unintended consequences, broad policy recommendations are inadvisable until further research has been conducted.

The CJRA might also benefit from a reporting scheme that takes into account a broader set of metrics, including metrics unrelated to speed. According to the

[57] Judicial Improvements Act of 1990 § 103, 28 U.S.C. § 476 (2012).

multitask model previewed in Section 1.3 (and further specified in Appendix Section **??**), the tendency to compromise on quality stems from disparities between competing goals (e.g. speed, fairness, and accuracy) with respect to both monitoring costs and the power of incentives. In other words, since speed is more easily monitored than quality, *and* since the six-month list rewards speed but not quality, judges may compromise quality. While quality is inherently hard to monitor, recent scholarship has sought to measure it. For example, recent articles by Judge William Young of the U.S. District Court for the District of Massachusetts and Professor Jordan Singer propose a new metric for judicial productivity, which they call "bench presence" (Young and Singer 2013; Singer and Young 2014). Bench presence measures the time that a district judge spends on the bench, actively presiding over cases. By incorporating more holistic measures of adjudicative quality into the CJRA's judicial reporting scheme, we may eliminate some of the incentive to sacrifice quality for the sake of speed.

Of course, the inherent danger of including additional metrics in the six-month list is that those metrics will simply create new biases in judicial behavior. Moreover, at least as a matter of public perception, monitoring judges on *how* they decide matters before them—and not merely on *when*—may be interpreted by some as an unacceptable intrusion into judicial independence. One possibility, which requires more research, is to include ostensibly "neutral" metrics. These metrics would be intended not to convey some notion of "quality," but rather to simply indicate that something may be amiss. In other words, these metrics would serve as the "canary in the coal mine." For example, we may not have a strong prior for whether judges should be qualifying more or fewer expert witnesses, but if we observe that a particular judge is a major outlier, that may be an indication that the judge is compromising on some aspect of adjudicative quality. Additional research would be necessary in order to identify which metrics, if any, are ideal for reporting. Still, this too could raise concerns, not least of which is the erosion of judicial independence.

1.8 Conclusion

This paper presents one of the first empirical analyses of the causal effects of the six-month list on the speed and quality of civil adjudication. Aided by an original large-N motion-level dataset and a novel identification strategy based on quasi-random variation in exposure to the six-month list, I uncover two important findings. First, "shaming" works. That is, the six-month list has effectively accomplished its ostensible goal of promoting speedy adjudications. Motions that are most exposed to the six-month list are adjudicated almost 15% faster than those that are least exposed, and overall cases are adjudicated almost 2% faster as a consequence. Second, improved speed does not appear to have been achieved at a significant cost with respect to the *quality* of civil adjudications. While district court judges are slightly less likely to grant summary judgment when the motion is more exposed to the six-month list, the effect is small, marginally significant, and not robust to all specifications. Effects on appellate outcomes are similarly small and statistically insignificant. On the other hand, after controlling for the direct effect on motion processing time, it does appear that greater exposure to the six-month list actually *prolongs* overall case duration, suggesting that the six-month list may be causing judges to inefficiently "cut corners." I interpret the above results as broadly consistent with models of judicial behavior that emphasize career concerns, procrastination, and judicial multitasking. In the previous section, I discussed the normative implications of my findings. In particular, I suggest reforms aimed both at making the effects of the six-month list more uniform across motions and cases and at mitigating the six-month list's potential for adverse effects on adjudicative quality.

In addition to the main results, I find evidence of considerable heterogeneity across judges in their responsiveness to the six-month list. In particular, I find that young judges, non-white judges, and female judges are among the most sensitive to the six-month list. These findings, while preliminary, call attention to the ways in which non-traditional workplace incentives—here, the use of social sanctions—interact with worker characteristics like race, age, and gender.

My analysis suggests several avenues for future research. In particular, my finding that judges respond heterogeneously to the six-month list highlights the importance of additional research on judges' sensitivity to career concerns, with a particular emphasis on differences across race, gender, and age. Additional work is also necessary in order to better conceptualize and measure adjudicative "quality." While my analysis has relied on relatively easy-to-measure proxies like modes of motion disposition and appellate outcomes, additional research should probe alternative proxies for adjudicative quality. Examples could include the frequency with which judges grant oral argument and the frequency, content, and citation rates of written judicial opinions. Moreover, insofar as the preceding analysis is limited to "within-motion" and "within-case" effects, additional research is necessary in order to properly account for the aggregate effects of the six-month list, including spillovers across motions within a case and across cases that are on the docket of the same judge.[58]

[58]In ongoing work I will implement a "bunching" estimator in order to estimate the effects of the six-month list on the aggregate distributions of motion durations and outcomes. Bunching estimators were pioneered in the empirical tax literature, where they have been used for such purposes as estimating the effects of tax policy on labor supply (Saez 2010; Chetty et al. 2011).

The six-month list resembles a discontinuous "notch" in judicial incentives (Kleven and Waseem, 2013).

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1 Bibliography

Acemoglu, Daron and Simon Johnson (2005) “Unbundling Institutions,” *Journal of Political Economy*, Vol. 113, pp. 949–995.

Akerlof, George A. (1991) “Procrastination and Obedience,” *American Economic Review*, Vol. 81, pp. 1–19.

Ashraf, Nava and Oriana Bandiera (2018) “Social Incentives in Organizations,” *Annual Review of Economics*, Vol. 10, pp. 439–463.

Ashraf, Nava, Oriana Bandiera, and B. Kelsey Jack (2014) “No margin, no mission? A field experiment on incentives for public service delivery,” *Journal of Public Economics*, Vol. 120, pp. 1–17.

Autor, David H., Nicole Maestas, Kathleen J. Mullen, and Alexander Strand (2015) “Does Delay Cause Decay? The Effect of Administrative Decision Time on the Labor Force Participation and Earnings of Disability Applicants,” *NBER Working Paper*, Vol. No. 20840.

Batchelder, Lily L. (2003) “Taxing the Poor: Income Averaging Reconsidered,” *Harvard Journal on Legislation*, Vol. 40, pp. 395–452.

Besley, Timothy and Stephen Coate (2003) “Elected Versus Appointed Regulators:

Theory and Evidence," *Journal of the European Economic Association*, Vol. 1, pp. 1176–1206.

Chemin, Matthieu (2012) "Does Court Speed Shape Economic Activity? Evidence from a Court Reform in India," *Journal of Law, Economics, and Organization*, Vol. 28, pp. 460–485.

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri (2011) "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records," *Quarterly Journal of Economics*, Vol. 126, pp. 749–804.

Chetty, Raj, Emmanuel Saez, and Laszlo Sandor (2014) "What Policies Increase Prosocial Behavior? An Experiment with Referees at the Journal of Public Economics," *Journal of Economic Perspectives*, Vol. 28, pp. 169–188.

Chevalier, Judith and Glenn Ellison (1999) "Career Concerns of Mutual Fund Managers," *The Quarterly Journal of Economics*, Vol. 114, pp. 389–432.

on Civil Justice Reform, Brookings Task Force (1989) "Justice for All: Reducing Costs and Delay in Civil Litigation,"Technical report, The Brookings Institution.

Connolly, Paul R. and Saundra Smith (1983) "The Litigant's Perspective on Delay: Waiting for the Dough," *The Justice System Journal*, Vol. 8, pp. 271–286.

de Figueiredo, Miguel, Alexandra D. Lahav, and Peter Siegelman (Forthcoming) "The Six Month List and the Unintended Consequences of Judicial Accountability," *Cornell Law Review*, Vol. 105.

Dessem, R. Lawrence (1993) "Judicial Reporting Under the Civil Justice Reform Act: Look, Mom, No Cases!," *University of Pittsburgh Law Review*, Vol. 54, pp. 687–716.

Dusek, Libor and Josef Montag (2017) "The Effects of a Simplified Criminal Procedure: Evidence from One Million Czech Cases," *ISE Working Paper*, Vol. No. 1.

Epstein, Lee, William M. Landes, and Richard A. Posner (2011) "Why (And When) Judges Dissent: A Theoretical And Empirical Analysis," *Journal of Legal Analysis*, Vol. 3, pp. 101–137.

——— (2013) *The Behavior of Federal Judges: A Theoretical and Empirical Study of Rational Choice*: Harvard University Press.

Falkoff, Marc D. (2012) "The Hidden Costs of Habeas Delay," *University of Colorado Law Review*, Vol. 83, pp. 339–407.

Fama, Eugene F. (1980) "Agency Problems and the Theory of the Firm," *Journal of Political Economy*, Vol. 88, pp. 288–307.

Frakes, Michael D. and Melissa F. Wasserman (2016) "Procrastination in the Workplace: Evidence from the U.S. Patent Office," *NBER Working Paper*, Vol. No. 22987.

Gauri, Varun, Julian C. Jamison, Nina Mazar, and Owen Ozier (2019) "Motivating bureaucrats through social recognition: External validity—A tale of two states," *Organizational Behavior and Human Decision Processes*.

Gelbach, Jonah (2014) "Rethinking Summary Judgment Empirics: The Life of the Parties," *University of Pennsylvania Law Review*, Vol. 162, pp. 1663–1689.

Gertner, Nancy (2012) "Losers' Rules," *Yale Law Journal*, Vol. 122, pp. 109–124.

Harris, Milton and Bengt Holmstrom (1982) "A Theory of Wage Dynamics," *The Review of Economic Studies*, Vol. 49, pp. 315–333.

Hausman, Catherine and David S. Rapson (2018) "Regression Discontinuity in Time: Considerations for Empirical Applications," *Annual Review of Resource Economics*, Vol. 10, pp. 533–552.

Holmstrom, Bengt (1999) "Managerial Incentive Problems: A Dynamic Perspective," *The Review of Economic Studies*, Vol. 66, pp. 169–182.

Holmstrom, Bengt and Paul Milgrom (1991) "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design," *Journal of Law, Economics, and Organization*, Vol. 7, pp. 24–52.

Huang, Bert I. (2011) "Lightened Scrutiny," *Harvard Law Review*, Vol. 124.

Imbens, Guido W. and Thomas Lemieux (2008) "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, Vol. 142, pp. 615–635.

Johnston, Patrick (1994) "Civil Justice Reform: Juggling Between Politics and Perfection," *Fordham Law Review*, Vol. 62, pp. 833–882.

Judicial Conference of the United States (1998) "Report of the Proceedings of the Judicial Conference of the United States,"Technical report.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan (2015) "Self-Control at Work," *Journal of Political Economy*, Vol. 123, pp. 1227–1277.

Kleven, Henrik J. and Mazhar Waseem (2013) "Using Notches to Uncover Optimization Frictions and Structural Elasticities: Theory and Evidence from Pakistan," *Quarterly Journal of Economics*, Vol. 128, pp. 669–723.

Kondylis, Florence and Mattea Stein (2018) "The Speed of Justice," *World Bank Policy Research Working Paper*, Vol. No. 8372.

Laibson, David (1997) "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, Vol. 112, pp. 443–477.

Lavie, Shay (2016) "Appellate Courts and Caseload Pressure," *Stanford Law & Policy Review*, Vol. 27.

Lee, David S. and Thomas Lemieux (2010) "Regression Discontinuity Designs in Economics," *Journal of Economic Literature*, Vol. 48, pp. 281–355.

Levy, Gilat (2005) "Careerist Judges and the Appeals Process," *The RAND Journal of Economics*, Vol. 36.

Mathauer, Inke and Ingo Imhoff (2006) "Health worker motivation in Africa: the role of non-financial incentives and human resource management tools," *Human Resources for Health*, Vol. 24, p. 24.

McCrary, Justin (2008) "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test," *Journal of Econometrics*, Vol. 142, pp. 698–714.

Olson, Walter K. (1992) *The Litigation Explosion: What Happened When America Unleashed the Lawsuit*: Truman Talley Books.

Peck, Jeffrey J. (1991) ""Users United": The Civil Justice Reform Act of 1990?" *Law and Contemporary Problems*, Vol. 54, pp. 105–118.

Posner, Richard A. (2010) *How Judges Think*: Harvard University Press.

Post, Robert (1998) "Judicial Management and Judicial Disinterst: The Achievements and Perils of Chief Justice William Howard Taft," *Journal of Supreme Court History*, Vol. 23, pp. 50–78.

Resnik, Judith (1982) "Managerial Judges," *Harvard Law Review*, Vol. 96, pp. 374–448.

Rubin, Alvin B. (1980) "Bureaucratization of the Federal Courts: The Tension between Justice and Efficiency," *Notre Dame Law Review*, Vol. 55, pp. 648–659.

Saez, Emmanuel (2010) "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy*, Vol. 2, pp. 180–212.

Simon, Herbert A. (1955) "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics*, Vol. 69, pp. 99–118.

Singer, Jordan M. and William G. Young (2014) "Bench Presence 2014: An Updated Look at Federal District Court Productivity," *New England Law Review*, Vol. 48, p. 565.

Thaler, Richard H. and Cass R. Sunstein (2009) *Nudge: Improving Decisions About Health, Wealth, and Happiness*: Penguin Books.

Tobias, Carl (2002) "The Expiration of the Civil Justice Reform Act of 1990," *Washington & Lee Law Review*, Vol. 59.

Visaria, Sujata (2009) "Legal Reform and Loan Repayment: The Microeconomic Impact of Debt Recovery Tribunals in India," *American Economic Journal: Applied Economics*, Vol. 1, pp. 59–81.

Wasby, Stephen L. (1994) "Is Court Delay Unconstitutional?" *The Justice System Journal*, Vol. 17, pp. 267–270.

——— (1997) "Delay as a Due Process Violation," *The Justice System Journal*, Vol. 19, pp. 235–240.

Yang, Crystal S. (2016) "Resource Constraints and the Criminal Justice System: Evidence from Judicial Vacancies," *American Economic Journal: Economic Policy*, Vol. 8, pp. 289–332.

Young, William G. and Jordan M. Singer (2013) "Bench Presence: Toward a More Complete Model of Federal District Court Productivity," *Penn State Law Review*, Vol. 118, pp. 55–98.

Chapter 1 Tables and Figures

Figure 1-1: Histogram of Summary Judgment Motion Dispositions (by calendar day)
All Federal Civil Cases, 2004-2014

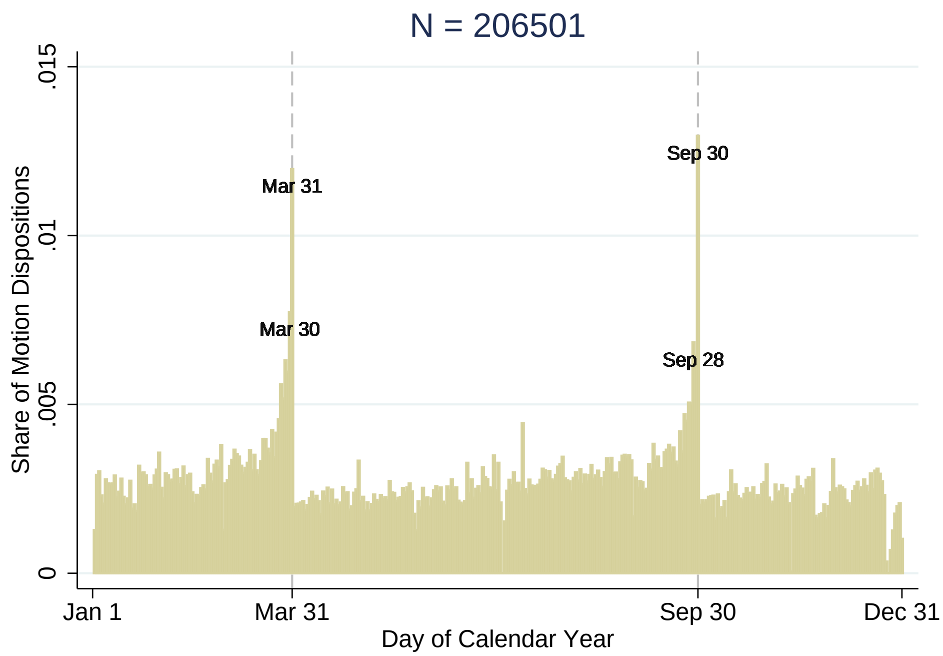


Table 1.1: Comparison of Means: Summary Judgments Decided Immediately Before 6-Month List Vs. All Others
All Civil Cases, (2004-2014)

	(1) Last Two Weeks	(2) All Other Weeks	(3) Difference in Means
Months Until Disposition	7.614	4.944	2.671
	(5.164)	(4.329)	[0.096]***
Reporting Time (months)	9.739	10.063	-0.323
	(1.749)	(1.743)	[0.042]***
% Granted	0.461	0.483	-0.022
	(0.498)	(0.500)	[0.006]***
% Granted in part	0.166	0.141	0.025
	(0.372)	(0.348)	[0.005]***
% Denied	0.361	0.363	-0.002
	(0.480)	(0.481)	[0.006]
% Decided for Plaintiff	0.258	0.278	-0.020
	(0.438)	(0.448)	[0.004]***
% Decided for Defendant	0.566	0.568	-0.002
	(0.496)	(0.495)	[0.005]
% Order Appealed	0.287	0.218	0.069
	(0.452)	(0.413)	[0.011]***
% Filed Pro Se	0.206	0.180	0.025
	(0.404)	(0.384)	[0.012]**
% In Forma Pauperis	0.196	0.154	0.042
	(0.397)	(0.361)	[0.013]***
N	32,058	449,204	481,262

This table presents a comparison of means between summary judgment motions decided in the two weeks immediately preceding the publication of a six-month list (that is, in the final two weeks of March and the final two weeks of September) and summary judgment motions decided in all other weeks of the calendar year. Details on the sample are provided in Section **??**. Columns (1) and (2) show sample means with standard deviations in parentheses, and column (3) shows differences in means with standard errors in brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.2: Summary Statistics, Summary Judgment Motions
All Civil Cases, (2004-2014)

	(1) Full Sample	(2) Ruled-On
% Filed by Plaintiff	0.28	0.29
	(0.449)	(0.455)
% Filed by Defendant	0.61	0.64
	(0.488)	(0.481)
Months Until 6-Month Report	10.04	10.03
	(1.745)	(1.748)
Months Until Disposition		5.36
		(4.572)
% Motion granted		0.48
		(0.500)
% Motion granted in part		0.14
		(0.352)
% Motion denied		0.36
		(0.481)
% Motion Decided for Plaintiff		0.28
		(0.447)
% Motion Decided for Defendant		0.57
		(0.495)
% Appealed		0.26
		(0.440)
Observations	481,262	206,513

This table presents summary statistics on the main motion-level dataset. Standard deviations are presented in parentheses below sample mean.

Figure 1-2: Histogram of Summary Judgment Motion Durations
(months until disposition)

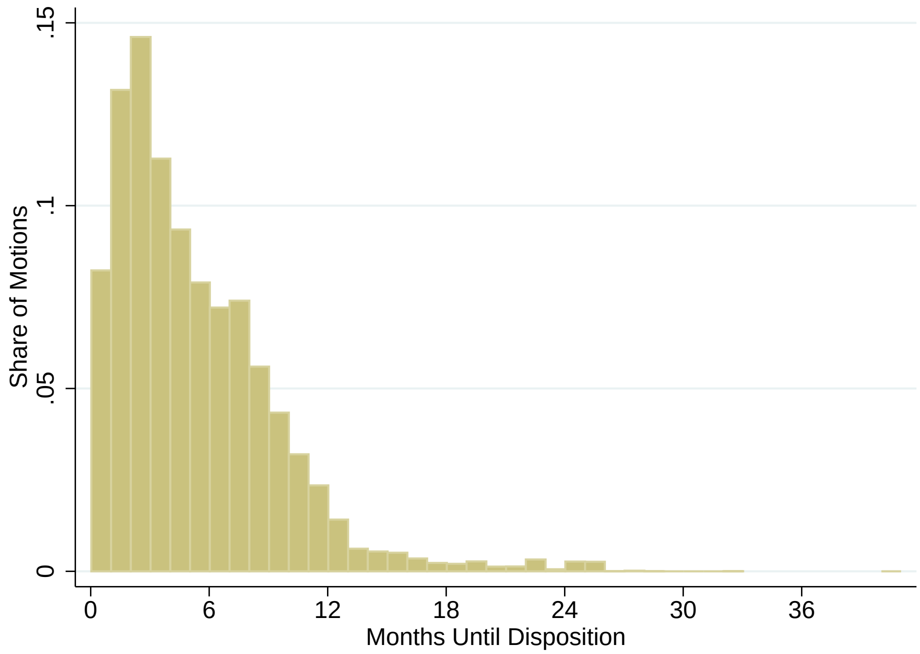


Figure 1-3: Examples of Six-Month List "Reporting Time"

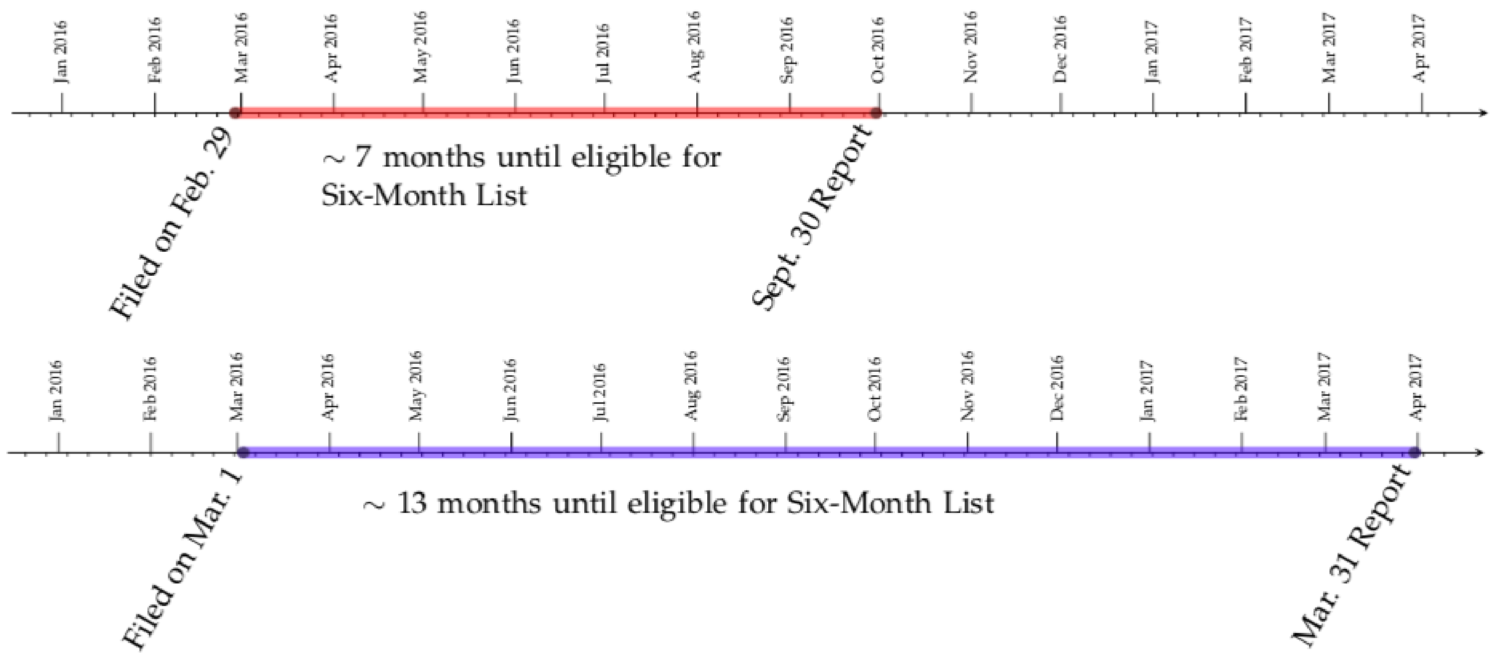


Figure 1-4: 6-month list "Reporting Time" as a function of filing date

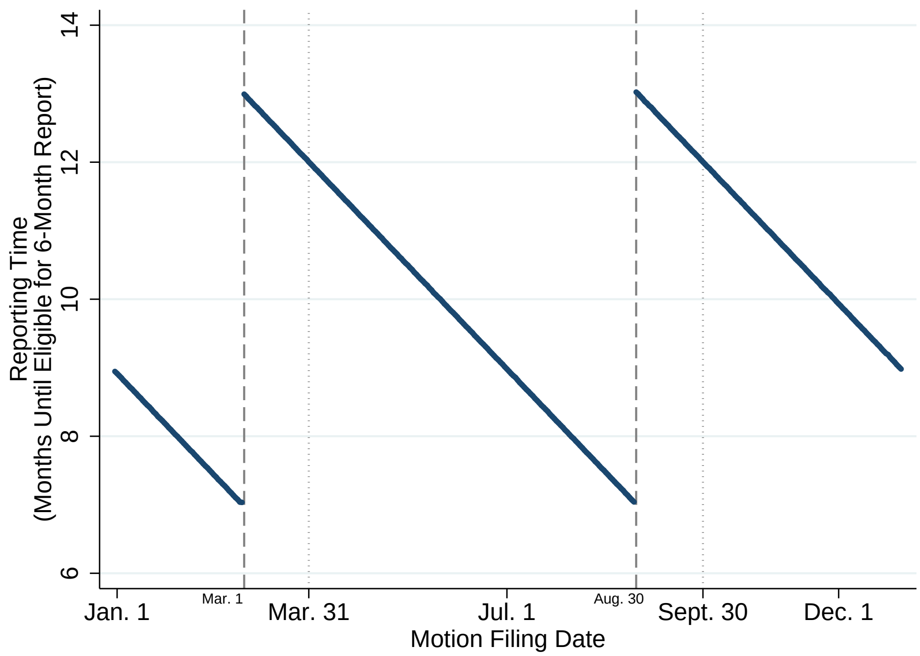


Figure 1-5: Histogram of Summary Judgment Motion Filings (by calendar day)
All Federal Civil Cases, 2004-2014

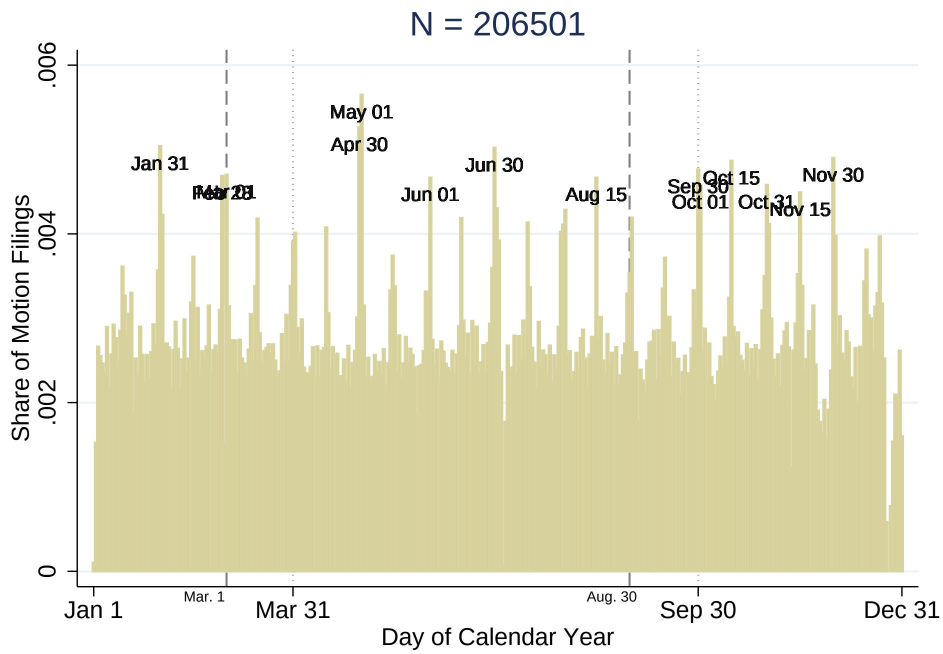
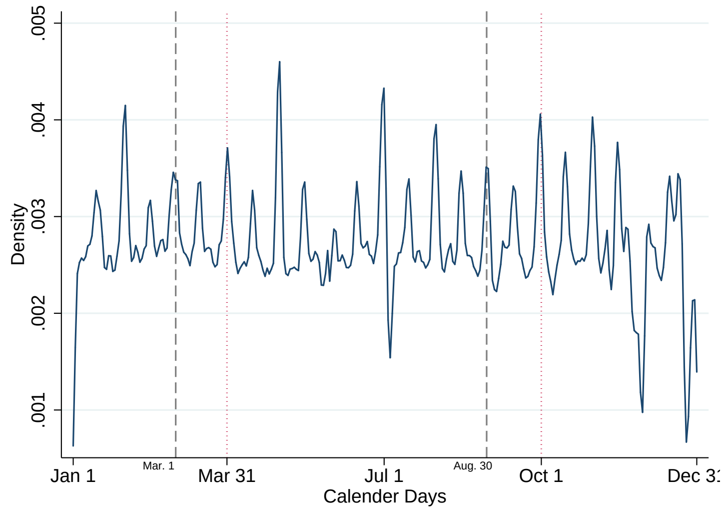
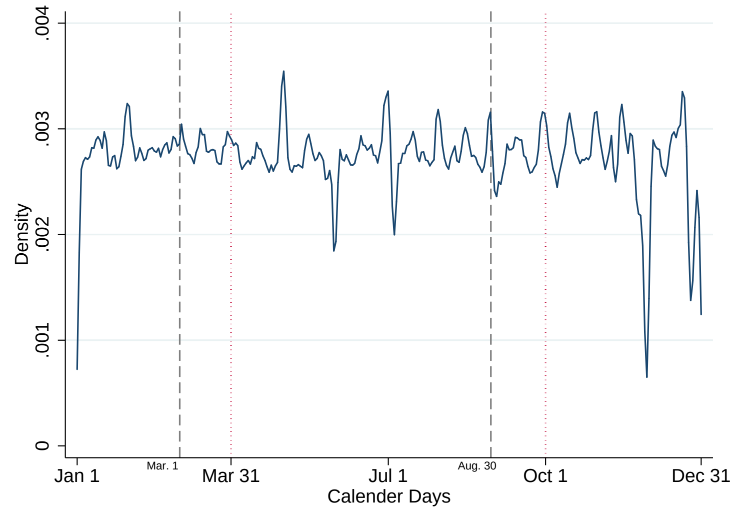


Figure 1-6: Summary Judgment Motion Filings (by calendar day)
All Federal Civil Cases, 2004-2014



(a) Raw Motion Filings



(b) Adjusted Motion Filings

Table 1.3: Comparison of Means: Low versus High Reporting Time Summary Judgment Motions, All Civil Cases, (2004-2014)

	(1) Low Reporting Time	(2) High Reporting Time	(3) Difference in Means
Reporting Time (months)	8.500	11.523	3.023
	(0.880)	(0.864)	[0.016]***
% Filed by Pltf.	0.280	0.281	0.001
	(0.449)	(0.450)	[0.003]
% Filed by Deft.	0.606	0.608	0.002
	(0.489)	(0.488)	[0.004]
% Pro Se	0.182	0.182	-0.001
	(0.386)	(0.386)	[0.002]
% I.F.P.	0.157	0.157	-0.000
	(0.364)	(0.363)	[0.002]
% Prisoner Rights	0.135	0.132	-0.002
	(0.341)	(0.339)	[0.002]
% Employment Discrim.	0.103	0.103	0.000
	(0.304)	(0.304)	[0.002]
% Personal Injury	0.120	0.123	0.003
	(0.325)	(0.328)	[0.007]
% Soc. Sec.	0.103	0.102	-0.001
	(0.304)	(0.303)	[0.002]
N	235,905	245,357	481,262

This table presents a comparison of means between summary judgment motions with low (i.e. less than 10 months) and high (i.e. at least 10 months) reporting time. Columns (1) and (2) show sample means with standard deviations in parentheses, and column (3) shows differences in means with standard errors in brackets.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1-7: Running Variable as a Function of Filing Date

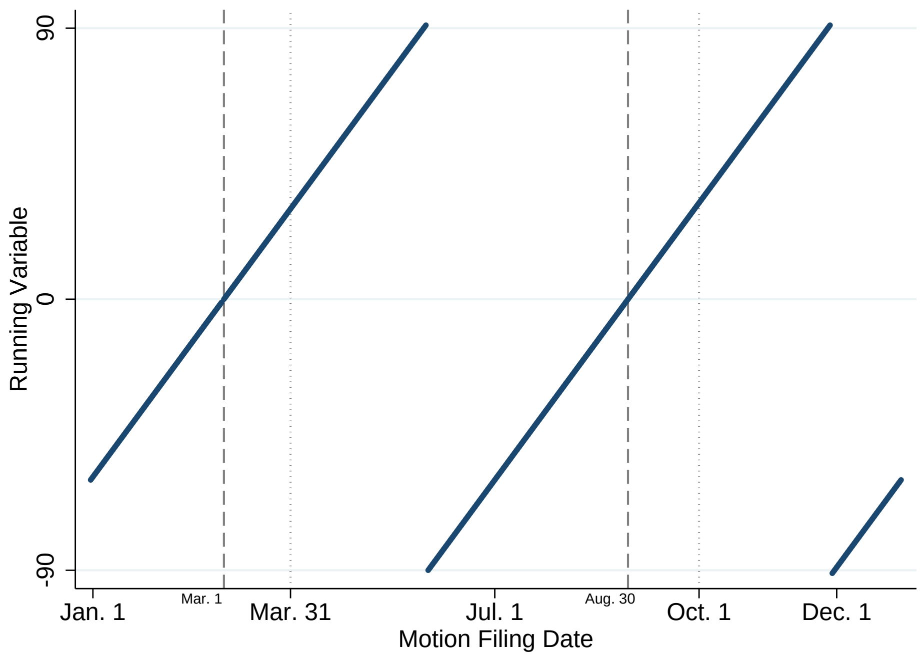
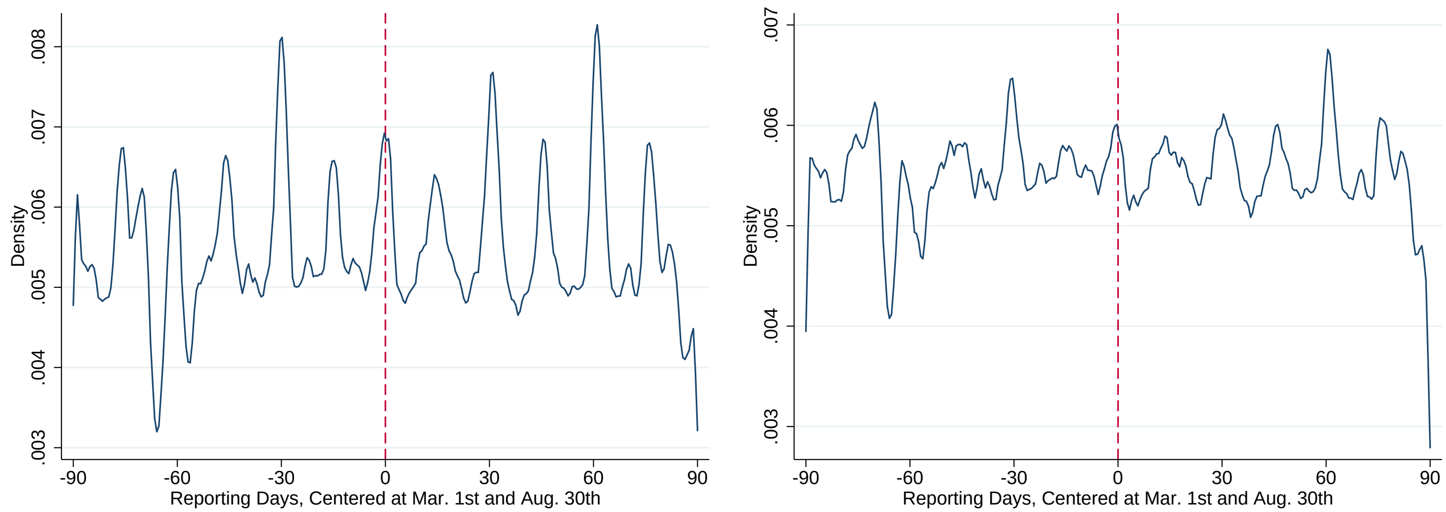


Figure 1-8: Distribution of Motion Filings by RD Running Variable



(a) Raw Motion Filings

(b) Adjusted Motion Filings

Figure 1-9: Distribution of Motion Duration, by Relative Reporting Time

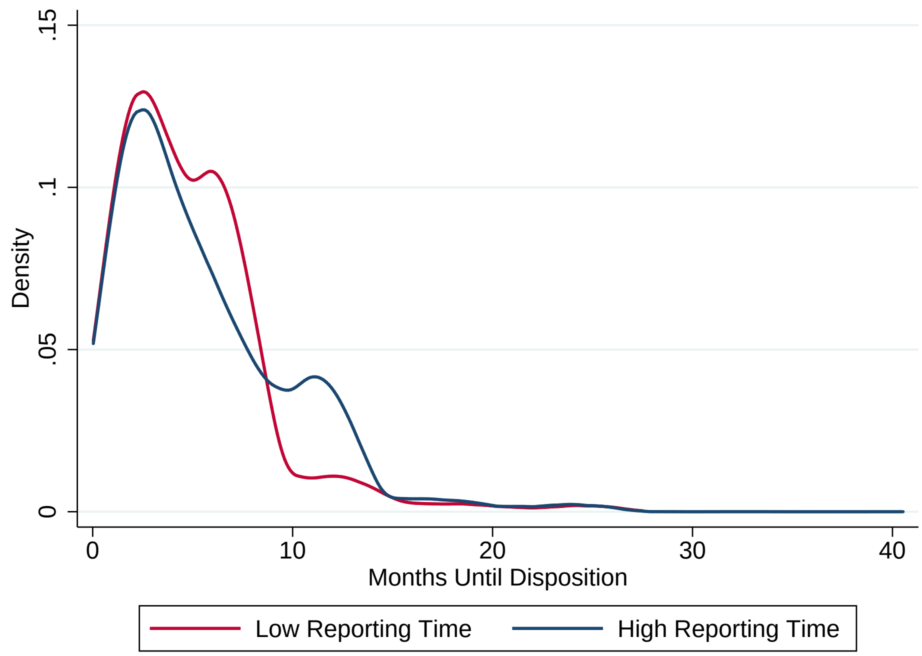


Table 1.4: Effect of Reporting Time on Months Until Motion Disposition
Summary Judgment Motions, All Civil Cases, (2004-2014)

	(1)	(2)	(3)	(4)
Months until Report	0.129***	0.133***	0.132***	0.132***
	(0.005)	(0.005)	(0.005)	(0.005)
Observations	206,187	206,187	206,151	206,151
Case & Motion Controls	Yes	Yes	Yes	Yes
Calendar Trends		Yes	Yes	Yes
District*Year FEs			Yes	Yes
Day-of-Month FEs				Yes
Mean of Dep. Variable	5.36	5.36	5.36	5.36
Mean of Indep. Var	10.03	10.03	10.03	10.03

This table presents OLS estimates of the effect of additional reporting time on months until motion disposition. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.5: Effect of Motion Reporting Time on Months Until Case Disposition Summary Judgment Motions, All Civil Cases, (2004-2014)

	(1)	(2)	(3)	(4)
Months until Report	0.062^{***}	0.082^{***}	0.086^{***}	0.080^{***}
	(0.018)	(0.019)	(0.019)	(0.019)
Observations	183,923	183,923	183,887	183,887
Case & Motion Controls	Yes	Yes	Yes	Yes
Calendar Trends		Yes	Yes	Yes
District*Year FEs			Yes	Yes
Day-of-Month FEs				Yes
Mean of Dep. Variable	23.38	23.38	23.37	23.37
Mean of Indep. Var	10.04	10.04	10.04	10.04

This table presents OLS estimates of the effect of additional motion reporting time on months until overall case disposition. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 1-10: Average Months Until Motion Disposition
by RD Running Variable

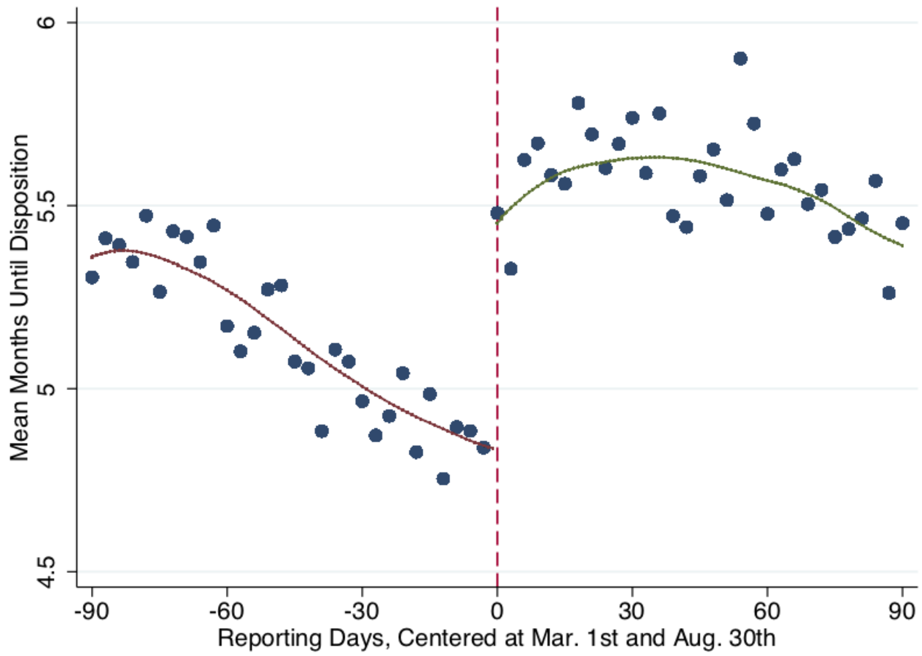


Table 1.6: Regression Discontinuity Estimates
Effect of Reporting Time on Average Months Until Disposition

	Parametric			Non-Parametric (Local Linear)	
	(1)	(2)	(3)	(4)	(5)
	Linear	Quadratic	Cubic	IK Bandwidth	CCT Bandwidth
Filed After Cutoff	0.849***	0.753***	0.630***	0.624***	0.363***
	[0.039]	[0.054]	[0.073]	[0.198]	[0.111]
Mean of Dep. Variable	5.36	5.36	5.36	5.22	5.26
Observations	204137	204137	204137	6826	51247

This table presents regression discontinuity (RD) estimates of the effect of additional reporting time on total motion duration. The running variable represents the motion filing date relative to the six-month list eligibility cutoff. Motions filed just before the cutoff are eligible for the current six month list, whereas motions filed just after the cutoff have an additional six months before they might appear on a list. Columns (1)-(3) are estimated parimetrically with linear, quadratic, and cubic polynomials, respectives. Columns (4)-(5) are estimated nonparametrically with local linear regressions, using the IK and CCT methods of optimal bandwidth selection, respectively. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.7: Effect of Motion Reporting Time on Probability of Selected Motion & Appellate Outcomes Motions Filed by Defendants

	Motion Outcomes			Appellate Outcomes		
	(1) Granted	(2) Denied	(3) Granted In Part	(4) Appealed	(5) Affirmed	(6) Reversed
Months until Report	0.0019**	-0.0009	-0.0006	0.0003	0.0018	-0.0013
	[0.0008]	[0.0007]	[0.0007]	[0.0006]	[0.0017]	[0.0009]
Observations	131,406	131,406	131,406	131,406	34,390	34,390
Case & Motion Controls	Yes	Yes	Yes	Yes	Yes	Yes
Calendar Trends	Yes	Yes	Yes	Yes	Yes	Yes
District*Year FEs	Yes	Yes	Yes	Yes	Yes	Yes
Day-of-Month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Motion Outcome Dummies				Yes	Yes	Yes
Mean of Dep. Variable	.57	.26	.16	.26	.53	.08
Mean of Indep. Var	10.03	10.03	10.03	10.03	10.04	10.04

This table presents OLS estimates of the effect of additional reporting time on probability of various motion-level outcomes for summary judgment motions filed by a defendant. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.8: Effect of Motion Reporting Time on Months Until Case Disposition Controlling for Direct Effect on Motion Duration

	(1)	(2)	(3)
Months until Report	0.080***	-0.052***	-0.048***
	(0.019)	(0.017)	(0.018)
Months until Motion Disposition		1.035***	1.001***
		(0.010)	(0.010)
Motion Granted			-2.740***
			(0.074)
Motion Granted in Part			1.874***
			(0.106)
Motion Mooted			-1.796***
			(0.170)
Appeal Filed			3.147***
			(0.081)
Observations	183,887	183,887	183,887
Case & Motion Controls	Yes	Yes	Yes
Calendar Trends	Yes	Yes	Yes
District*Year FEs	Yes	Yes	Yes
Day-of-Month FEs	Yes	Yes	Yes
Mean of Dep. Variable	23.38	23.38	23.38
Mean of Indep. Var	10.04	10.04	10.04

This table presents OLS estimates of the effect of additional motion reporting time on months until overall case disposition. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.9: Effect of Reporting Time on Months Until Motion Disposition by Judge Characteristics

	(1)	(2)	(3)	(4)	(5)
Months until Report	0.128***	0.126***	0.131***	0.135***	0.133***
	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)
Reporting Time * Judge Under 55	0.041***				
	(0.014)				
Reporting Time * Non-White Judge		0.059***			
		(0.016)			
Reporting Time * Female Judge			0.023*		
			(0.013)		
Reporting Time * Chief Judge				0.006	
				(0.017)	
Reporting Time * Same-party Judge					0.007
					(0.011)
Observations	170,950	170,950	170,950	170,950	170,950
Case & Motion Controls	Yes	Yes	Yes	Yes	Yes
Calendar Trends	Yes	Yes	Yes	Yes	Yes
District*Year FEs	Yes	Yes	Yes	Yes	Yes
Day-of-Month FEs					
Mean of Dep. Variable	5.44	5.44	5.44	5.44	5.44
Mean of Indep. Var	10.03	10.03	10.03	10.03	10.03

This table presents OLS estimates of the heterogeneous effects of additional reporting time on months until motion disposition for various judge characteristics, including whether the judge is under 55 years old, non-white, female, whether the judge is the Chief Judge of a district court, and whether the judge was appointed by a President of the same party as the current President at the time of the motion filing. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Where the judge characteristic is time-varying (e.g. judge's age, or whether judge is of same party as the President), the un-interacted judge characteristic is also included. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 2

I Want You! (But Not You): Selection in Military Retention

Joint with Christina Patterson and William Skimmyhorn

2.1 Introduction

The public sector is a large and important part of the economy. Approximately 15% of U.S. workers are employed by the federal, state, or local governments and the public sector also produces public goods that are key to economic growth. Existing studies document the impact of public sector worker quality on a variety of important public sector outputs including education (Chetty et al., 2014), nursing (Aiken et al., 2003), law enforcement (Rydberg and Terrill, 2010), and political leadership (Besley et al., 2011). However, the public sector is unusual in the constraints it imposes on the compensation and management of personnel and in its relative insulation from direct competition. As a result, the determinants of selection into the public sector has been a longstanding question in economics, spanning fields from labor and public finance (Katz and Krueger (1991); Borjas (2002)) to development and political economy (Dal Bo et al. (2013); Deserranno (2019)) and national security (Friedman (1967); Simon and Warner (2007); Korb and Segal (2011)). Existing research has focused primarily on understanding how differences in the levels of compensation across the public and private sectors affect who decides to enter government service (see, e.g., Dal Bo et al. (2013); Finan (2017); Nickell and Quintini (2002); Bacolod (2007)).[1]

In this paper, we bring new evidence to this literature and provide well-identified estimates of the effects of commonly used public sector compensation policies on the quality of public sector employees. We also expand the scope of this line of research by studying these effects in the context of retention policies, as opposed to the better-studied effect of wages on the entry margin. Because public sector personnel managers typically lack the same tools as private sector managers to individually adjust compensation, they instead frequently rely on a limited menu of retention policies and incentives, including retention bonuses and retirement incentives. These policies, almost all of which feature lump-sum cash payouts, are known to be effective at increasing the quantity of retained workers.[2] In this paper, however, we show that they also meaningfully affect the types of workers who elect to remain in the public sector. In particular, we study how key reten-

[1]There is a modest related literature on military recruitment and retention, almost all of which has analyzed (Brown (1985); Warner et al. (2003); and Gelber (2007)) or modeled (Gotz and McCall (1984) and Daula and Moffitt (1995)) enlistment and retention quantities, with little attention to worker quality. Among the papers studying military personnel, our work is most closely related to Warner and Pleeter (2001) and Simon et al. (2015), who estimate personal discount rates using military drawdown policies—including those studied here. However, our paper is the first to establish the causal effects of these different types of compensation on the quality of retained workers and the implications this has for the aggregate workforce quality.

[2]See, e.g., Asch et al. (2010).

tion incentives affect worker sorting in the U.S. military. In contrast to much of the literature showing that higher levels of compensation induce higher quality workers to enter the public sector, we find that more generous lump-sum retention incentives actually induce lower ability workers to remain. Our findings highlight that the structure, rather than just the level, of compensation matters in determining the quality of retained public servants. We show that these effects are large enough to affect the average ability level of the organization's overall workforce, a finding that should draw increased attention to how commonly used retention policies are designed and deployed.

Our results are somewhat striking in light of both the existing empirical literature and predictions made by the simplest models of selection. In a simple model in which returns to individual ability are higher in the private sector than in the public sector,[3] and where workers differ only in their ability, one would expect any increases in public sector compensation—even those that are not specifically targeted towards higher ability workers—to increase the average ability of those who select into the public sector. Indeed, this prediction that higher wages attract higher quality workers is consistent with the selection patterns documented by Dal Bo et al. (2013) and throughout much of the literature on the personnel of the state (Finan, 2017). However, our results on the retention margin are inconsistent with this simple model of selection. Instead, we find that, because low-ability workers are more responsive than their higher ability peers to a lump-sum retention bonus, generous retention incentives can actually reduce average ability levels. These results support a richer model with additional dimensions of worker heterogeneity, and they demonstrate that the design of retention policies can be crucial for retaining high-ability workers.

[3]See Borjas (2002); Katz and Krueger (1991).

Our setting is the U.S. Army, where we combine rich micro-data with a policy environment that generates plausibly exogenous variation in the relative returns to continued employment in the military. Specifically, we study how soldiers of different ability levels respond to two common types of retention policies: 1) lump-sum retention bonuses and 2) offers of early retirement benefits. The U.S. military provides a useful setting in which to study questions relating to the public sector more broadly, as key features of military compensation are relatively common across the public sector but comparatively rare in the private sector. First, the military sets wages according to a highly standardized pay scale with minimal variation based on individual abilities. Second, the military offers a generous but cliff-vested (at 20 years of service) defined benefit pension, which substantially shifts compensation to the future and creates unique retention incentives.[4] Third, the military often uses large recruitment and retention bonuses as relatively blunt tools for either growing or shrinking the overall size of the force. These three features are prevalent across other public sector organizations at the federal and state levels. For example, defined benefit pensions remain more common today in the public than private sectors (Poterba et al., 2007), and the retention bonuses and early retirement incentives we study are frequently relied upon by other public sector organizations seeking to affect retention (e.g., the U.S. Postal Service, Social Security Administration, and the U.S. Border Patrol).[5] The

[4]In 2018 the military replaced its defined benefit system with a "blended" defined benefit and defined contribution system. Our data are confined to the 1992-2016 time period, when the military relied on a pure defined benefit system.

[5]As of January 2018, the U.S. Postal Service, Social Security Administration, Small Business Administration, and Environmental Protection Agency all offered early retirement policies to thousands of employees. See `http://www.fedweek.com/fedweek/usps-offering-round-early-retirements`; `https://www.govexec.com/management/2017/10/agency-jobs-watch-how-will-your-agency-cut-its-workforce/137905/`. Additionally, members of Congress have recently proposed greater use of recruitment and retention bonuses in the United States Border Patrol, which is

military is especially intriguing because military retirement often occurs in middle age (Kamarck, 2018). In contrast to the existing retirement literature, which has been primarily concerned with workers at the very end of their careers, studies of the military may enhance our understanding of how retirement incentives affect the transitions of skilled workers in the mid-to-late parts of their careers.[6]

Not only does the military mirror many of the dynamics affecting public sector organizations at large, but given its size and economic importance,[7] the military is also worth studying in its own right. Recently, policymakers have expressed concern that the U.S. military in particular is failing to retain its best and brightest members, particularly among commissioned and non-commissioned officers, who comprise the middle and upper-level "management" of the military.[8] In fact, our own data validates their concerns and shows that the enlisted soldiers who stay in the Army the longest tend to be the ones with the lowest average scores on pre-enlistment aptitude tests (see Figure 2-1). Compared to soldiers who exit the Army after a single enlistment, soldiers who serve 20 years or more have an average Armed Forces Qualification Test (AFQT) score that is almost half of a standard deviation lower. Military analysts have suggested that the military's retention policies should be designed to optimize not only quantity retained, but also the quality of those retained, as they argue that retaining a more talented workforce increases productivity, boosts morale, and ultimately saves costs (Wardynski et al.

said to be experiencing a "brain drain". See `https://www.foxnews.com/us/border-patrol-brain-drain-agency-losing-more-agents-than-it-can-hire`.

[6]Specifically, our paper contributes to a larger literature quantifying the effects of retirement programs on labor supply, which has focused primarily on the relationship between retirement decisions and pensions (e.g., Brown (2013)). We add to this literature by studying mid-career workers and by studying the heterogenous response of workers of different ability levels.

[7]Including civilian employees, the Department of Defense is the world's single largest employer. See `https://www.forbes.com/sites/niallmccarthy/2015/06/23/the-worlds-biggest-employers-infographic/#78410ba5686b`.

[8]See, e.g., Wardynski et al. (2010). See, also, Kane (2012).

(2010), Wallace et al. (2015)). However, there is little concrete empirical evidence on the nature of selection in military retention. In Appendix B.1, we show that the key parameter to inform policy makers of how retention policies will affect the average quality of retained soldiers is precisely the object we estimate—the differential sensitivity of soldiers of varying abilities to potential reenlistment incentives.

Our empirical strategy leverages two sources of quasi-random variation in the financial returns to reenlisting in the military. First we study Selective Reenlistment Bonuses (SRBs), which offer a lump-sum payment to soldiers who choose to reenlist. SRB offers fluctuate frequently in response to changes in the Army's demand for soldiers of different ranks and skill sets, but importantly for our purposes, they are offered to all soldiers of a given rank and specialty regardless of individual ability. Second, we study early retirement incentives, which offer soldiers immediate (but reduced) retirement benefits in exchange for early exit from the military. Like the reenlistment bonuses, they were applied without regard to individual ability.

Our analysis shows that low-ability soldiers are more responsive to both types of near-term reenlistment incentives. Specifically, a 10 point decrease in a soldier's AFQT score (approximately equivalent to one-half of a standard deviation) is associated with a nearly one percentage point increase in the effect of a $10,000 SRB offer on a soldier's probability of reenlistment. Even more striking, soldiers with upper quintile AFQT scores are totally unresponsive to bonus offers. We find similar results using a soldier's speed of promotion as an alternative measure of ability. We also find that lower ability soldiers are more responsive to early retirement programs, and that of the soldiers who leave the military in direct response to early retirement programs, almost two-thirds have below-median AFQT scores.

We show that the increased sensitivity of low-ability soldiers to lump-sum bonuses is not consistent with a simple model in which the return to ability is lower in the military than in the civilian sector. Rather, we show that this excess sensitivity could be due to differences in unobservable taste for the military. We also show that the observed selection patterns persist even after controlling for variables proxying for soldiers' access to credit and discount factors. This finding suggests that differences in liquidity constraints and personal discount rates are not the primary explanations for the excess sensitivity of low-ability soldiers to lump-sum cash incentives.

The rest of the paper proceeds as follows. Section 2.2 describes our institutional setting and Section 2.3 describes our data. We present our empirical strategies and results in Section 2.4. Section 2.5 explores explanations for our primary finding, and Section 2.6 concludes

2.2 Institutional Setting

We analyze the reenlistment decisions of enlisted members of the all-volunteer U.S. Army between 1992 and 2016. Reenlistment is uniquely important in the military given its restricted lateral entry. Unlike private firms, which are free to hire at all levels, the military cannot simply hire more Sergeants or more Generals; instead, it must promote from within. Enlisted soldiers serve for fixed terms, and the typical first term of service lasts four years. At the end of each term, soldiers deemed eligible to reenlist (based on their previous performance) meet with a counselor to discuss their options which normally include opportunities to reenlist for an additional term of between two and six years. The counselors will also discuss the monetary and other potential benefits of remaining in the Army as well as potential opportunities in the civilian labor market. While reenlistment policies have changed some over time, eligible soldiers can typically reenlist between 12 months and 90 days prior to the end of their term.[9] Just after basic training, soldiers receive their Military Occupational Specialty (MOS), which corresponds to the job they will perform in the Army. A soldier's MOS is one of the most salient and important features of her individual experience in the Army, and while mid-career changes are possible, they are the not common.

We utilize two measures of individual ability—the AFQT score and the soldier's speed of promotion in their first term. A substantial body of previous research has established that a soldier's cognitive ability affects her on-the-job performance.[10] Wigdor and Green (1991) undertook an ambitious study of U.S. military performance and found that a soldier's AFQT is highly correlated with both hands-on performance and written knowledge of her job. Observed correlations range from 0.10 to almost 0.70, and the highest correlations tend to be in combat occupations. (See Appendix Table B.1.) Other studies have documented that AFQT scores explain individual and group performance in technical fields such as communications (Winkler et al. 1992; Fernandez 1992), air defense systems (Orvis et al., 1992), and automotive and helicopter maintenance (Mayberry and Carey, 1997). AFQT scores also predict early service attrition (Flyer and Elster 1983; Teachout and Pellum 1991; Horowitz and Sherman 1980). Finally, while most of the existing studies have focused on enlisted personnel, recent military research highlights the importance of cognitive ability for military officers as well (Condly et al., 2017).

[9]Figure B-6 in the appendix shows the distribution of the gap between the beginning of the eligibility window and the expected end of service. For the large majority of soldiers, this is either 12, 15, or 24 months. See Appendix Section B.2.1 for more details.

[10]For a review of the literature on human capital and military performance, see Kavanagh (2005).

Like many public sector compensation schemes, the military pay system has some unique features that distinguish it from the private sector. Military basic pay is a function of only rank,[11] years of service, and dependents status. The military also offers generous additional benefits, such as enlistment bonuses, periodic retention bonuses, education benefits, housing allowances, and a generous retirement program. The military's pension system is especially distinctive. Prior to 2018 (and throughout our sample), the U.S. military offered only a defined benefit plan to servicemembers. Active duty service members were eligible for a retirement pension only after 20 years of service, and soldiers who separated prior to 20 years received no retirement pay whatsoever. A soldier who separated with 20 years of service received an annual pension valued at approximately 50 percent of her final annual salary, and soldiers who retired after more than 40 years received up to 100 percent of their final salary. Notably, a retired soldier begins receiving her annual pension immediately upon retirement from the military, regardless of the soldier's age or employment status. Since many soldiers enlist at just 18 to 20 years of age, a soldier as young as 38 can be "retired" and receiving a military pension.[12]

2.2.1 Variation in Military Retention Policies

We leverage two particular military retention policies that generate quasi-random variation in the relative return to continued military service. Our first policy is

[11]Throughout this paper we refer to ranks by their corresponding pay grades. A pay grade consists of a letter—"E" for enlisted personnel, and "O" for commissioned officers—followed by a number, denoting the relative position of the rank. For example, an E-5 (Sergeant) is superior by two ranks to an E-3 (Private First Class).

[12]Although the purely defined benefit system was replaced with a "blended" defined benefit and defined contribution system in 2018, the defined benefit portion still cliff vests at 20 years, and it will likely still account for the majority of most servicemembers' retirement savings.

the Army's Selective Reenlistment Bonus (SRB) program. SRBs are cash bonuses offered to certain reenlistment-eliglble soldiers nearing the end of an enlistment term in order to encourage reenlistment. SRB offers vary by the soldier's current rank, the MOS that the soldier chooses to fill upon reenlistment, the soldier's total years of service, certain specialty skills the soldier might possess (for example, "airborne" qualification), the number of years for which the soldier reenlists, and the location in which the soldier is willing to be stationed. Depending upon her characteristics, a soldier may be eligible for a menu of several different SRB offers, and it is up to the soldier which SRB offer (if any) she accepts. SRB offers generally range from $0 to as high as $20,000. In our sample, the average SRB bonus received was $1,891, but among the 11% of soldiers who received a non-zero bonus, the average was $9,150. Compared to a soldier's base pay (e.g., in 2015, an E-4 with four years of service earned just over $28,000 annually), SRBs frequently represent a sizeable share of overall compensation.

The second set of policies we consider comprises the military's early retirement programs. In the early 1990s, after the Cold War ended, the Department of Defense implemented two programs—Voluntary Separation Incentives and Special Separation Benefits (VSI/SSB), and the Temporary Early Retirement Authority (TERA) program—as part of a larger "drawdown" strategy. Both programs were offered in two waves over the course of the early 1990s. In addition to reducing its overall size, the Army sought to reshape its force for the post-Cold War era by directing separation and retirement incentives at certain MOS and rank combinations.

We specifically study the second wave of the TERA program (August 1994 through July 1995), which offered early retirement to soldiers with at least 18 but less than 20 years of service who also met specific service requirements within their occupation and rank. The program was small overall, with only 1,731 eligible soldiers, which reflects 0.6 percent of all soldiers serving at that time and 6.8 percent of soldiers with at least 15 years of experience (see Appendix Table B.5). The benefits bestowed by TERA were generous. While soldiers are generally ineligible for retirement benefits prior to 20 years of service, TERA entitled recipients to an immediate military pension, albeit at a slightly reduced rate. Specifically, a soldier retiring under TERA had her military pension reduced by approximately 5% for each year less than 20.[13]

We also exploit variation from the VSI/SSB program, which offered inducements to mid-career soldiers who were willing to voluntarily separate from the Army pre-retirement. We focus our VSI/SSB analysis on the second wave of the program (August 1993 through June 1995). The VSI/SSB program was offered to soldiers who had 1) completed their first full term of service and 2) had accrued more than 6 but less than 20 years of service as of December 5, 1991.[14] Among that set of soldiers, eligibility was further restricted to certain occupation and rank combinations. The VSI/SSB programs were significantly larger than the TERA program—7,326 soldiers were eligible, covering 3.8 percent of all soldiers serving at that time and 11.7 percent of soldiers with at least 6 years of experience.

The VSI and SSB programs shared identical eligiblity rules, but the benefits provided by the two programs differed significantly, with VSI offering an annuity payment and SSB offering a single lump-sum payment upon separation. Soldiers had the option of choosing between the two programs. A soldier electing the VSI program received an annual payment equal to 2.5% of the soldier's final base

[13]More specifically, the retirement pay formula for TERA is $0.025 *$ years of service $*$ final base pay $*$ reduction factor, where the reduction factor is $\frac{m}{240}$ and where m is the number of full months served as of the retirement date.

[14]Both programs also requires that the soldier enter the reserves for several years.

annual pay multiplied by her total years of service, paid out once a year for twice the number of years of service. A soldier electing the SSB program received a single payment valued at 15% of her final base annual pay multiplied by her total years of service (i.e., a soldier with 7 years of service had a SSB payment just larger than her annual salary). For mid-career and senior soldiers, VSI/SSB and TERA eligibility had a major effect on the relative returns to continued military service.[15]

2.3 Data

We use the U.S. Army's Total Army Personnel Database (TAPDB) to construct a panel of enlistment spells from 1992 to 2016. Each observation (or "spell") corresponds to a single enlistment term for a soldier (e.g., a soldier who has served a single enlistment of four years will have just one observation, while a soldier in her tenth year of service will have multiple observations). We exclude all current enlistment spells (approximately 6%) since we do not observe their conclusion. We provide summary statistics for our sample in Table 2.1. The sample is primarily male with an average age of 28 and an average service duration of 6.33 years. For all analyses, we restrict our attention to those soldiers eligible to reenlist at the end of the term (Column 2), who look observably similar to the overall sample. The last two columns show the average characteristics of individual spells

[15]Before being granted the benefits of either TERA or VSI/SSB, eligible soldiers who decided to take up the program had to be approved by their commander. Eligible soldiers were able to apply to these programs at any time, regardless of whether they were in their reenlistment window or not. One may be concerned that although all soldiers within a rank, occupation and year of service bin were eligible, the approving commander may take the soldier's performance and aptitude into account when granting approval. While this is possible, evidence from Army archives suggest this was not the case. In fact, according to the Army's Fiscal Year 1992 "Historical Summary," 100% of on-time VSI/SSB applications were approved that year (see `http://www.history.army.mil/books/DAHSUM/1992/ch07.htm`).

that end in the soldier choosing to leave the Army (Column 3) or with the soldier reenlisting (Column 4). Around 50 percent of soldiers serve for only a single enlistment, and the average number of enlistments per soldier is 2.8.[16] On average, soldiers deciding to reenlist are more likely to be married and slightly younger than those who do not.

Our primary measure of ability is a soldier's AFQT score, which reflects the soldier's vocabulary, reading comprehension, and mathematical skills. The military uses the AFQT for initial selection (i.e., eligibility to enlist) and classification (i.e., eligibility for certain occupations), and labor economists have used these scores widely as a measure of individual cognitive ability (e.g., Grilliches and Mason 1972). AFQT scores range from 0-99, corresponding to the percentile of the applicant's raw test score.[17] Table 2.1 shows that soldiers eligible to reenlist have higher scores than those who are ineligible (Column 1 vs. 2), and that those who choose to reenlist have lower scores than those who leave (Column 3 vs. 4). Indeed, Appendix Figure B-8 shows that at every year of service, lower AFQT soldiers are more likely to reenlist.

While evidence suggests that AFQT scores are good predictors of military performance, cognitive measures may not capture all dimensions of ability relevant to the military. For that reason, we complement AFQT scores with a variable related to the speed of a soldier's promotions, which is commonly used to measure military aptitude. In particular, we observe the number of months in a soldier's first term that she spent below the rank of Sergeant, with larger numbers reflecting slower advancement. As expected, Appendix Figure B-7 shows that AFQT

[16]See Appendix Figure B-5 for the full distribution of the number of enlistments per soldier.

[17]Note that percentiles are determined with respect to the full population of test-takers. Because the military restricts enlistments to those above a minimum score—typically in the vicinity of 30—the median and mean AFQTs within the military exceed 50.

and speed of promotion are positively correlated both overall and within a range of occupations.[18]

In addition to personnel data, we collect monthly SRB offers and eligibility criteria for the VSI/SSB and TERA programs from publicly available policy announcements ("U.S. Army Military Personnel Messages").[19] We record the amount of the offer and the eligibility requirements (i.e., MOS, rank, years of service, and any special conditions) for each SRB. We construct the SRB offer data to isolate the exogenous aspects of the program (i.e., the variation in SRB offers that is uncorrelated with soldiers' choices). Specifically, we define the soldier's SRB offer as the bonus that is available for a 4-year reenlistment with the soldier's current occupation, rank, skill level, and tenure. This assignment process abstracts from the variation in SRBs that results from soldiers switching occupations in order to take advantage of a high SRB offer in a different occupation.[20] We exclude SRB offers that require moving to a particular location or unit, as they might reflect endogenous location preferences.[21] Finally, since monthly bonus offers may vary throughout the reenlistment window, we expect that soldiers may delay reenlist-

[18] We have also explored several alternative specifications of soldier promotion speed and find very similar results across alternative parameterizations. We chose the time the soldier took to get to rank E-5 (Sergeant) as a baseline because it is highly predictive of future promotion speeds and has a reasonable amount of variation among first term soldiers (See Table B.2).

[19] We are grateful to the authors of Greenstone et al. (2018), who shared with us the bonus offer data for the period 1997-2010. We have extended the dataset through 2016. Eligibility criteria for the VSI/SSB and TERA programs were announced in two separate Military Personnel Messages, both published in 1993. Unfortunately, these memoranda were not stored electronically, and copies of the final messages were destroyed in the Pentagon during the 9/11 attack. We therefore constructed the eligibility criteria from a pair of draft messages, which the Army had preserved. While we are are confident that the final rules were similar to the draft messages, we cannot be certain that they were identical.

[20] In fact, 23 percent of soldiers in our sample switch occupations upon reenlistment, and the average reenlistment term in the sample is 4.18 years. Appendix Table B.3 shows that SRB offers are highly correlated across the length of reenlistment terms.

[21] Appendix Table B.4 shows that general bonus offers and simultaneously offered location-specific bonus offers are highly correlated.

ment if they anticipate that a higher bonus offer is imminent, and this sort of behavior may be more common among high-ability soldiers. To eliminate this strategic timing of reenlistment, we assign each soldier the SRB offer that was available in the first month of their reenlistment window.[22] Despite these abstractions, our assigned SRB offers are highly predictive of the actual received bonus amount for those who take up SRB offers.[23]

2.4 Empirical Strategy & Results

The following section provides evidence on the selection on ability induced by two of the Army's lump-sum retention policies—Selective Reenlistment Bonuses, which provide cash bonuses to soldiers who stay, and early-retirement programs, which provide cash bonuses to soldiers who leave. In Appendix Section B.1, we show that the differential response of soldiers to lump-sum bonuses is the key statistic for understanding how the average ability of the military is affected by these reenlistment programs.

2.4.1 Evidence from Selective Reenlistment Bonuses (SRBs)

We begin by comparing the reenlistment decisions of soldiers according to the bonus amounts they are offered. In particular, we estimate the following equation:

$$\text{Stay}_{it} = \beta_0 + \beta_1 \text{SRB}_{it} + \beta_2 \text{SRB}_{it} * \text{AFQT}_i + \beta_3 \text{AFQT}_i + \gamma_{MOS,rank,yos} + \mu_t + \delta \mathbf{X}_{it} + \epsilon_{it}, \tag{2.4.1}$$

[22]We show, however, that our results are not sensitive to the timing assumption for the SRBs. See Appendix Tables B.11 and B.12.

[23]The coefficient of a regression of actual bonuses on SRB offers is 0.236 and is highly statistically significant ($p < 0.01$).

where Stay_{it} is an indicator for whether soldier i chooses to reenlist at time t; SRB_{it} represents a soldier's SRB offer as described above, and $AFQT_i$ is the soldier's raw AFQT score percentile. We expect β_1, which estimates the average effect of SRB offers on reenlistment, to be positive, since SRBs are designed to increase soldier retention. Our coefficient of interest is β_2, which reflects the differential responsiveness of high- and low-ability soldiers to reenlistment bonus offers.

The identification assumption underlying the estimation of β_2 is that SRBs are conditionally randomly assigned, and thus unrelated to both individual ability and non-monetary factors affecting the reenlistment decision. Since SRB offers vary by occupation, rank, year of service, and date, all of our specifications include offer-date fixed effects and MOS$\times$rank$\times$years-of-service fixed effects. We also include controls ($\mathbf{X}_{it}$) for marital status, gender, race, age, and special military skills designations. While the demographic controls are not necessary for identification, they nonetheless improve the precision of our estimates. Although we are unable to test whether SRBs are correlated with unobservable soldier characteristics, such as their taste for military service, in Columns (1) and (4) of Appendix Table B.17 we document that, conditional on occupation, tenure, and rank, SRBs are not offered to cohorts of soldiers that are higher ability. This test on observables strongly supports the identifying assumption, since the finding that SRBs are uncorrelated with our rich set of observables makes it unlikely that they are nonetheless correlated with potential unobservable characteristics (Altonji et al., 2005).

Given these controls, β_2 will be identified off of relatively high-frequency variation in SRB offers that vary across MOS, rank, and years of service within a date. While it is difficult to know precisely what drives this time-series variation, anecdotal and observational evidence suggests that variation in SRBs arises from a combination of “inside” factors—namely, the military’s operational and strategic requirements—and “outside factors”—namely, labor market conditions and other economic trends affecting civilian labor market opportunities. For example, SRB offers for Patriot missile operators (MOS 14T) appear to have been largely driven by operational requirements (i.e., air defense requirements during the first Gulf War) and large-scale changes to the Army’s overall force structure (i.e., growth of the total air defense capability). In contrast, SRB offers for infantrymen (MOS 11B)—the largest MOS in the military—appear to vary more closely with secular trends (e.g., macroeconomic conditions, post-9/11 surges in enlistments, and increased demand to support the wars in Afghanistan and Iraq). Insofar as outside economic conditions affect SRB offers, they will only threaten our identification if they vary at a high frequency and in a manner that is specific to soldiers of a particular MOS, rank, and tenure. Appendix Section B.2.2 provides case studies for the time series variation driving other specific occupations.

In Figure 2-2 we provide descriptive evidence for the effect of SRBs on selection. Both the left and right panels depict the residualized AFQT distributions for soldiers who reenlist compared to those who stay. We residualize the AFQT scores by the soldier’s occupation, rank, years of service, and the date of the reenlistment decisions—the very same variables that are used to determine a soldier’s eligibility for the military’s various incentive programs. This residualization removes, for example, any differences stemming from the fact that soldiers of higher ranks tend to have higher AFQT scores, are more likely to reenlist, and may also be eligible for different reenlistment incentives. Figure 2-2a plots the AFQT distributions for soldiers who were offered *no* SRB at the time of reenlistment, while Figure **??** plots the distributions for soldiers who were offered an SRB of at least $8,000. In both panels the stayer distribution (drawn in dashed lines) is shifted left relative to the leaver distribution (drawn in solid lines), meaning that the average ability of the soldiers who choose to reenlist is lower than those who chose to leave the military.[24] This comports with Table 2.1, which indicated that soldiers who reenlist tend to have lower AFQT scores than those who leave, but the residualized distributions plotted in Figure 2-2 show that, even within detailed occupation, rank, and tenure bins, soldiers at the higher end of the AFQT distribution are less likely to stay in the military. What is key from Figures 2-2a and **??**, however, is that the disparity between stayers and leavers is even greater for soldiers who receive a large SRB offer than it is for soldiers who receive no SRB offer. This suggests that when the SRB is higher, either lower ability soldiers are even more likely to stay, or higher ability soldiers are even more likely to leave.[25]

In Table 2.2 we formalize this descriptive result with a regression analysis. Column 1 first shows a benchmark specification relating bonus offers to average reenlistment without including the interaction between a soldier's AFQT score and their bonus offer. The coefficient on a soldier's AFQT score in Column 1 reiterates that soldiers with higher AFQT scores are less likely to reenlist—for each additional percentile point in the raw AFQT score, soldiers are 0.1141 percentage points less likely to reenlist. The Column 1 results also show that SRBs work as intended: on average, a $10,000 bonus increases soldier retention by 1.5 percentage points (2.3 percent).[26]

However, as depicted in Figure 2-2, soldiers across the ability distribution are

[24] Appendix Figure B-11 shows the raw distribution of AFQT scores by reenlistment status.

[25] Appendix Figure B-10 shows a similar pattern using a soldier's speed of promotion in their first term as their measure of quality.

[26] Note that the average non-zero SRB offer is $9,151 in 2015 dollars. About 75% of soldiers face no SRB offer in their current MOS at the beginning of their reenlistment window. This baseline estimate of the effect of SRB offers on reenlistment probabilities is similar to those reported in Greenstone et al. (2018).

not uniformly responsive to SRBs. Column 2 of Table 2.2 corresponds to our baseline specification in Equation 2.4.1, and it shows that a soldier's responsiveness to the bonus offer is decreasing in her AFQT score. The point estimate on the interaction of the SRB offer and the soldier's AFQT score is negative and statistically significant – a soldier who has an AFQT score that is 10 percentiles higher is more than 0.7 percentage points less responsive to a $10,000 SRB bonus offer. Indeed, as we show in additional results below, soldiers with AFQT scores above the 80th percentile are not at all responsive to the SRB offer.

In Columns 3 and 4 of Table 2.2 we estimate the same model with additional fixed effects that control for potential confounders. Column 3 includes nonparametric time trends for each soldier's commuting zone of record (i.e., place of residence immediately prior to initial enlistment) to control for any reenlistment differences that are correlated with the soldier's local area. The point estimates are smaller, but, as we show in Appendix Table B.7, this difference is entirely driven by changes in the sample induced by the additional fixed effects. Even so, the main pattern of lower responsiveness by higher-ability soldiers remains sizable and statistically significant. Column 4 includes nonparametric time trends for each occupation. This model identifies SRB effects from the differential time variation across ranks and tenures within an occupation and thus sweeps out anything that varies at the occupation level (e.g., changes in mortality risk, changes in outside employment opportunities for a given occupation). Once again, we find that soldiers with higher AFQT scores are less responsive to SRB offers. In Column 5 we measure a soldier's ability not by her AFQT score but by the number of months that the soldier spent below sergeant in her first term. Higher numbers imply slower promotion speeds and therefore lower military performance. Our results show that that soldiers who are promoted less quickly are more responsive to SRB offers, consistent with the AFQT findings in Columns 2-4. In Appendix Tables B.7 and B.8 we document that the Table 2.2 results are robust to various alternative specifications and sample restrictions, including using the log rather than the level of the SRB offer, restricting to the 10 largest occupations, and dropping the Iraq War "surge" years (2007-2009).

Equation 2.4.1 imposes a linear relationship between a soldier's ability and her responsiveness to bonus offers. We relax this assumption in Figure 2-3 and depict the effects of an SRB offer throughout the ability distribution. The left panel presents results using the AFQT scores, where we interact the SRB offer with dummies for each AFQT score decile. We use equally sized decile bins to reflect the soldier's relative position among those eligible to reenlist. The figure reveals that that the relationship is close to linear and decreasing throughout the distribution. Soldiers in the bottom decile are almost 5 percentage points more likely to reenlist when offered a \$10,000 SRB versus no SRB, while soldiers in the middle of the distribution are only about 1 percentage point more likely to reenlist when facing the same incentive. Beginning at the 80th percentile of this AFQT distribution, we can no longer reject the hypothesis that SRBs have no effect on reenlistment rates. We find similar results in the right panel of Figure 2-3, which uses our speed-of-promotion-based ability measure. The effect of SRBs on reenlistment is almost entirely driven by soldiers in the highest three deciles (i.e., those with the slowest promotions).

Effect Magnitudes

So far we have compared how bonuses affect the reenlistment decisions of individual soldiers at different ability levels, but an alternative method for assessing the magnitude of the selection induced by SRBs is to ask how the "marginal" reenlisters differ from the average reenlisters, and how these two groups vary with bonus offers. This approach mirrors that of Gruber et al. (1999), who analyzed the effects of legalized abortion on children's average living standards.

Figure 2-2 shows that, on average, the soldiers who choose to leave the military are of higher ability than those who choose to reenlist. Therefore, if the effect of SRBs on reenlistment were constant across the ability distribution, offering higher SRBs would *increase* the average quality of soldiers in the military. However, as we explore at length in Appendix B.4, the pattern of self-selection that we document in Table 2.2 is large enough that increasing SRB offers actually *decreases* the average quality of retained soldiers. Specifically, the estimates in Figure 2-3 imply that if the Army offered an average cohort a $10,000 SRB, it would retain an additional 195 soldiers. However, of those retained soldiers, about 150 (77 percent) would come from below the 50th percentile of the AFQT distribution, and the average AFQT percentile of those marginally retained soldiers would be 46, a full 10 points lower than the average AFQT score of the average reenlisting cohort in our data (where most soldiers receive no SRB, and the average SRB offer is just $1,891).

2.4.2 Evidence from early retirement incentives

While SRBs offer cash to those who choose to stay in the military, early retirement programs offer lump-sum payouts to those who choose to leave the military. Our analysis of the Army's early retirement programs is conceptually similar to our preceding SRB analysis, but the program details and structure of the data require a slightly modified approach. Rather than evaluating whether a soldier reenlists at the end of her spell, we evaluate whether or not she remained in the Army throughout the duration of the drawdown program eligibility window. This modification pools together soldiers who actively decide to reenlist with those who were not up for reenlistment during the program window but who nonetheless declined to take-up the early retirement program and leave the Army. We restrict our sample to spells that are active 6 months before the introduction of the early retirement program, thus counting each individual soldier only once. We make a few additional sample restrictions (described below) to isolate soldiers who are most similar to the eligible soldiers.

We first document that the program accomplished its objective of encouraging eligible soldiers to exit the military by estimating Equation (2.4.2):

$$\text{Stay}_{i,t_T} = \beta_0^T + \beta_1^T \text{ELIG}_i + \beta_2^T YOS_{i,t_0} + \gamma_{MOS,rank}^T + \delta^T \mathbf{X}_i + \epsilon_i^T, \tag{2.4.2}$$

where ELIG_i is an indicator for soldier i's eligibility for either VSI/SSB or TERA, YOS_{i,t_0} is the soldier's years of service as of the program eligibility date t_0, and Stay_{i,t_T} is an indicator for the solider remaining in the Army T months after the early retirement program went into effect (t_T). For example, the estimate for $\beta_1^{T=3}$ shows the relative probability of being in the military, by program eligibility, 3 months after the program went into effect. We re-estimate Equation (2.4.2) for a range of T values in order to analyze the effects of the early retirement program prior to and while it is in active effect. We include occupation$\times$rank fixed effects to capture any average differences in retention probabilities, and we control for the soldier's tenure since reenlistment probabilities generally decrease with tenure. We identify the effect of program eligibility by comparing soldiers of different service tenures within an occupation-by-rank bin and by comparing soldiers with the same years of service across different occupation-by-rank bins. The basic identifying assumption is that, after controlling for these observable determinants of program eligibility, eligibility for an early retirement program is correlated with neither an individual's ability level nor with the various unobservable determinants of her reenlistment decision. This assumption implies that, absent program implementation, reenlistment rates for eligible and ineligible groups would have followed parallel trends.

We present our regression results in Figure 2-4. In Panel A, we first document the effects of the retirement programs on average retention. The left graph depicts the results for the VSI/SSB programs, which offered separation incentives to mid-career soldiers. Note that the small and statistically insignificant coefficient left of the zero-month threshold shows that, prior to the official implementation of the VSI/SSB program, soldiers who were eventually eligible for the program had the same probability of staying in the military as those who would never be eligible, validating the primary parallel trends assumption underlying this specification. However, once the program comes into effect, eligible soldiers are more likely to leave the military, and by the time the VSI/SSB program expires, eligible soldiers were almost 15 percentage points less likely to remain in the military compared to ineligible soldiers. The right graph in Panel A depicts a similar analysis for TERA (which affected late-career soldiers). While the results are noisier because the program was significantly smaller, the overall pattern is similar—retention rates for eligible and ineligible soldiers moved in parallel prior to the program, but after implementation, TERA induced eligible soldiers to retire at higher rates.

In Panel B of Figure 2-4 we present the retirement program effects by ability levels (specifically, upper and lower AFQT score terciles).[27] The left panel depicts

[27] The estimates from these two groups were jointly estimated in a single regression, with sol-

the results for the VSI/SSB program. As before, there are no pre-program differences in reenlistment probabilities for each ability group, and both groups are more likely to leave the Army when offered early retirement. However, higher ability soldiers responded less to the early retirement offer than lower ability soldiers, as demonstrated by the coefficients for the bottom-tercile soldiers lying below the coefficients for top-tercile soldiers at all times after program implementation. The right panel documents similar results for the TERA program. Soldiers with lower AFQT scores are more responsive to the program than soldiers with comparatively higher scores..[28] In Appendix Figure B-12, we show that patterns are similar when we split not by AFQT score but instead by soldiers' speed of promotion in their first term. Appendix Tables B.13 and B.14 provides regression estimates from a version of Equation (2.4.2) where VSI/SSB or TERA program eligibility is interacted with a soldier's ability, further documenting that high ability soldiers are less responsive to these programs.

The Figure 2-4 estimates imply that a VSI offer to 1,000 soldiers would induce an additional 90 soldiers to retire. Almost two-thirds of those soldiers would be below the median AFQT score. Thus, in contrast to the SRB result, the self-selection induced by this policy is large enough to *increase* the average quality of retained soldiers, since the lowest ability soldiers disproportionately take up the cash offer to leave the military. See Appendix B.4 for a more formal analysis of the effect of both retirement programs on average soldier ability levels.

diers belonging to the middle AFQT tercile as the omitted category.

[28]There are several reasons why the results would be stronger for the VSI/SSB program than the TERA program. As shown in Table B.5, the VSI/SSB program affected more soldiers. Additionally, the VSI/SSB program ran for longer than the TERA program, perhaps giving soldiers more time to react. However, the programs also differed in the type of benefit—soldiers eligible for the VSI/SSB program had the option to get a large lump-sum payment while soldiers in TERA were only entitled to the retirement annuity. Indeed, most soldiers who took up the VSI/SSB program chose the lump-sum payment rather than the annuity.

2.5 Explanatory Mechanisms

The previous sections document that the sensitivity of reenlistment decisions to near-term cash incentives is decreasing in individual soldier ability. These results are perhaps surprising. First, this selection pattern would seem to work against the positive effect of base wages on quality of civil service recruits documented throughout much of the literature (e.g., Dal Bo et al. 2013). Furthermore, in Appendix B.1, we demonstrate that this pattern of selection is not consistent with a simple workhorse model of selection in which soldiers differ only along one dimension—their ability—and in which the wage profile in the military is less sensitive to ability than in the private sector.

In this section, we empirically explore the degree to which the selection patterns we document are driven by the specific lump-sum structure of the retention payments, which, unlike changes in base wages, alter both the level and timing of compensation. First, we assess whether low-ability soldiers are more credit constrained and thus value the liquidity provided by the lump-sum payments more than their higher-ability peers. Second, we explore whether higher ability soldiers are more patient (as measured by lower personal discount rates) and consequently less responsive to promises of immediate lump-sum transfers. Our results suggest that, while both access to credit and discount factors are associated with reenlistment, neither is likely to be driving the differential responsiveness of high- and low-ability soldiers to cash incentives. Rather, we hypothesize that the selection patterns we document could be driven by features of the programs other than the timing of their payments. We show formally in Appendix B.1 that this selection on ability may result from an idiosyncratic "taste for service" that is distributed such that high-ability soldiers tend to be inframarginal relative to their low-ability peers. It could also be that low-ability soldiers have lower expected permanent incomes, and thus any fixed nominal payment represents a larger relative income shock for them. Finally, since individuals may utilize hyperbolic (or quasi-hyperbolic) discounting (Laibson (1997)), and since those with lower cognitive abilities may be more likely to do so (Benjamin et al. (2013), Shamosh and Gray (2008), Parker and Fischhoff (2005)), our differential responsiveness to near term incentives by ability may reflect these alternative time preferences. Our data are unfortunately not well suited to formally test either of these possibilities, so in the next sections, we focus on empirical evidence exploring whether differences in either credit constraints or personal discount factors are able to explain the main result.

2.5.1 Credit Constraints

Low-ability soldiers may exhibit differential sensitivity to cash incentives because they are more credit constrained than their high-ability peers. Given that family resources account for a large share of the variation in AFQT scores (Neal and Johnson, 1996) and that AFQT scores are themselves strongly correlated with future labor market outcomes (Heckman et al., 2006), access to credit—which is a function of both current assets and future income—is likely to be correlated with cognitive ability. These differences in liquidity by ability may cause lower ability soldiers to respond to near term incentives for several reasons. First, they may place a higher value on cash for precautionary savings or to finance a larger household expenditure. In the case of the early retirement programs, more credit-constrained households may also value the liquidity as it enables them to prolong and optimize their job search in the civilian labor market.[29]

We explore the degree to which differences in liquidity across the ability distribution explain our main results by adding additional controls to our baseline SRB regressions from Section 2.4.1. If differences in access to credit are driving soldiers' differential responsiveness to SRBs by ability level, then we anticipate that directly controlling for measures of credit constraints in our baseline regressions will correct for omitted variable bias and consequently eliminate the interaction between SRBs and our measures of ability. In order to control for access to credit, we match soldiers to their individual credit scores and balances, which were obtained from one of the major credit reporting agencies for soldiers who were eligible for reenlistment at any point between April 2007 and March 2015.[30] In our data, we confirm that there is a positive correlation between ability and credit scores.[31]

The first three columns of Table 2.3 present regression coefficients from equation (2.4.1) after controlling for soldiers' credit scores. Column 1 replicates our main SRB result in the subsample of soldiers with non-missing credit scores. Column 2 shows both that the coefficient on the SRB $*$ AFQT interaction is robust to simply controlling for credit scores and that soldiers with higher credit scores are less likely to reenlist on average. However, the estimates in Column 3 show that soldiers with more credit are less responsive to SRBs, as theory would predict, but that the coefficient on SRB $*$ AFQT remains unchanged, suggesting that credit

[29] A few recent papers have addressed the important role that worker liquidity constraints can play in labor markets. For example, Giannetti (2011) find that liquidity constraints can also affect occupational choice; individuals with a higher probability of facing liquidity constraints are less likely to be self-employed, and they are more likely to be employed in the public vs. private sector (see, also, Bianchi and Bobba 2013).

[30] Our match rate is high (nearly 90%) for our main sample of reenlistment-eligible soldiers.

[31] The correlation coefficients between credit score and AFQT and months spent below sergeant are 0.21 and -0.14, respectively.

constraints are not driving our main finding. In Appendix Table B.16 we show that these patterns are robust to alternative proxies for credit constraints, including an indicator for whether a soldier has a credit score of at least 680 (which most lenders consider to be "prime" credit).

2.5.2 Personal Discount Factors

An alternative explanation for our patterns is that high- and low-ability soldiers' differential sensitivity to cash incentives may reflect behavioral differences in decision making between high- and low-ability individuals. Previous research has demonstrated that cognitive ability is strongly correlated with a variety of decision-making characteristics, such as greater patience and higher risk tolerance (see, e.g., Frederick (2005) and Benjamin et al. (2013)). Importantly, similar relationships have been documented previously for the military. Warner and Pleeter (2001) estimate servicemembers' personal discount rates (PDRs) using take-up of early 1990s military drawdown programs (namely, VSI and SSB, discussed above). Their estimates suggest average discount rates as high as 17%, and they document higher rates for enlisted members, less educated members, and those with lower AFQT scores. Simon et al. (2015) estimate PDRs using more recent military retirement programs and find smaller PDRs of around 7% for enlisted soldiers and 2-4% for officers. Both studies document a negative correlation between AFQT scores and PDRs. We do not attempt to replicate these analyses, but we do confirm similar patterns in our sample, which is restricted to the Army enlisted force given our desire to exploit variation in SRBs (which were not offered to officers) and early retirement incentives. In Appendix Table B.15 we provide OLS estimates from regressions of VSI/SSB take-up on soldier ability and show that for each 10 additional points of AFQT score a soldier is approximately 2% less likely to select the SSB lump-sum payment over the VSI annuity, which has a higher net present value for standard discount rates. The relationship is robust, albeit smaller, after including various demographic controls and MOS and rank fixed effects.

As with credit constraints, we explore the role that differences in personal discount rates play in driving the heterogeneous responses to SRB offers by adding additional proxies for discount factors to our baseline SRB regressions. We proxy for soldiers' discount factors with two additional variables, both of which are directly observable within the Army's personnel data. First, we generate an indicator variable for whether, at the time of initial enlistment, soldiers made an upfront investment (known as the GI Bill "buy-up")in order to enlarge their future GI Bill educational benefits.[32] Second, we construct a variable measuring soldiers' participation in the Thrift Savings Plan (TSP), which is an optional 401(k)-style retirement savings plan offered to members of the military since 2001.[33] We proxy for a soldier's relative patience with her TSP contributions as a % share of her total base pay over the course of an enlistment spell. Both of these proxies capture the degree to which soldiers choose to transfer resources from the present to the future.

Columns 4-6 and 7-9 of Table 2.3 present regression coefficients from equation

[32]The "GI Bill" is a general term used to describe a series of federal government programs that have funded civilian higher education for returning military veterans since 1944. We focus on the "Montgomery GI Bill" (MGIB), which was passed into law in 1984 and remains in place today. Although most present-day veterans now use the "Post-9/11 GI Bill", the MGIB remained the dominant source of veterans' education benefits until at least 2008. We focus in particular on the MGIB's $600 "buy-up" option, which promises soldiers a larger future GI Bill benefit in exchange for a $600 deduction from their first year's salary. Under 2016 rates, veterans using the MGIB are eligible to receive a baseline monthly tuition benefit of $1,857. Soldiers who elected to participate in the buy-up are eligible to receive an additional $150 per month.

[33]Additional details on both the MGIB buy-up and the TSP are provided in Appendix Section B.2.1.

(2.4.1) after controlling for participation in the GI Bill "buy-up," and TSP contributions, respectively. As with credit scores, soldier ability is indeed positively correlated with our proxies for soldiers' relative patience.[34] However, Table 2.3 suggests that the correlation between discount factors and soldier ability cannot fully explain our main result. Columns 4 and 7 replicate our main SRB result in the subsamples of soldiers with non-missing GI Bill and TSP data, respectively. Columns 5 and 8 show that the coefficient on the SRB $*$ AFQT interaction is robust to controlling for each of these variables individually, and columns 6 and 9 show that the coefficient on SRB $*$ AFQT is also robust to including the interaction between SRBs and either of our two discount factor proxies. Specifically, we find that soldiers who are more patient (as reflected by our proxy measures) are indeed less responsive to SRBs. However, the coefficient on SRB $*$ AFQT remains unchanged across the columns, suggesting that differences in discount factors across ability levels are also unlikely to be driving our main finding. In Appendix Table B.16 we show that these patterns are largely robust to alternative specifications, including when we proxy for discount factors not with participation in the MGIB buy-up option but with enrollment in the baseline MGIB,[35] and when we proxy not with the share of a soldier's salary that she contributes to the TSP but with whether she makes any TSP contribution whatsoever.

[34]The correlations between buy-up participation and AFQT/months-below-sergeant are 0.07 and -0.04, respectively; and, the correlations between TSP contributions and AFQT/months-below-sergeant are 0.09 and -0.03, respectively.

[35]Because even the baseline MGIB benefits require an initial contribution of $1,200 at the time of enlistment, enrollment in the MGIB is itself a reasonable proxy for a soldier's relative patience. However, enrollment in standard MGIB is far more common than participation in the buy-up, and it may be the default for most soldiers. Whereas more than 93% of all eligible soldiers elect to enroll in the standard MGIB, just 3% of those soldiers make the additional $600 contribution necessary to participate in the buy-up.

2.6 Conclusion

This paper explores the nature of selection in public sector employee retention with evidence from the U.S. Army. Our paper extends the literature on worker sorting between the public and private sectors. Relative to the existing research, which has tended to emphasize differences in the levels of compensation at the initial entry margin, our paper brings new attention to the retention margin, and in particular to the structure of commonly used retention incentives. Using variation in reenlistment bonuses and early retirement programs, we have shown that low-ability soldiers are more sensitive to immediate lump-sum transfers than their higher-ability peers. On the margin, lump-sum bonus offers induce lower-ability soldiers to reenlist, while early retirement programs induce lower-ability soldiers to leave the Army. We provide suggestive evidence that these patterns do not arise from differences in either credit constraints or discount factors across the ability distribution. We nonetheless estimate that these effects are large enough to affect the average ability level of the military. Insights from this project are relevant not only to the U.S. military but also to the many other public sector organizations that lack the private sector's ability to target incentives to high-performing workers but are nonetheless tasked with recruiting and retaining a high-quality workforce.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2 Bibliography

Aiken, Linda H., Sean P. Clarke, Robyn B. Cheung, Douglas M. Sloane, and Jeffrey H. Silber (2003) "Educational Levels of Hospital Nurses and Surgical Patient Mortality," *JAMA: The Journal of the American Medical Association*, Vol. 290, pp. 1617–1623.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005) "Selection on Observed and Unobserved Variables: Assessing the Effectiveness of Catholic Schools," *Journal of Political Economy*, Vol. 113, pp. 151–184.

Asch, Beth J., Paul Heaton, James Hosek, Paco Martorell, Curtis Simon, and John T. Warner (2010) "Cash Incentives and Military Enlistment, Attrition, and Reenlistment,"Technical report, The RAND Corporation.

Bacolod, Marigee (2007) "Do Alternative Opportunities Matter? The Role of Female Labor Markets in the Decline of Teacher Quality, 1960-1990," *Review of Economics and Statistics*, Vol. 89, pp. 737–751.

Benjamin, Daniel J., Sebastian A. Brown, and Jesse M. Shapiro (2013) "Who is 'Behavioral'? Cognitive Ability and Anomalous Preferences," *Journal of the European Economic Association*, Vol. 11, pp. 1231–1255.

Besley, Timothy, Jose G. Montalvo, and Marta Reynal-Querol (2011) “Do Educated Leaders Matter?” *The Economic Journal*, Vol. 121, pp. F205–227.

Bianchi, Milo and Matteo Bobba (2013) “Liquidity, Risk, and Occupational Choices,” *Review of Economic Studies*, Vol. 80, pp. 491–511.

Borjas, George J. (2002) “The Wage Structure and the Sorting of Workers into the Public Sector,” *NBER Working Paper Series*, Vol. No. 9313.

Brown, Charles (1985) “Military Enlistments: What Can We Learn from Geographic Variation?” *American Economic Review*, Vol. 75, pp. 228–234.

Brown, Kristine M. (2013) “The link between pensions and retirement timing: Lessons from California teachers,” *Journal of Public Economics*, Vol. 98, pp. 1–14.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff (2014) “Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood,” *American Economic Review*, Vol. 104, pp. 2633–2679.

Condly, Steven, Arthur Coumbe, and William Skimmyhorn (2017) *Still Soldiers and Scholars*: Strategic Studies Institute.

Dal Bo, Ernesto, Frederico Finan, and Martin A. Rossi (2013) “Strengthening State Capabilities: The Role of Financial Incentives in the Call to Public Service,” *The Quarterly Journal of Economics*, Vol. 128, pp. 1169–1218.

Daula, Thomas and Robert Moffitt (1995) “Estimating Dynamic Models of Quit Behavior: The Case of Military Reenlistment,” *Journal of Labor Economics*, Vol. 13, pp. 499–523.

Deserranno, Erika (2019) "Financial Incentives as Signals: Experimental Evidence from the Recruitment of Village Promoters in Uganda," *American Economic Journal: Applied Economics*, Vol. 11, pp. 277–317.

Fernandez, Judth C. (1992) "Soldier Quality and Job Performance in Team Tasks," *Social Science Quarterly*, Vol. 73, pp. 253–265.

Finan, Frederico (2017) "The Personnel Economics of the Developing State," in Abhijit Banerjee and Esther Duflo eds. *Handbook of Field Experiments*, Vol. II: North Holland.

Flyer, Eli S. and Richard S. Elster (1983) "First Term Attrition Among Non-Prior Service Enlisted Personnel: Loss Probabilities Based on Selected Entry Factors,"Technical report, Navy Personnel Research and Development Center.

Frederick, Shane (2005) "Cognitive Reflection and Decision Making," *Journal of Economic Perspectives*, Vol. 19, pp. 25–42.

Friedman, Milton (1967) "Why Not a Voluntary Army," in Sol Tax ed. *The Draft: A Handbook of Fracts and Alternatives*: University of Chicago Press.

Gelber, Alexander (2007) "The Supply of Military Enlistments," *Unpublished manuscript*.

Giannetti, Mariassunta (2011) "Liquidity Constraints and Occupational Choice," *Finance Research Letters*, Vol. 8, pp. 37–44.

Gotz, Glenn A. and John J. McCall (1984) "A Dynamic Retention Model for Air Force Officers,"Technical Report R-3028-AF, The RAND Corporation.

Greenstone, Michael, Kyle Greenberg, Stephen P. Ryan, and Michael Yankovich (2018) "The Value of a Statistical Life: Evidence from Military Retention Incentives and Occupation-Specific Mortality Hazards," *Unpublished working paper*.

Grilliches, Zvi and William M. Mason (1972) "Education, Income, and Ability," *Journal of Political Economy*, Vol. 80, pp. S74–S103.

Gruber, Jonathan, Phillip Levine, and Douglas Staiger (1999) "Abortion Legalization and Child Living Cicumstances: Who is the "Marginal Child"?" *Quarterly Journal of Economics*, Vol. 114, pp. 263–291.

Heckman, James J., Jora Stixrud, and Sergio Urzua (2006) "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior," *Journal of Labor Economics*, Vol. 24, pp. 411–482.

Horowitz, Stanley A. and Allan Sherman (1980) "A Direct Measure of the Relationship between Human Capital and Productivity," *The Journal of Human Resources*, Vol. 15, pp. 67–76.

Kamarck, Kristy N. (2018) "Military Retirement: Background and Recent Developments,"Technical report, Congressional Research Service.

Kane, Tim (2012) *Bleeding Talent: How the U.S. Military Mismanages Great Leaders and Why It's Time for a Revolution*: Palgrave Macmillan US.

Katz, Lawrence F. and Alan B. Krueger (1991) "Changes in the Structure of Wages in the Public and Private Sectors," *NBER Working Paper Series*, Vol. No. 3667.

Kavanagh, Jennifer (2005) "Determinants of Productivity for Military Personnel: A Review of the Findings on the Contribution of Experience, Training, and Aptitude to Military Performance,"Technical Report TR-193-OSD, The RAND Corporation.

Korb, Lawrence J. and David R. Segal (2011) "Manning & Financing the Twenty-First- Century All-Volunteer Force," *Dædalus, the Journal of the American Academy of Arts & Sciences*, Vol. 104, pp. 75–87.

Laibson, David (1997) "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, Vol. 112, pp. 443–477.

Mayberry, Paul W. and Neil B. Carey (1997) "The Effect of Aptitude and Experience on Mechanical Job Performance," *Educational and Psychological Measurement*, Vol. 57, pp. 131–149.

Neal, Derek A. and William R. Johnson (1996) "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, Vol. 104, pp. 869–895.

Nickell, Stephen and Glenda Quintini (2002) "The consequences of the decline in public sector pay in Britain: a little bit of evidence," *The Economic Journal*, Vol. 112, pp. F107–F118.

Orvis, Bruce R., Michael T. Childress, and J. Michael Polich (1992) "Effect of Personnel Quality on the Performance of Patriot Air Defense System Operators,"Technical Report R-3901-A, The RAND Corporation.

Parker, Andrew and Baruch Fischhoff (2005) "Decision-Making Competence: External Validation Through an Individual-Differences Approach," *Journal of Behavioral Decision Making*, Vol. 18, pp. 1 – 27.

Poterba, James M., Joshua Rauh, Steven Venti, and David A. Wise (2007) "Defined Contribution Plans, Defined Benefit Plans, and the Accumulation of Retirement Wealth," *Journal of Public Economics*, Vol. 91, pp. 2062–2086.

Rydberg, Jason and William Terrill (2010) "The Effect of Higher Education on Police Behavior," *Police Quarterly*, Vol. 13, pp. 92–120.

Shamosh, Noah A. and Jeremy R. Gray (2008) "Delay discounting and intelligence: A meta-analysis," *Intelligence*, Vol. 36, pp. 289 – 305.

Simon, Curtis J. and John T. Warner (2007) "Managing the All-Vounteer Force in a Time of War," *The Economics of Peace and Security Journal*, Vol. 2, pp. 20–29.

Simon, Curtis J., John T. Warner, and Saul Pleeter (2015) "Discounting, Cognition, and Financial Awareness: New Evidence from a Change in the Military Retirement System," *Economic Inquiry*, Vol. 53, pp. 318–334.

Teachout, Mark S. and Martin W. Pellum (1991) "Air Force Research to Link Standards for Enlistment to On-the-Job Performance,"Technical report, Air Force Human Resources Laboratory, Brooks AFB TX.

Wallace, Roy A., Michael J. Colarusso, Andrew O. Hall, David S. Lyle, and Michael S. Walker (2015) "Paid to Perform: Aligning Total MIlitary Compensation with Talent Management," *Strategic Studies Institute Officer Corps Strategy Series*.

Wardynski, Casey, David S. Lyle, and Michael J. Colarusso (2010) "Towards a U.S. Army Officer Corps Strategy for Success: Retaining Talent," *Strategic Studies Institute Officer Corps Strategy Series*.

Warner, John T. and Saul Pleeter (2001) "The Personal Discount Rate: Evidence from Military Downsizing Programs," *American Economic Review*, Vol. 91, pp. 33–53.

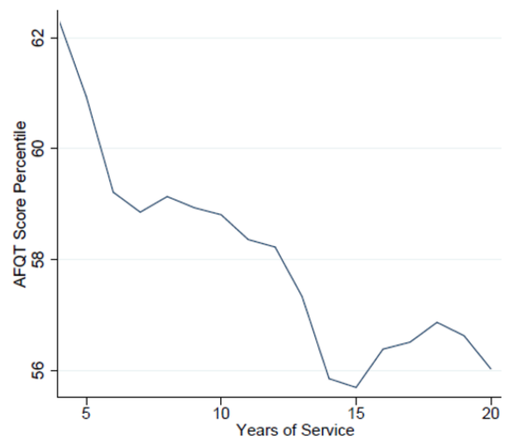
Warner, John T., Curtis Simon, and Deborah Payne (2003) "The Military Recruiting Productivity Slowdown: The Roles of Resources, Opportunity Cost and the Tastes of Youth," *Defence and Peace Economics*, Vol. 14, pp. 329–342.

Wigdor, Alexandra K. and Bert F. Green eds. (1991) *Performance Assessment for the Workplace*, Vol. 1: National Academies Press, National Research Council.

Winkler, John D., Judth C. Fernandez, and J. Michael Polich (1992) "Effect of Aptitude on the Performance of Army Communications Operators,"Technical Report R-4143-A, The RAND Corporation.

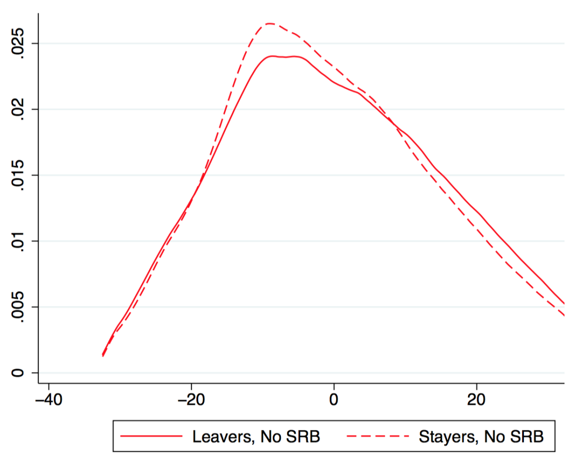
Chapter 2 Tables and Figures

Figure 2-1: Average AFQT Score Percentile by Tenure with the Army

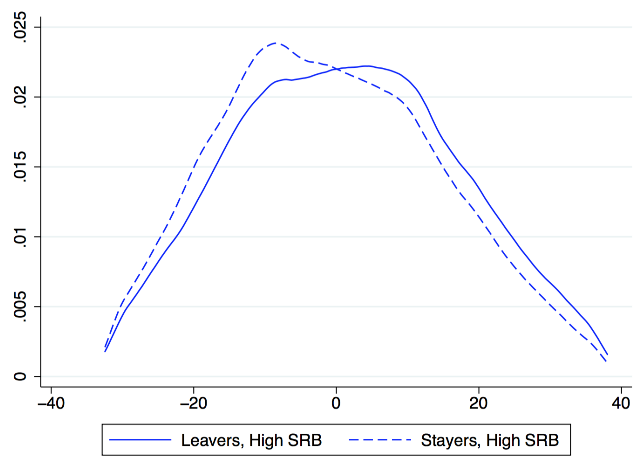


Notes: The figure plots the average AFQT Score Percentile of enlisted soldiers in the Army from 1992-2016, excluding soldiers who are currently serving. Years of service is defined as a soldier's total tenure with any branch of the military. Years of service is measured at the time of separation, or, for soldiers still serving, in the current period.

Figure 2-2: The Distribution of AFQT Scores for Soldiers, Split by Reenlistment Decisions and SRB Offers



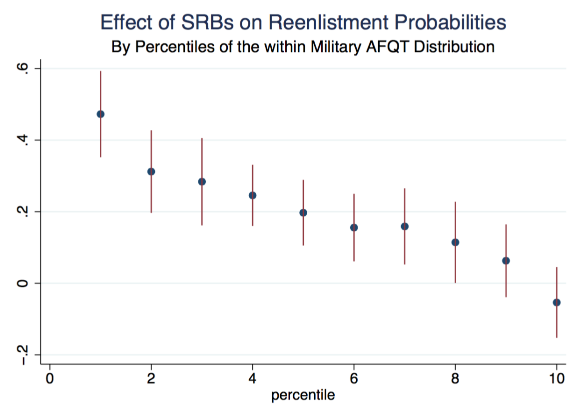
(a) Soldiers without an SRB offer

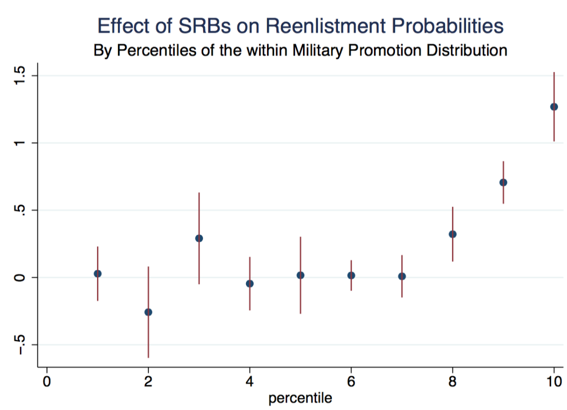


(b) Soldiers offered an SRB of at least $8,000

Notes: Figures 2-2a and **??** plot the residuals of a regression of AFQT score on MOS*rank*YOS dummies as well as date dummies. The sample includes only those soldiers who have a choice to reenlist. The left panel plots the distributions for the set of soldiers who do not have a SRB available at the start of their reenlistment window. The right panel shows the distributions for the set of soldiers who have an offered SRB of at least $8,000. The left figure includes 1.7 million observations (75% of the sample) while the right panel includes 300,000 observations (13% of the sample). Each distribution is truncated at the top and bottom 1%.

Figure 2-3: The Effect of Selective Reenlistment Bonuses on Soldier Retention by Soldier Quality: Nonlinear Specifications



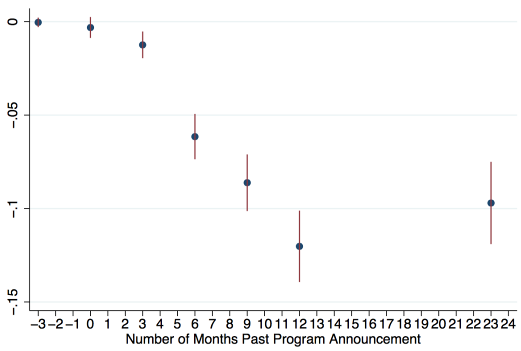


Notes: The left panel of this figure plots the coefficient estimates on the interaction of SRB offers and a dummy for each percentile of the AFQT score distribution. These are 10 equallly sized percentile bins, and correspond to the distribution of those soldiers who are eligible to reenlist, not the overall distribtuion of AFQT percentiles. The right panel plots similar regressions using the distribution of soldier's promotion speeds instead of AFQT scores. The promotion speed is measured by the number of months the soldier spend at a rank below a sergeant. In both panels, the red bars show 95 percent confidence intervals, clustering the standard erros at the MOS*rank*yos level. Reenlistment probabilities (the y-axis) are scaled by 100 and SRB values are in terms of thousands of U.S. dollars.

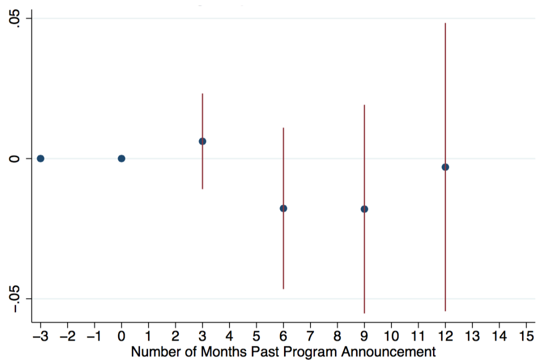
Figure 2-4: The Effect of Early Retirement Programs on Soldier Selection

Panel A: The Effect of Early Retirement Programs on Soldier Retention

(a) VSI/SSB

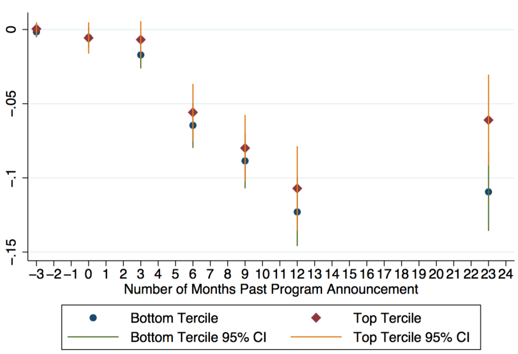


(b) TERA

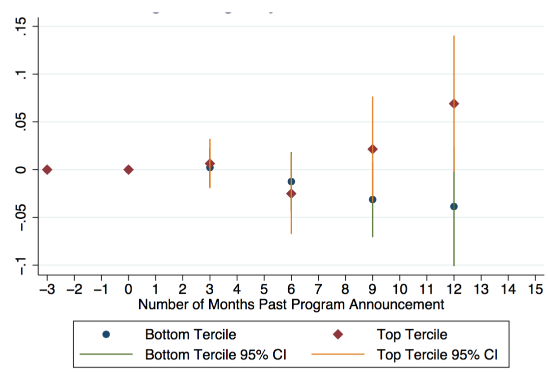


Panel B: The Effect of Early Retirement Programs on Retention by AFQT Scores

(c) VSI/SSB



(d) TERA



Notes: The left graph of each panel (VSI/SSB) shows the probability of remaining in the Army for each month relative to August 1, 1993, the start of the VSI/SSB program and includes soldiers with at least 6 years of experience. In Panel A, blue dots show the coefficient estimate on program eligibility from separate regressions on the probability of remaining in the military in period t. In Panel B, we split soldiers into terciles of the AFQT score distribution. In each time period, we run a regression of program eligibility interacted with the soldier's AFQT tercile on the probability of remaining in the military in period t. The right figures shows similar specifications, but defines the sample and the time period relative to August 31, 1994, the day the TERA program was introduced and includes only soldiers in the affected ranks and occupations, who have tenures that put them within 1 year of being eligible. In panel B, blue circles plot the coefficient on program eligibility interacted with the bottom tercile, and red diamonds plot the coefficient on program eligibility interacted with the top tercile. The middle tercile was also included in the regression but is not plotted here. Across all figures, regressions also includes occupation and rank fixed effects, a control for the soldier's tenure as of the program start date, dummies for the soldier's AFQT score tercile, and demographic controls (age, marital status, gender and race). Lines show the 95% confidence intervals, with standard errors clustered at the occupation*rank*year of service bin.

Table 2.1: Summary Statistics

	(1) Full Sample	(2) Soldiers with Reenlistment Choice	(3) Spells ending in exit	(4) Spells ending in Reenlistment
Fraction Male	0.85	0.85	0.85	0.86
Age	28.37	29.02	29.71	28.66
Years of Service	6.33	6.98	7.96	6.46
Fraction Married	0.57	0.60	0.52	0.64
AFQT Percentile	57.94	58.25	59.68	57.48
Months as Sergeant in First Term	2.51	2.99	1.95	3.55
Number of Soldiers	1,626,298	1,180,179	726,930	715,153
Number of Spells	2,765,755	2,102,206	734,972	1,367,234

Notes: Sample in Column 1 includes the enlistment spells for all enlisted soldiers from 1992-2016. Column 2 restricts to the enlistment spells at the end of which soliders have the option to reenlist. Column 3 includes the set of spells at the end of which the soldier decides to exit the military. Column 4 includes the set of spells that are followed by another term in the Army. Years of service are defined as of the end of the spell, and AFQT scores are measured at the time of entrance into the Army.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.2: The Effect of SRBs on Soldier Retention, by Soldier Ability

	(1)	(2)	(3)	(4)	(5)
Ability Measure:	AFQT Score				Months Below Sergeant in First Term
SRB	0.158***	0.615***	0.327***	0.359***	-0.607***
	(0.042)	(0.078)	(0.066)	(0.085)	(0.108)
SRB * AFQT		-0.710***	-0.281***	-0.745***	0.015***
		(0.116)	(0.102)	(0.117)	(0.002)
AFQT	-11.411***	-9.347***	-14.312***	-9.127***	0.309***
	(0.873)	(0.868)	(0.648)	(0.914)	(0.024)
R^2	0.157	0.157	0.189	0.195	0.171
Year * Month FE	Y	Y	N	N	Y
Year * Month * CZ FE	N	N	Y	N	N
Year * Month * MOS FE	N	N	N	Y	N
MOS * Rank * YOS FE	Y	Y	Y	Y	Y
Demographic Controls	Y	Y	Y	Y	Y
Avg. Reenlistment Rate	65.10	65.10	66.72	65.13	66.30
Avg. SRB	2.89	2.89	3.26	2.9	3.02
Observations	1,761,615	1,761,615	1,422,783	1,757,584	1,708,425

Standard errors are reported in parentheses. They are two-way clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. SRBs are in $1000s of 2015 dollars. Demographic controls include gender, age, marital status, race, and special skill dummies. "Ability" is defined as AFQT score for columns (1)-(4) and months below Sergeant for column (5). AFQT is on a scale from 0-1. See Table C1, Column 1 for evidence that the average SRB in a given period is conditionally uncorrelated with the average ability of the eligible soldiers.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.3: The Effect of SRBs on Soldier Retention, by AFQT
Including Credit Score, Montgomery GI Bill, and Thrift Saving Program Interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Credit Score			GI Bill "Buy-up"			TSP Contribution %	
SRB	0.477***	0.472***	1.488***	0.198*	0.169	0.163	0.365***	0.363***	0.367***
	(0.145)	(0.145)	(0.221)	(0.114)	(0.116)	(0.116)	(0.093)	(0.094)	(0.093)
SRB * AFQT	-0.847***	-0.839***	-0.707***	-0.483***	-0.427***	-0.368***	-0.708***	-0.706***	-0.692***
	(0.188)	(0.186)	(0.179)	(0.122)	(0.123)	(0.123)	(0.132)	(0.133)	(0.133)
AFQT	-9.652***	-8.511***	-3.126	-17.491***	-16.735***	-16.876***	-10.171***	-10.806***	-10.212***
	(0.955)	(0.884)	(3.585)	(0.813)	(0.814)	(0.815)	(0.907)	(0.905)	(0.924)
Mechanism Var.		-0.248***	-0.161***		-12.662***	-9.727***		29.542***	56.600***
		(0.020)	(0.041)		(0.797)	(1.388)		(1.303)	(3.769)
SRB * Mechanism Var.			-0.017***			-0.584***			-0.656***
			(0.002)			(0.107)			(0.167)
AFQT * Mechanism Var.			-0.088			-0.608			-36.592***
			(0.056)			(1.787)			(4.937)
R^2	0.207	0.209	0.209	0.221	0.223	0.223	0.232	0.232	0.232
Year * Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
MOS * Rank * YOS FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year * Month * MOS FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Demographic Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Avg. Reenlistment Rate	68.28	68.42	68.42	52.05	52.05	52.05	64.62	64.62	64.62
Avg. SRB	2.06	2.06	2.06	3.33	3.33	3.33	2.70	2.70	2.70
Observations	606,350	600,688	600,688	1,000,035	1,000,035	1,000,035	1,168,621	1,168,621	1,168,621

Standard errors are reported in parentheses. They are two-way clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. Samples for columns (1)-(3) are further restricted to soldiers with non-missing credit scores. Samples for columns (4)-(6) are restricted to soldiers with non-missing GI Bill participation data. Samples for columns (7)-(9) are restricted to soldiers with non-missing TSP contribution data. SRBs are in $1000s of 2015 dollars. Demographic controls include gender, age, marital status, race, and special skill dummies. AFQT and TSP contribution % are on scales from 0-1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Do Military Housing Allowances Inflate Local House Prices?

Joint with Paul Goldsmith-Pinkham

3.1 Introduction

An open question in public economics is *how* governments should provide the public goods they do. For any given government program, policymakers might choose between in-kind transfers, cash transfers, or alternative arrangements like public-private partnerships. Economists have traditionally favored cash transfers—and among cash transfers, lump-sum payments—since they minimize distortions. Less attention has been given to the potential price effects (or "pecuniary externalities") associated with in-kind cash transfers. Theory would suggest that in-kind transfers of a merit good (for example, health care, education, or housing) may tend to reduce prices by expanding supply of the good. Conversely, cash transfers may *increase* prices by boosting demand for the good.

This paper analyzes the pecuniary effects of military housing allowances, which represent a large unconditional cash transfer to certain military households. We find evidence of sizeable price effects, particularly in areas with relatively dense military populations, and particularly in areas where housing supply is especially inelastic. Methodological challenges advise that our results should be interpreted with caution. Nonetheless, our results suggest that policymakers should give more consideration to in-kind transfer schemes, especially where expenditures on a particular merit good represent a large share of household spending, and especially where market conditions allow sellers to raise prices without sacrificing profits.

The theoretical literature has put forward a number of potential justifications for in-kind transfer programs. Common arguments for in-kind transfers include paternalism (e.g., Besley 1988), interdependent preferences (Daly and Giertz 1972; Garfinkel 1973), and targeting (Nichols and Zeckhauser, 1982). Pecuniary externalities offer another possible rationale for in-kind transfers. If in-kind transfers stand to reduce the price of a good by boosting supply, then in-kind transfers can be especially effective at increasing consumption of a merit good. Moreover, subsidies or lump-sum cash transfers could even create an opposite pecuniary externality. If an influx of cash generates additional demand for the merit good, then the price of the good may rise, causing non-treated individuals to substitute away from the merit good. Cash transfers are more likely to generate pecuniary externalities when the cash must be spent on the merit good (for example, Section 8 vouchers, which must be spent on rental housing, or SNAP benefits, which must be spent on eligible food purchases). Nonetheless, even unconditional lump-sum transfers can generate price effects as long as the income effect is sufficiently large.

This paper studies the potential for pecuniary externalities in the government's provision of military housing benefits. In 2019 the Department of Defense (DoD) proposed spending $21 billion in cash housing cost allowances for approximately 1 million U.S. service members.[1] By contrast, the same proposal called for less than $1.6 billion in direct provision of family housing.[2] The Basic Allowance for Housing (BAH) is a lump-sum transfer paid to active duty U.S. service members to compensate them for local area housing costs when on-base housing is not provided. We can think of military housing allowances and on-base housing as just two among a long list of alternative options for delivering adequate housing to military personnel. But how does the federal government choose between alternative policy instruments, and what are the relevant tradeoffs?

In fact, there is some anecdotal evidence that the U.S. military housing policy causes pecuniary externalities. The debate over military housing policy is especially pronounced in places like Hawaii and Southern California, where the cost-of-living is high and military bases loom large. With more than 100,000 active duty service members and dependents— almost 10% of the state's population— Hawaii has the highest military density of any state in the nation. Meanwhile, Hawaii's fair market rent for a two-bedroom apartment exceeds the national average by 67%. An abundance of media reports claim that the link between military presence and the high cost of housing is not accidental.[3]

State and local officials are not insensitive to the potential externalities of mil-

---

[1]*See* DoD press release from Dec. 14, 2018: `https://dod.defense.gov/News/News-Releases/News-Release-View/Article/1714524/dod-releases-2019-basic-allowance-for-housing-rates/fbclid/IwAR205uOAWOVbxgt9qgLXBwg4GROTa3Igkhqmzbk9sGBKUl3xkKBFXbjUuMU/`.

[2]*See* DoD's 2019 budget proposal: `https://comptroller.defense.gov/Portals/45/Documents/defbudget/fy2019/FY2019_Financial_Summary_Tables.pdf`.

[3]*See, e.g.*, *Military housing allowance and the rental market*, Hawaii News Now (July 31, 2015).

itary housing policy on civilian communities. A 2014 report commissioned by the City and County of Honolulu (Cassiday, 2014) names "military absorption of local rental stock" as the first among several factors limiting the availability of affordable housing, and another 2014 report by the Hawaii Department of Business, Economic Development, and Tourism cites Hawaii's military presence as one of the three primary factors driving excess housing demand in the state. For its part, in October 2015 the U.S. Department of Housing and Urban Development published a report on the Military Housing Privitization Initiative, which is part of an effort to incentivize private developers to build new on-base military housing.

We look for evidence of pecuniary externalities resulting from military housing allowances. In particular, we ask whether increases to military housing allowances inflate local housing prices. The key parameter of interest is the elasticity of local area house prices with respect to BAH rates. Isolating the causal impact of military allowances on local house prices is challenging, especially because BAH rates are mechanically related to local area cost of living. We exploit changes to the military housing allowance formula from the late 1990s and early 2000s in order to identify the effect.

This paper proceeds as follows. Section 3.2 surveys the related literature and provides background on the U.S. military's housing policies. Section 3.3 presents a simple model in order to build intuition and benchmark our empirical exercise. Section 3.4 describes the data and discusses the empirical strategy. Section 3.5 presents a discussion of the results, and Section 3.6 concludes.

3.2 Background

3.2.1 Related Literature

This paper contributes to a growing empirical literature on the choice of policy instruments. This literature is largely concerned with explaining the prevalence of in-kind transfers in government programs. Currie and Gahvari (2008) provide a survey of the most compelling theoretical rationales for in-kind transfer programs. Pecuniary externalities are one such rationale. The reasoning is that in-kind transfers of a merit good can lower its price by shifting out the supply curve, which benefits all consumers of the good. Coate (1989) and Coate et al. (1994) provide examples of when pecuniary effects can cause in-kind transfers to be preferable to cash transfers.

A substantial empirical literature has studied the "choice of instrument" question, including the tradeoffs between in-kind and cash transfer programs. This literature has touched on healthcare, nutrition, and housing, among other government programs. A more modest literature has specifically explored the pecuniary effects of cash versus in-kind transfers. Cunha et al. (2015) find evidence for pecuniary effects in an experimental nutrition program in Mexico. Specifically, they compare villages receiving in-kind transfers of food with villages receiving equivalently-valued cash transfers and a third group of village receiving no transfer at all. They find that in-kind transfers reduce prices substantially, while cash transfers cause a positive but negligible increase in prices. Reducing the price of food was an explicit goal of the experiment in Cunha *et al.*, but pecuniary effects do not always operate as intended. Finkelstein (2007) documents evidence of pecuniary effects in the U.S. health insurance market, where the introduction of

Medicare drove up overall health care costs by making a large group of consumers (namely, the elderly) insensitive to prices.

Several other papers have studied price effects specifically in the context of low-income housing. Murray (1983) and (1999) study subsidized housing in the United States, finding evidence that subsidized housing programs lead to substantial crowd out of unsubsidized rental housing. Similarly, Eriksen and Rosenthal (2010) find evidence of crowd out due to supply-side construction subsidies.

Several papers have studied the incidence of housing vouchers, including Gibbons and Manning (2006) and, more recently, Collinson and Ganong (2014). Our paper is perhaps most similar to the latter paper. They study the incidence of Housing Choice Vouchers in the United States. They find that increases in the generosity of vouchers accrue mostly to landlords who receive higher rents; they find little evidence of a substitution to higher housing quality. Similar to our setting, they face the challenge of identifying the effect of a righthand side variable (rent ceilings) that is mechanically correlated with a lefthand side variable (rental prices). For identifying variation, they exploit changes to U.S. Department of Housing and Urban Development (HUD) policies, as well as HUD's policy of "rebenchmarking" Fair Market Rents once every decade, which uses the Decennial Census to correct for a decade worth of accumulated forecast error.

An important distinction between their paper and our own is that they assume that the share of voucher recipients is too small to influence prices in the overall housing market, whereas we are particularly interested on overall price effects within a local housing market. Voucher recipients account for just 2% of the U.S. housing market. While this is roughly on par with the share of military members in the housing market, military members tend to be concentrated in a limited number of housing markets near military bases, whereas voucher recipients are more uniformly distributed across housing markets.

Additional research on the pecuniary effects of in-kind versus cash transfers may offer insights into current debates over a so-called universal basic income ("UBI")—also known as a basic income guarantee ("BIG"). The details of universal basic income programs vary from proposal to proposal, but they generally consist of unconditional periodic cash transfers to all households. The transfers are unconditional in the sense that they are not means-tested—that is, all households are eligible, and all households receive the same benefit, regardless of other sources of income. Little evidence exists on the macroeconomic effects of UBI programs, and as Hoynes and Rothstein (2019) have recently argued, current pilot studies of the UBI may be poorly suited to better understanding the likely consequences of a guaranteed income. Some critics have argued, for example, that basic income programs are likely to generate high rates of inflation, potentially canceling out the benefits to families. Others have argued that inflation concerns are overblown, and they have sometimes supported their arguments with research from the policy instrument literature. A 2017 article from Vox, for example, extrapolated from the results of the Cunha et al. (2015) study in order to argue that the UBI's effects on inflation are likely to be negligible.[4] The program studied here—the U.S. military's Basic Allowance for Housing (BAH)—is similar to a UBI in that it is essentially unconditional, and families may spend it any way that they choose. Insofar as we find evidence of pecuniary externalities, our research may provide additional insights into the likely price effects of a UBI.

[4] *See* `https://www.vox.com/policy-and-politics/2017/9/20/16256240/mexico-cash-transfer-inflation-basic-income`.

3.2.2 Institutional Background

The modern military housing allowance is known as the Basic Allowance for Housing (BAH). The BAH system has been in place since 1998. Service members are eligible for BAH as long as 1) they are assigned to duty within the 50 United States, and 2) they have not been furnished with government-provided housing. On most major military installations, junior enlisted members without dependents are required to live on base in government housing. More senior service members, and any members with dependents, are generally afforded the choice of living on or off base. Members who live on base receive no BAH, while members who live off base receive an allowance in accordance with their rank, geographic location, and dependency status[5]. Importantly, BAH rates are set at the geographic level of a Military Housing Area (MHA). An MHA is defined as a collection of 5-digit ZIP codes. MHAs are roughly comparable to commuting zones, and they are designed to group together residential areas where members assigned to a particular base are most likely to live.

According to the DoD's own *BAH Primer*, the goal of the BAH program is to "help [military] members cover the costs of housing in the private sector." Specifically, BAH rates are intended to reflect the costs of rental housing and utilities, including electricity, heating fuel, water, and sewer. BAH rates are intended to match the average rental and utilities expenditures of civilians in the same local area with an equivalent annual income. In practice, the DoD has set separate housing standards for each rank and dependency status. For example, BAH rates for an E-6 with dependents are set according to the average cost of a three-bedroom townhouse, whereas an O-1 without dependents receives a BAH equivalent to the

[5]For the purpose of BAH determination, dependency status is a binary variable indicating members with at least one spouse or child.

cost of renting a two-bedroom apartment.

BAH rates are updated at least once per year to reflect the current cost of living in a given area. The DoD contracts with Robert D. Niehaus, Inc. (RDN), a private firm based in California, to collect annual data on the actual cost of rental housing and utilities in each of approximately 300 MHAs. The data collection process is surprisingly detailed, including telephone interviews with landlords, conversations with base commanders regarding which neighborhoods are most suitable for military members, and even on-site evaluations of housing units. The DoD publishes changes to the BAH schedule at least once per year, typically in December, in which case the rates take effect in the following calendar year. Importantly, BAH payments are exempt from State and Federal income taxes as well as Social Security taxes.

Throughout the remainder of the paper, the term BAH will be used to refer to the BAH rate for an E-6 with dependents, since this appears to be close to the median BAH received by all members. For robustness, we have replicated our analysis with higher and lower BAH rates as well. Appendix Table 1 shows the basic monthly pay and average BAH in 2016 for a sampling of rank and dependency status combinations.

Another important feature of the BAH system is its so-called "out-of-pocket" shares. Out-of-pocket shares were intended to slow the growth in total military housing allowance expenditures. Since the inception of the BAH system, service members have been expected to pay some fraction of their housing costs out-of-pocket. The exact fraction has varied over time, with a high of almost 20% in the early years of the program, and a low of 0% between 2005 and 2015. Currently, BAH-eligible service members are expected to pay 1% of their housing costs. Of course, individual service members' out-of-pocket contributions will depend on individual housing choices, with some members spending less than the allowance, and others spending more. We will discuss this feature in more detail in Section 3.

3.3 Conceptual Framework

In the following section we will attempt to estimate the pecuniary effects of military housing allowances on local housing markets. Since at least the late 1990s, military housing allowances have taken the form of monthly lump-sum transfers. They are indexed to the local cost of rental housing, but they are independent of a service member's actual housing consumption. As such, military housing allowances represent purely an income effect. Presented below is a simple static model of equilibrium housing supply and demand in a single local housing market. The goal of the model is to provide intuition and benchmark our expectation of the responsiveness of local housing prices to changes in military housing allowances.

Consider a local housing market consisting of two types of agents: civilian and military. Suppose there are N^c homogeneous civilians, N^m homogenous members of the military, and $N = N^c + N^m$ total agents. Civilians have the following indirect utility function:

$$V^c(w^c, r) = \max_{x,h} u^c(x, h) \quad \text{s.t.} \quad x + rh - w^c = 0, \tag{3.3.1}$$

where w^c is the civilian market wage, r is the price of housing, x is total non-housing consumption, and h represents quality-weighted housing consumptions.

Members of the military have the following indirect utility function:

$$V^m(w^m, b^m, r) = \max_{x,h} u^m(x,h) \quad \text{s.t.} \quad x + rh - w^m = 0, \tag{3.3.2}$$

where the military wage w^m includes both basic pay and special allowances (including a geographically-specific military housing allowance). Total demand for housing is given by the sum of civilian and military Marshallian demand curves:

$$H^* = N^c h^c(w^c, r) + N^m h^m(w^m, r) \tag{3.3.3}$$

Housing supply conforms to Diamond (2017), except that we continue to use quality-weighted quantities of housing. Quality-weighted housing units are sold at their marginal cost of production to absentee landlords, who rent units to local civilian and military residents. The housing supply curve is given by

$$r = \alpha + \gamma \ln(H), \tag{3.3.4}$$

where γ reflects the slope of the inverse supply curve. The elasticity of housing supply is therefore decreasing in γ.

Substituting total housing demand into the housing supply function yields the housing market equilibrium condition:

$$r - \alpha - \gamma \ln(N^c h^c(w^c, r) + N^m h^m(w^m, r)) = 0 \tag{3.3.5}$$

Applying the implicit function theorem, we can derive the responsiveness of rental prices r to military wages w^m:

$$\frac{d}{dw^m}\left[r-\alpha-\gamma\ln\left(N^c h^c(w^c,r)+N^m h^m(w^m,r)\right)\right]=0 \tag{3.3.6}$$

Assume that $\frac{dw^c}{dw^m}=0$. This is a justifiable assumption if we suppose either that military members comprise a small fraction of the labor force in a local housing market or that military workers tend to be imperfectly substitutable for civilian workers. [6]

Further assume that civilian and military agents share preferences so that $\frac{\partial h^c}{\partial r}=\frac{\partial h^m}{\partial r}\equiv\frac{\partial h}{\partial r}$. This is a stronger assumption, especially if we consider that military service members tend to be involuntarily assigned to a particular base for a fixed period of time. Whereas civilians can respond to an increase in rental prices along the intensive margin (i.e by consuming less quality-weighted housing) or the extensive margin (i.e. by moving to another housing market altogether), military members are generally constrained to an intensive margin response. For the sake of simplicity, we'll maintain this assumption.

Taking the total derivative of equation (3.3.6) yields an elasticity of local rental prices with respect to the military housing allowance:

$$\epsilon_{r,w^m}\approx\frac{-\gamma N^m\frac{\partial h^m}{\partial w^m}}{-1+\frac{\gamma}{H^*}N\frac{\partial h}{\partial r}}\frac{w^m}{r}>0, \tag{3.3.7}$$

We can use equation (3.3.7) to benchmark the response of local rental prices to a change in the military BAH rates. More generally, equation (3.3.7) confirms the intuition that we should expect see a larger price response where 1) the military share of the population is higher, 2) housing supply is relatively inelastic

[6]In fact, uniformed service members comprise just 0.8% of the labor force in the average county and just 38% of the labor force in even the most militarily dense counties.

(represented here by a large γ), or 3) housing demand is relatively inelastic.

The model also provides some insight into the effects of other unconditional cash transfer programs like the UBI, which we discussed briefly in section 3.2.1. Namely, we should expect the price effects of a UBI program to be largest where markets for large-expenditure items (for example, housing or healthcare) are less competitive (for example, in places where legal restrictions or geographic constraints combine to make housing supply inelastic).

3.4 Data Overview and Empirical Approach

3.4.1 Data Sources and Descriptive Statistics

We compile a panel dataset incorporating annual ZIP code-level information on military housing allowances, house prices, rental prices, and a variety of controls. Historical BAH data is publicly available from the DoD's Defense Travel Management Office.[7] An advantageous feature of this data is that we observe BAH rates for all U.S. ZIP codes, regardless of the proximity of a U.S. military installation. Although service members tend to be geographically concentrated near U.S. military bases, the DoD nevertheless sets BAH rates for all other areas in order to be prepared should any member or dependent establish BAH eligibility in those areas.[8] We observe annual BAH rates for all U.S. ZIP codes, ranks, and dependency statuses from 1998 to present.

As previously mentioned, BAH rates are generally set at the level of the MHA,

[7]For BAH rates, see: `http://www.defensetravel.dod.mil/site/pdcFiles.cfm?dir=/Allowances/BAH/PDF/`

[8]In fact, this has proved a useful practice since 2009, when veterans' educational benefits were overhauled through the so-called "Post-9/11 G.I. Bill." The new G.I. bill paid a BAH to veteran college students based upon the location of their college or university.

which is a collection of ZIP codes. Although BAH rates are established for all U.S. ZIP codes, it is not the case that all ZIP codes belong to an MHA. Rather, MHAs are established only for ZIP codes in the vicinity of major military installations. All other ZIP codes are grouped into County Cost Groups (CCGs). CCGs are not geographic units, but rather collections of ZIP codes with similar rental and utilities costs. In fact, as long as they share a similar cost of living, one ZIP code could share a CCG with another ZIP code many states away. In 2016 there were 303 MHAs and 39 CCGs. Figure 3-1 shows the boundaries of MHAs, with labels for several major military installations.

Our house price data come from the Zillow Home Value Index (ZHVI) All Homes Time Series,[9], which is a public use dataset published by online real estate database firm Zillow. Zillow's ZIP code-level data cover approximately 13,000 U.S. ZIP codes with quarterly estimates of the local home price index since 1996.

Since Zillow's rental data are available only since 2010, we instead rely on the U.S. Department of Housing and Urban Development's (HUD) Fair Market Rent (FMR) data. Quarterly estimates of the county-level average FMR are available since 1983, and we perform a crosswalk from U.S. county to U.S. ZIP code. Although FMR estimates are available for a variety of rental property types, we choose to emphasize the FMR for 3-bedroom apartments, since this appears consistent with the housing standard for the median BAH recipient. One limitation of the FMR data is that the FMR must be measured at a particular point in the rental price distribution, and the chosen point has varied over time and across counties. Since 1995, the vast majority of FMR estimates are made for the 40th percentile of local rental prices, but since 2001 some FMRs have been measured at the 50th per-

[9]Zillow data are available at `http://www.zillow.com/research/data/#median-home-value`

centile of local rents. Without knowledge of the full distribution of rental prices, we cannot perfectly convert between different measurement percentiles, so for the time-being we have dropped FMRs that are measured at anywhere but the 40th percentile.

For a variety of reasons, we choose to focus on the local house price index (HPI) rather than local Fair Market Rents. In addition to the limitations of the FMR cited above, we have observed that rental prices appear to lag behind house prices in our data. One possible explanation for this is that houses are transacted continuously, whereas large portions of the rental market tend to clear on an annual or semi-annual basis. With a limited time series of data, we have chosen to emphasize house prices, which appear to respond more fluidly. Nonetheless, we will present many of our results replacing local house prices with local rental prices.

Our empirical strategy leverages additional data on the locations of military bases throughout the United States. We determined the locations of military bases by referring to the DoD's annual *Demographics Reports*, which list military installations and their populations by ZIP code.[10] For robustness, we have also attempted specifications where we bin ZIP codes according to military share of the overall population.

Table 3.1 presents simple descriptive statistics, including the evolution of key variable since 1998 (the first year for which we have BAH data). In addition to BAH, FMR, and HPI, we show county-level median income estimates, which are available from the Census Bureau. Table 3.2 shows year-over-year rates of change in BAH, FMR, and HPI. One initial observation is that BAH rates have increased

[10]DoD *Demographics Reports* are available at: `http://www.militaryonesource.mil/footer?content_id=279104`.

monotonically since 1998, whereas FMR and HPI have experienced episodes of decline. Another observation is that BAH rates tend to have much lower within-period variance than either rental or house prices.

3.4.2 Potential Data Sources

In the future, we hope to augment our aggregate analysis with an individual analysis of microdata. We propose linking transaction-level real estate data with individual-level military personnel databases. Transactional real estate data are available for purchase from CoreLogic (formerly known as DataQuick). Individual-level military personnel records by Freedom of Information Act (FOIA) request to the DoD's Defense Manpower Data Center (DMDC), or directly from each branch[11]. By linking the data, we could potentially identify the housing characteristics of each BAH-eligible service member, including whether the service member rents or owns the unit. We would also observe house price data with much more precision. In addition to studying the elasticity of local house prices with respect to the BAH, we could also ask how individual service members respond to the income effect of a housing allowance. Moreover, we could investigate how service members select into on-base versus off-base housing.

[11]For example, the Total Army Personnel Database (TAPDB) is potentially accessible through the Office of Economic and Manpower Analysis (OEMA) at West Point, with which MIT has recently entered into a data agreement.

3.4.3 Empirical Approach

The simplest empirical approach regresses a local housing price index (HPI) in area c at time t on the current local Basic Allowance for Housing (BAH):

$$\log \text{HPI}_{ct} = \alpha + \beta \log \text{BAH}_{ct} + \epsilon_{ct} \tag{3.4.1}$$

where the coefficient of interest, β, can be interpreted as the $\%$ change in local house prices for a 1% change in the local BAH amount.

As discussed in the preceding section, an advantage of our BAH data is that we observe BAH rates for all U.S. ZIP codes, regardless of the presence of BAH-eligible service members. A slightly more sophisticated specification takes advantage of this feature of the data by interacting local BAH rates with an indicator variable for the presence of a military base:

$$\begin{aligned}\log \text{HPI}_{ct} = \alpha + \beta_0 \log \text{BAH}_{ct} + \beta_1 \text{Military Base}_{ct} + \\ \beta_3 \left(\log \text{BAH}_{ct} \times \text{Military Base}_{ct}\right) + \epsilon_{ct}\end{aligned} \tag{3.4.2}$$

The coefficient of interest is β_3. Intuitively, if the BAH formula is applied identically for ZIP codes with and without a military base, then the uninteracted BAH term effectively controls for the mechanical relationship between BAH and local housing prices. Insofar as we find a significant coefficient on the interaction term, this would represent the additional association between BAH rates and local house prices in areas where military members are likely to have an impact on equilibrium prices. Of course, the assumption that the BAH formula is applied uniformly is a strong one. In fact, we know that the housing cost data collected for MHAs is far richer than that collected for CCGs, which suggests that BAH rate-setting formula is unlikely to be evenly applied.[12]

Another empirical prediction is that local housing prices should be more responsive to BAH rate adjustments in areas with especially inelastic housing supply. Intuitively, if an increase in BAH rates represents a pure income effect, then the BAH rate adjustment shifts out the housing demand curve. This results in a large equilibrium price adjustment when housing supply is relatively inelastic. In order to test this prediction, we estimate the following specification, which interacts local BAH rates with an instrument for housing supply elasticity from Saiz (2010).

$$\begin{aligned} \log \mathrm{HPI}_{ct} = \alpha + \beta_0 \log \mathrm{BAH}_{ct} + \beta_1 \text{Supply Elasticity}_{ct} + \\ \beta_3 \left(\log \mathrm{BAH}_{ct} \times \text{Supply Elasticity}_{ct} \right) + \epsilon_{ct}, \quad (3.4.3) \end{aligned}$$

The coefficient of interest remains β_3, which we expect to be negative. This would confirm our prediction that local house prices are less responsive to BAH rates in areas where housing supply is inelastic. An important assumption is that the BAH formula is not constructed differentially across high- and low-elasticity areas.

In practice, we will estimate equations (3.4.2) and (3.4.3) with individual ZIP code-level fixed effects. We will also take first differences of the logged left- and righthand-side variables. In practice, this does not alter the interpretation of the coefficients, but it does allow us to control flexibly for year-pair fixed effects, which permits the national time trend to change across years. For example, our

[12] According to the BAH Primer, CCGs are established specifically because "collecting rental data for all such locations [without significant military personnel] is not practical." Instead, the DoD groups together non-military counties with similar HUD FMRs, and it then collects data on just a small sample of the counties within a CCG.

baseline specification with a Military Base interaction is:

$$\begin{aligned}\Delta_{t,t-1}\log \text{HPI}_c = \alpha + \beta_0\Delta_{t,t-1}\log \text{BAH}_c + \beta_1\text{Military Base}_{ct} + \\ \beta_3\left(\Delta_{t,t-1}\log \text{BAH}_c \times \text{Military Base}_{ct}\right) + \mathbf{X}_{ct}\gamma + \mu_c + \delta_{t,t-1} + \epsilon_{c,t,t-1}, \quad (3.4.4)\end{aligned}$$

where $\mathbf{X}_{ct}$ represents a vector of ZIP code- and time-varying controls. After first-differencing, the constant represents a linear time trend, and the year-pair fixed effects represent annual deviations from the linear trend.

Unfortunately, none of the specifications above address the fundamental challenge to our identification, which is that BAH rates are determined as an unknown function of local rental and utilities prices. Although the specifications in equations (3.4.2) and (3.4.3) provide casual evidence of pecuniary effects from military housing allowances, in order to credibly estimate the elasticity of local house prices with respect to BAH rates, we need a source of exogenous variation in BAH rates. The following section discusses one potential source of variation.

3.4.4 "Cohen Initiatives" Instruments

For identification, we exploit changes to the BAH formula from the late 1990s and early 2000s. In fact, there were two major sets of reforms during this period. First, in 1998, an earlier system of allowances known as Basic Allowance for Quarters/Variable Housing Allowance (BAQ/VHA) was replaced with the modern BAH system. Second, between 2001 and 2005, the DoD gradually reduced service members' out-of-pocket housing share from 19% to 0%. Following a shorthand used in some contemporary media accounts, we will refer to this entire series of reforms as the "Cohen Initiatives," named after then Secretary of Defense William

Cohen.

Overhaul of BAQ/VHA System

Prior to 1998, military housing allowances were paid out in two distinct payments: the Basic Allowance for Quarters (BAQ) and a Variable Housing Allowance (VHA). The BAQ was a lump-sum amount determined as a function of pay grade and dependent status (but without respect to geographic location); eligibility for the BAQ was identical to eligibility for the modern BAH. Additionally, service members residing in high housing cost areas received a VHA payment, which was a function of local housing costs. Specifically, the VHA was equal to the greater of either 1) the amount by which the local median monthly cost of housing exceeded 80% of the national median monthly cost of housing, or 2) the amount by which the service member's BAQ fell short of the local "adequate housing allowance floor."[13] Importantly, not all service members received a VHA payment. Service member were determined to be eligible for a VHA only when the first amount (i.e. median monthly local cost of housing minus 80% of national median monthly housing costs) was a positive number.[14] Abstracting from adequate housing allowance floors, the total allowance for area c in year t was approximately:

$$\begin{aligned}\text{Total Allowance}_{ct} &\approx \text{BAQ}_t + \mathbb{1}\{\widetilde{\text{cost-of-housing}}_{ct} - 0.80* \\ &\overline{\text{cost-of-housing}}_t > 0\} * \left[\widetilde{\text{cost-of-housing}}_{ct} - 0.80 * \overline{\text{cost-of-housing}}_t\right], \quad (3.4.5)\end{aligned}$$

[13]"Adequate housing allowance floors" were determined by the Secretary of Defense, but they were typically set equal to 85% of the HUD Fair Market Rent for a one bedroom apartment in this same area.

[14]Original rules for both the BAQ/VHA and BAH systems are specified in United States Code, Title 37 ("Pay and Allowances of the Uniformed Services"), Chapter 7 ("Allowances").

where $\widetilde{\text{cost-of-housing}}_{ct}$ and $\overline{\text{cost-of-housing}}_{t}$ represent the DoD's estimated median monthly cost of housing for same-rank service members and the national median monthly cost of housing, and cost-of-housing is determined as a function of local rental and utility prices: $\text{cost-of-housing}_{ct} = f(r_{ct}, u_{ct})$.

The popular perception was that the BAQ/VHA system was disadvantaging service members living in high cost areas. A related concern was that estimated cost-of-housing for service members was determined by surveying active duty service members about their actual housing costs, and VHA payments were in fact revised according to these annual surveys. Under this system, deviations from the true average cost of housing (for example, if many service members decided to economize on inexpensive housing) could lead to upward or downward spirals in the local VHA. Another unpopular feature was that individual service members were required to report their actual monthly housing expenditures, and the government kept 50% of any unspent allowance.

From 1998 onward, the BAQ/VHA system was replaced with the modern BAH system. The BAH system consists of a single lump-sum payment determined as a function of pay grade, dependency status, and local rental/utility costs. Initially, BAH payments were not intended to fully compensate service members for the cost of housing. Specifically, until 2000, the BAH rate was set equal to the difference between the monthly cost of adequate housing in a given area and 15% of the national average monthly cost of adequate housing in the United States.[15] In other words, BAH was calculated by

$$\text{BAH}_{ct} \approx \widehat{\text{cost-of-housing}}_{ct} - 0.15 * \overline{\text{cost-of-housing}}_{t}, \tag{3.4.6}$$

[15] https://www.gpo.gov/fdsys/pkg/USCODE-1997-title37/html/USCODE-1997-title37-chap7-sec403.htm

where $\widehat{\text{cost-of-housing}}_{ct}$ still represented the DoD's estimated cost-of-housing, but the DoD's estimates were now formed from an intensive data collection effort (and importantly, the DoD's estimates were independent of service members' actual housing expenditures). The second term represented the member's "out-of-pocket" payment. The % change in total allowance between the new BAH system and the previous BAQ/VHA system was approximately equal to

$$\%\Delta\text{Allowance}_c^{1997,1998} \approx \frac{\widehat{\text{cost-of-housing}}_{c,1998} - 0.15 * \overline{\text{cost-of-housing}}_{1998} - \text{BAQ}_{1997}}{\text{BAQ}_{1997}} \tag{3.4.7}$$

For service members living in areas just below 80 percent of the national median cost of housing—that is, those who were narrowly ineligible for location-specific VHA payments—the new BAH system was clearly much preferred. In future work we hope to exploit this non-linearity in housing allowances. Unfortunately, current data limitations (namely, the unavailability of pre-1998 housing allowance rates) make it extremely difficult to analyze the BAQ/VHA-BAH transitional period. Nonetheless, if equation (3.4.7) is correct, then we should see that the annual % change in total allowance is *positively* correlated with the baseline (pre-1998) cost-of-housing. Absent these policy changes, it is not *a priori* obvious why we should see any correlation between baseline cost-of-housing and subsequent rate of change in allowances. Moreover, it is not obvious why we should expect the rate of change of local house prices to vary with baseline housing costs.

Out-of-Pocket Reductions

In 2000, after just two years of the new BAH system, the DoD developed plans to gradually eliminate service members' out-of-pocket share of housing costs. Specifically, the Fiscal Year 2001 National Defense Authorization Act rewrote the law governing military housing allowances so that BAH rates would be commensurate with the full cost of "adequate housing determined for the area." The elimination of the out-of-pocket requirement forms the core of the reforms that came to be called the "Cohen Initiatives". If we interpret the statutory changes literally, then we can write the 2000-2001 % change in area c's BAH as

$$\%\Delta\text{BAH}_c^{2001,2000} \approx \frac{\Delta\widehat{\text{cost-of-housing}}_c^{2001,2000} + 0.15 * \overline{\text{cost-of-housing}}_{pre-2000}}{\widehat{\text{cost-of-housing}}_{c,2000} - 0.15 * \overline{\text{cost-of-housing}}_{2000}} \tag{3.4.8}$$

The elimination of out-of-pocket shares yields yet another empirical prediction. If equation (3.4.8) is correct, then we should expect to see that the annual % change in BAH rates is *negatively* correlated with the baseline (pre-2000) cost-of-housing. Intuitively, since out-of-pocket housing costs were calculated as a percentage of *national average* housing costs, the elimination of the out-of-pocket payment represented the same $ increase in all areas (at least conditional on pay grade and dependency status). This should have generated a larger % change in housing allowances for low-cost areas, since they start from a lower base. Again, absent these policy changes, it is not obvious why we should expect to see any such relationship.

We will attempt to leverage both the transition to the BAH system and the elimination of out-of-pocket shares as sources of identifying variation. An unusual aspect of our instrument is that its predicted first-stage impact varies over time. Namely, we expect baseline cost-of-housing in a given area to be positively correlated with the rate of BAH growth following the transition to the BAH system, and we expect the same baseline cost-of-housing to be negatively correlated with BAH growth after the elimination of out-of-pocket shares. Moreover, we know from contemporary media accounts that both policy changes were implemented gradually over a period of more than five years. In practice, we interact an area's Fair Market Rents in a baseline period (1996) with individual year dummies. The first-stage regression can be written in the following event study-style notation:

$$\Delta_{t,t-1} \log \text{BAH}_c = \alpha + \sum_t \delta_t \left(\log \text{FMR}_{c,1996} * \mathbb{1}Year_t \right) + \mathbf{X}_{ct}\beta + \mu_c + \lambda_{t,t-1} + \epsilon_{i,t,t-1}, \tag{3.4.9}$$

where $\mathbf{X}_{ct}$ represents a vector of ZIP code- and time-varying controls, and μ_c and $\lambda_{t,t-1}$ represent ZIP code and year-pair fixed effects, respectively. By interacting the baseline rental prices with individual year dummies, we allow the first-stage relationship to evolve flexibly over time.

The validity of our instrument rests on the assumption that baseline Fair Market Rents affect the subsequent rate of change for local house prices only through their impact on the military housing allowance, at least conditional on observable characteristics. Appendix Table 2 compares observable characteristics across quartiles of the baseline (1996) Fair Market Rent distribution. Ideally, we would see balance across quartiles, but this is not the case. Admittedly, this suggests that the exclusion restriction may be a strong assumption.

3.5 Results and Discussion

Our conceptual framework suggests that the pecuniary effects of BAH rate changes will be most substantial where 1) the military share of the population is high, and 2) housing supply is relatively inelastic. We find evidence for both of these claims. Table 3.3 presents our baseline OLS results. As per equation (3.4.2), we interact the change in log BAH rates with a dummy for the presence of a nearby military base. In the vicinity of a military base, a 1% increase in BAH rates is associated with a 0.06% increase in local house prices, in addition to the 0.10% increase that we see across all ZIP codes.

Although equation (3.4.2) presumes that the price effects of BAH rate adjustments accrue instantaneously, this is almost certainly not the case. In fact, we find evidence of considerable persistence. Table 3.4 presents OLS estimates of equation (3.4.2) including lagged righthand-side variables. We find large and significant effects up to three years after a change in BAH.

Table 3.5 presents OLS estimates of equation (3.4.3), which interacts changes in log BAH with an area's housing supply elasticity. As predicted, we find a negative coefficient on the interaction term, which suggests that an increase in the local BAH rate generates less of a price effect when local housing supply is highly elastic. Columns (3) and (4) show estimates from specifications including three-way interactions between BAH, housing supply elasticity, and the presence of a nearby military base. As predicted, we find that an increase in BAH is associated with the largest house prices increases in areas with a nearby military base and relatively inelastic supply of housing.

Table **??** presents first stage IV results corresponding to equation (3.4.9). These same results are depicted in Figure 3-2, which plots the event study coefficients from column (4) of Table **??**. Ideally, we would like to see the absence of pre-trends prior to the transition to the BAH system in 1998. However, since our data are limited to the period 1998-2009, we cannot effectively evaluate pre-trends. As predicted, we see a substantial relationship between baseline Fair Market Rents and the change in log BAH. Also as predicted, baseline FMR is positively correlated with BAH in early years and negatively correlated after approximately 2003. This appears consistent with what we know about the policy reforms encompassed in the so-called Cohen Initiatives. Intuitively, it appears as if the adoption of the new BAH system in 1998 had a disproportionately positive impact on local housing allowance rates in areas with a high baseline cost of housing. This effect was countered starting in 2001 by the gradual elimination of out-of-pocket housing contributions, which tends to favor areas with a low baseline cost of housing. By 2003, baseline FMRs were actually negatively correlated with the growth rate of BAH.

In Table **??**, it is worth noting that the three-way interaction terms (baseline FMR interacted with individual year dummies and the presence of a nearby military base) are all small and insignificant. In fact, this is positive news for our empirical strategy, since it supports our claim that the BAH formula is applied uniformly across areas with and without a nearby military base.

Our basic reduced form results are presented in Table 3.7, which shows the effects of baseline FMR on local house prices over time; coefficients from column (4) are plotted in Figure 3-2. We see evidence of a substantially lagged effect. Whereas Table **??** suggests that the elimination of out-of-pocket contributions started to favor low-baseline FMR areas as early as 2003, the reduced form effects on local house prices do not appear until approximately 2007. This is somewhat consistent with BAH-house price dynamics we found in Table 3.4. The coefficients on the interaction terms in columns (3) and (4) of Table 3.7 are all quite noisy, but they do tend to have the predicted signs.

Finally, Tables 3.8 and 3.9 present instrumental variables results corresponding to Tables 3.3 and 3.5, respectively. Column (4) of Table 3.8 presents our preferred IV estimates. We find that a 1% increase in military housing allowances generates more than a 0.25% increase in local house prices in areas with a nearby military base. This suggests substantial which entail large rents to landlords and a potentially significant burden for civilian renters. However, these IV results should be considered with caution. The exclusion restriction rests on the strong assumption that the baseline cost of housing affects subsequent house price growth only through the channel of BAH rates. Since our models are overidentified, we can perform Sargan-Hansen tests for overidentifying restrictions. Indeed, across all four specifications the J-statistics are perilously high, and in the preferred specifications we very nearly reject the null.

The results in Table 3.9 are decidedly mixed. We continue to find that pecuniary are strongest in areas with either a military presence or inelastic housing supply, but the coefficients on the three-way interaction terms in columns (3) and (4) are positive, counter to our prediction.

3.6 Conclusion

How should the government deliver services, and what are the tradeoffs of various policy instruments? This paper addresses the "choice of instrument" question in the context of a major cash benefit paid to U.S. servicemembers. In particular, we assess the potential for pecuniary externalities relating to military housing allowances. In its provision of housing, the military chooses between direct in-kind provision and cash transfers. In addition to limited on-base housing, the U.S. military pays geographically-specific lump-sum housing allowances to servicemembers throughout the United States. Especially in areas with a high military share of the overall population and inelastic housing supply, the additional demand generated by military housing allowances holds the potential to inflate prices for military members and civilians alike. Anecdotally, this very mechanism is partially to blame for skyrocketing rents and limited affordable housing options in places like San Diego and Honolulu.

With panel data on the evolution of ZIP code-level military housing allowances and rental and house prices, we study the impacts of military housing allowances on local housing markets. As a source of identifying variation, we exploit a suite of policy reforms known as the "Cohen Initiatives," which rewrote the military housing allowance formulas in the late 1990s and early 2000s. Using our preferred IV estimates, we find that increases to local military housing allowance rates are accompanied by potentially large pecuniary effects, with a 1% increase in military housing allowances leading to a 0.25% increase in local house prices in areas with a nearby military base. We find additional evidence that this effect is strongest in areas where housing supply is especially inelastic. However, concerns remain over the validity of our instrument, and our results should be considered with caution. In the future, we hope to extend our analysis to individual-level micro data.

Chapter 3 Bibliography

Besley, Timothy J. (1988) "A Simple Model of Merit Good Arguments," *Journal of Public Economics*, Vol. 35, pp. 371–383.

Cassiday, Ricky (2014) "Honolulu Rental Market: Affordable Rental Housing Study Update, 2014,"Technical report, Department of Community Services, City & County of Honolulu, http://dbedt.hawaii.gov/hhfdc/files/2015/02/RENTAL-HOUSING-STUDY-2014-UPDATE-CITY-COUNTY-OF-HONOLULU.pdf.

Coate, Stephen (1989) "Cash Versus Direct Food Relief," *Journal of Development Economics*, Vol. 30, pp. 199–224.

Coate, Stephen, Stephen Johnson, and Richard Zeckhauser (1994) "Pecuniary Redistribution through In-Kind Programs," *Journal of Public Economics*, Vol. 55, pp. 19–40.

Collinson, Robert and Peter Ganong (2014) "The Incidence of Housing Voucher Generosity," *Mimeo*.

Cunha, Jesse M., Giacomo De Giorgi, and Seema Jayachandran (2015) "The Price

Effects of Cash versus In-Kind Transfers," *Federal Reserve Bank of New York Staff Reports*, Vol. Staff Report No. 735.

Currie, Janet and Firouz Gahvari (2008) "Transfers in Cash and In-Kind: Theory Meets the Data," *Journal of Economic Literature*, Vol. 46, pp. 333–383.

Daly, George and Fred Giertz (1972) "Welfare Economics and Welfare Reform," *American Economic Review*, Vol. 62, pp. 131–138.

Defense Travel Management Office (2016) *A Primer on the Basic Allowance for Housing (BAH)*.

Diamond, Rebecca (2017) "Housing Supply Elasticity and Rent Extraction by State and Local Government Workers," *American Economic Journal: Economic Policy*, Vol. 9, pp. 74–111.

Eriksen, Michael D. and Stuart S. Rosenthal (2010) "Crowd out effects of place-based subsidized rental housing: New evidence from the LIHTC program," *Journal of Public Economics*, Vol. 94, pp. 953–966.

Finkelstein, Amy (2007) "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare," *The Quarterly Journal of Economics*, Vol. 122, pp. 1–37.

Garfinkel, Irwin (1973) "Is In-Kind Redistribution Efficient?" *Quarterly Journal of Economics*, Vol. 87, pp. 320–330.

Gibbons, Stephen and Alan Manning (2006) "The Incidence of UK Housing Benefit: Evidence from the 1990s Reforms," *Journal of Public Economics*, Vol. 90, pp. 799–822.

Hoynes, Hillary W. and Jesse Rothstein (2019) "Universal Basic Income in the U.S. and Advanced Countries," *NBER Working Paper Series*, Vol. Working Paper No. 25538.

Murray, Michael P. (1983) "Subsidized and Unsubsidized Housing Starts: 1961-1977," *Review of Economics and Statistics*, Vol. 65, pp. 590–597.

——— (1999) "Subsidized and Unsubsidized Housing Stocks 1935 to 1987: Crowding Out and Cointegration," *Journal of Real Estate Finance and Economics*, Vol. 18, pp. 107–124.

Nichols, Albert L. and Richard J. Zeckhauser (1982) "Targeting Transfers through Restrictions on Recipients," *American Economic Review*, Vol. 72, pp. 372–377.

Saiz, Albert (2010) "The Geographic Determinants of Housing Supply," *Quarterly Journal of Economics*, Vol. 125, pp. 1253–1296.

State of Hawaii Department of Business, Economic Development & Tourism (2014) "Construction and Hawaii's Economy: An Analysis of the Past and Forecast of the Future,"Technical report, Research & Economic Analysis Division, http://files.hawaii.gov/dbedt/economic/data_reports/construction_industry.pdf.

Chapter 3 Tables and Figures

Table 3.1: Descriptive Statistics, 1998-2009 ($ US)

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
BAH (mean)	642	660	724	830	906	985	1,046	1,129	1,176	1,209	1,299	1,397
(sd)	129	146	209	282	322	347	351	364	373	380	405	431
(min)	536	476	476	481	515	523	598	684	731	767	771	874
(max)	1,333	1,273	1,489	2,411	2,656	2,656	2,656	2,656	2,774	2,714	2,673	2,945
(n)	45,309	45,393	45,440	45,496	45,498	45,584	45,610	45,644	45,692	45,722	45,794	41,572
FMR (mean)	692	707	720	747	781	818	832	851	877	913	972	1,003
(sd)	206	214	227	258	282	320	335	300	293	311	330	334
(n)	39,063	39,097	39,111	39,132	39,132	39,220	39,230	39,258	39,284	39,291	39,320	39,318
HPI (mean)	128,642	139,241	151,974	164,996	177,579	194,894	219,274	249,915	266,815	262,468	242,683	223,863
(sd)	81,305	92,761	111,091	125,505	135,768	151,560	177,275	207,258	219,416	214,180	198,313	181,154
(n)	11,413	11,790	11,817	11,886	11,913	11,969	12,164	12,301	12,359	12,383	12,409	12,521
Income (mean)	37,463	38,105	40,121	39,527	39,765	40,617	41,880	43,479	45,256	47,448	49,072	47,525
(sd)	9,366	9,678	10,402	10,589	10,703	10,494	10,706	11,750	12,332	12,870	13,501	13,057
(n)	39,085	39,118	39,132	39,157	39,157	39,235	39,249	39,267	39,283	39,290	39,326	39,337

Figures presented in nominal dollars. All data is defined at the zipcode level. House Price Indices (HPI) are obtained from Zillow. Fair Market Rents (FMR) are obtained from the U.S. Department of Housing and Urban Development. FMRs are defined at the 40th percentile of local rental prices.

Figure 3-1: Military Housing Areas in 2016

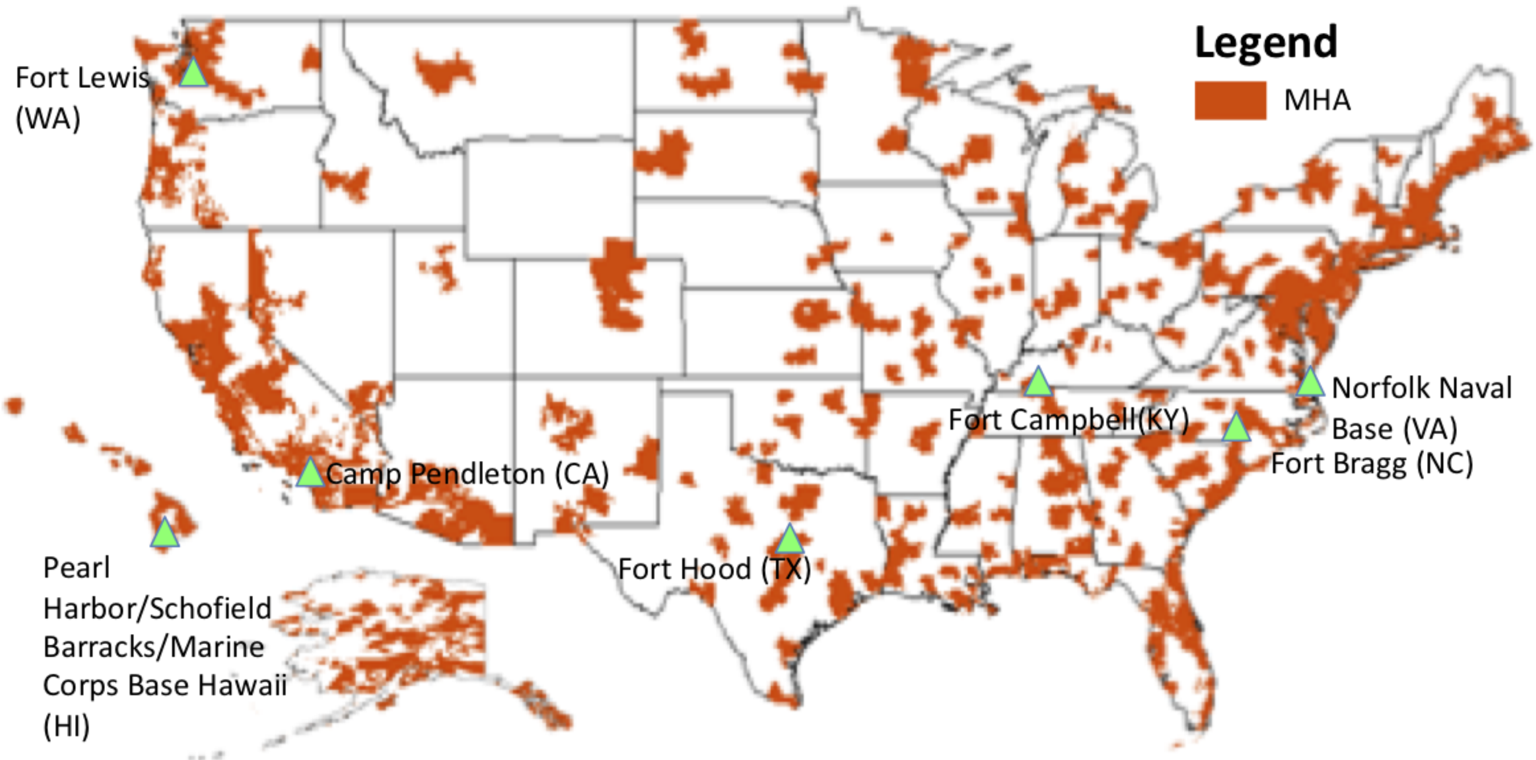


Table 3.2: Annual Percent Change in BAH, House Prices, and Rents, 1999-2009

	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
% Chg. BAH (mean)	2.3	7.9	13	8.4	8.3	6.5	7.8	4.2	2.9	7.2	7
(sd)	6.4	8	7.9	4.5	6.8	5.4	6.9	6.8	4	6.6	5.4
(min)	-27	-35	0	-52	-60	-79	-35	-68	-42	-21	-27
(max)	40	64	84	56	80	107	80	51	83	57	70
% Chg. FMR (mean)	2	1.5	3	4.1	3.8	1.2	3.6	3.6	3.9	6.3	3.3
(sd)	2.6	2.5	4	2.9	3.8	1.8	9.7	7.5	1.8	4.8	2.8
% Chg. HPI (mean)	5.7	6.7	6.8	6.5	8	9.3	12	6.4	-.55	-6.5	-6.9
(sd)	4.7	6.1	5.8	6	6.7	7.9	9	7.1	6.7	8.5	11

Figures show % year-over-year change in BAH, house price index, and fair market rents. All data is defined at the zipcode level.

Table 3.3: OLS Estimates of the Effect of BAH on Local House Prices

	(1)	(2)	(3)	(4)
Δ log (BAH) * Military Base	0.056**	0.054**	0.067***	0.060***
	[0.023]	[0.021]	[0.022]	[0.021]
Δ log(BAH)	0.133***	0.125***	0.108***	0.098***
	[0.019]	[0.016]	[0.020]	[0.017]
Military Base	0.003	0.001	-0.016*	-0.015
	[0.003]	[0.002]	[0.009]	[0.010]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs			Yes	Yes
Controls		Yes		Yes

This table presents OLS estimates of the effect of % change in BAH on % change in local house prices. Changes in BAH rates are interacted with a dummy variable the presence of a military base in the same 3-digit zipcode. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: OLS Estimates of the Effect of BAH on Local House Prices, Including Lags

	(1)	(2)	(3)	(4)
Δ log(BAH) * Military Base	0.089**	0.084***	0.093**	0.087**
	[0.035]	[0.030]	[0.038]	[0.033]
L.Δ log(BAH) * Military Base	0.065**	0.073**	0.080**	0.087**
	[0.030]	[0.028]	[0.035]	[0.035]
L2.Δ log(BAH) * Military Base	0.060**	0.073***	0.073**	0.089**
	[0.027]	[0.027]	[0.032]	[0.034]
L3.Δ log(BAH) * Military Base	0.019	0.025	0.037	0.043
	[0.028]	[0.029]	[0.030]	[0.034]
Δ log(BAH)	0.127***	0.126***	0.099***	0.091***
	[0.025]	[0.020]	[0.024]	[0.022]
L.Δ log(BAH)	0.080***	0.082***	0.056*	0.056*
	[0.027]	[0.026]	[0.030]	[0.029]
L2.Δ log(BAH)	0.059	0.064*	0.044	0.056
	[0.040]	[0.038]	[0.044]	[0.040]
L3.Δ log(BAH)	0.065*	0.080**	0.052	0.074*
	[0.036]	[0.034]	[0.040]	[0.038]
Military Base	-0.015	-0.014*	-0.027***	-0.025***
	[0.010]	[0.008]	[0.010]	[0.009]
L.Military Base	-0.003	-0.007	-0.006	-0.010
	[0.013]	[0.013]	[0.014]	[0.013]
L2.Military Base	0.005	0.012	0.007	0.013
	[0.015]	[0.014]	[0.015]	[0.014]
L3.Military Base	0.003	-0.007	-0.004	-0.009
	[0.016]	[0.015]	[0.015]	[0.013]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs			Yes	Yes
Controls		Yes		Yes

This table presents OLS estimates of the effect of change in log BAH on change in log HPI. Changes in BAH rates are interacted with a dummy variable the presence of a military base in the same 3-digit zipcode. Explanatory variables are lagged up to three periods. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.5: OLS Estimates of the Effect of BAH on Local House Prices, including Supply Elasticity

	(1)	(2)	(3)	(4)
Δ log(BAH) * Supp. Elast.	-0.050*	-0.056**	-0.031	-0.036
	[0.026]	[0.026]	[0.031]	[0.030]
Δ log(BAH)	0.225***	0.222***	0.184***	0.181***
	[0.050]	[0.049]	[0.057]	[0.056]
Δ log(BAH) * Military Base * Supp. Elast.			-0.058**	-0.060**
			[0.028]	[0.026]
Δ log(BAH) * Military Base			0.128***	0.124***
			[0.042]	[0.039]
Military Base			-0.010	-0.007
			[0.007]	[0.008]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes
Controls		Yes		Yes
Military Interactions			Yes	Yes

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This table presents OLS estimates of the effect of change in log BAH on change in log HPI. Changes in BAH rates are interacted with estimates of local housing supply elasticity from Saiz (2010). Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

Table 3.6: Effect of Cohen Initiative Instruments on BAH Rates (First Stage)

	(1)	(2)	(3)	(4)
1996 FMR * Year 2000	0.158***	0.158***	0.161***	0.160***
	[0.023]	[0.023]	[0.027]	[0.026]
1996 FMR * Military Base * Year 2000			-0.003	-0.000
			[0.027]	[0.028]
1996 FMR * Year 2001	0.099***	0.100***	0.089**	0.090**
	[0.032]	[0.033]	[0.035]	[0.036]
1996 FMR * Military Base * Year 2001			0.010	0.011
			[0.051]	[0.051]
1996 FMR * Year 2002	0.029	0.032	0.020	0.023
	[0.026]	[0.027]	[0.026]	[0.027]
1996 FMR * Military Base * Year 2002			0.014	0.015
			[0.029]	[0.029]
1996 FMR * Year 2003	-0.020	-0.016	-0.035	-0.031
	[0.024]	[0.025]	[0.026]	[0.028]
1996 FMR * Military Base * Year 2003			0.030	0.030
			[0.027]	[0.027]
1996 FMR * Year 2004	-0.053**	-0.050**	-0.070***	-0.068***
	[0.022]	[0.022]	[0.021]	[0.021]
1996 FMR * Military Base * Year 2004			0.027	0.028
			[0.031]	[0.032]
1996 FMR * Year 2005	-0.078***	-0.082***	-0.085***	-0.090***
	[0.029]	[0.028]	[0.031]	[0.031]
1996 FMR * Military Base * Year 2005			0.008	0.010

Continued on next page

Table 3.6 – continued from previous page

	(1)	(2)	(3)	(4)
			[0.049]	[0.049]
1996 FMR * Year 2006	-0.033	-0.031	-0.056	-0.054
	[0.035]	[0.035]	[0.036]	[0.036]
1996 FMR * Military Base * Year 2006			0.066	0.066
			[0.049]	[0.050]
1996 FMR * Year 2007	-0.027	-0.024	-0.034	-0.031
	[0.031]	[0.033]	[0.033]	[0.034]
1996 FMR * Military Base * Year 2007			0.010	0.013
			[0.023]	[0.022]
1996 FMR * Year 2008	-0.060*	-0.058*	-0.060*	-0.058
	[0.030]	[0.031]	[0.035]	[0.035]
1996 FMR * Military Base * Year 2008			-0.044	-0.042
			[0.039]	[0.040]
1996 FMR * Year 2009	-0.054***	-0.052***	-0.052***	-0.050***
	[0.017]	[0.017]	[0.018]	[0.018]
1996 FMR * Military Base * Year 2009			-0.028	-0.027
			[0.025]	[0.026]
1996 FMR * Military Base			-0.002	-0.002
			[0.002]	[0.002]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes
Population & Income Controls		Yes		Yes

Continued on next page

Table 3.6 – continued from previous page

	(1)	(2)	(3)	(4)

This table presents First Stage estimates of the effect of baseline log Fair Market Rent on change in log BAH where the baseline is measured in 1996. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3-2: Event Study Plot of First Stage Coefficients
Effect of Baseline Log FMR on Log BAH, 1999-2009

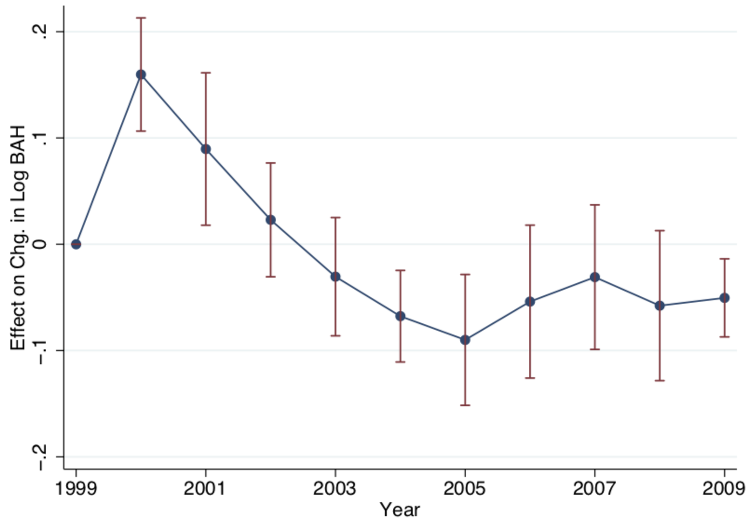


Table 3.7: Effect of Cohen Initiative Instruments on Local House Prices (Reduced Form)

	(1)	(2)	(3)	(4)
1996 FMR * Year 2000	0.104***	0.078***	0.101***	0.073***
	[0.012]	[0.016]	[0.012]	[0.015]
1996 FMR * Military Base * Year 2000			0.018*	0.026
			[0.011]	[0.016]
1996 FMR * Year 2001	0.086***	0.064***	0.081***	0.058***
	[0.013]	[0.012]	[0.014]	[0.012]
1996 FMR * Military Base * Year 2001			0.029**	0.026*
			[0.014]	[0.014]
1996 FMR * Year 2002	0.101***	0.091***	0.095***	0.084***
	[0.022]	[0.018]	[0.019]	[0.015]
1996 FMR * Military Base * Year 2002			0.029	0.027
			[0.021]	[0.020]
1996 FMR * Year 2003	0.125***	0.125***	0.119***	0.119***
	[0.014]	[0.011]	[0.012]	[0.010]
1996 FMR * Military Base * Year 2003			0.031	0.025
			[0.026]	[0.022]
1996 FMR * Year 2004	0.140***	0.132***	0.125***	0.116***
	[0.026]	[0.028]	[0.029]	[0.030]
1996 FMR * Military Base * Year 2004			0.050	0.048
			[0.034]	[0.029]
1996 FMR * Year 2005	0.118***	0.080***	0.103***	0.064***
	[0.025]	[0.027]	[0.019]	[0.021]
1996 FMR * Military Base * Year 2005			0.047	0.045

Continued on next page

Table 3.7 – continued from previous page

	(1)	(2)	(3)	(4)
			[0.034]	[0.030]
1996 FMR * Year 2006	0.016	0.010	0.013	0.006
	[0.027]	[0.024]	[0.025]	[0.021]
1996 FMR * Military Base * Year 2006			-0.001	-0.005
			[0.041]	[0.037]
1996 FMR * Year 2007	-0.079***	-0.082***	-0.074***	-0.079***
	[0.019]	[0.017]	[0.017]	[0.014]
1996 FMR * Military Base * Year 2007			-0.019	-0.019
			[0.037]	[0.035]
1996 FMR * Year 2008	-0.142***	-0.153***	-0.133***	-0.144***
	[0.035]	[0.034]	[0.031]	[0.032]
1996 FMR * Military Base * Year 2008			-0.016	-0.019
			[0.037]	[0.032]
1996 FMR * Year 2009	-0.174***	-0.188***	-0.158***	-0.172***
	[0.033]	[0.031]	[0.030]	[0.030]
1996 FMR * Military Base * Year 2009			-0.042	-0.048
			[0.037]	[0.035]
1996 FMR * Military Base			0.003***	-0.001
			[0.001]	[0.001]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes
Controls		Yes		Yes

Continued on next page

Table 3.7 – continued from previous page

	(1)	(2)	(3)	(4)

This table presents Reduced Form estimates of the effect of baseline log Fair Market Rent on the change in log House Price Index, where the baseline is measured in 1996. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 3-3: Event Study Plot of Reduced Form Coefficients
Effect of Baseline Log FMR on Log HPI, 1999-2009

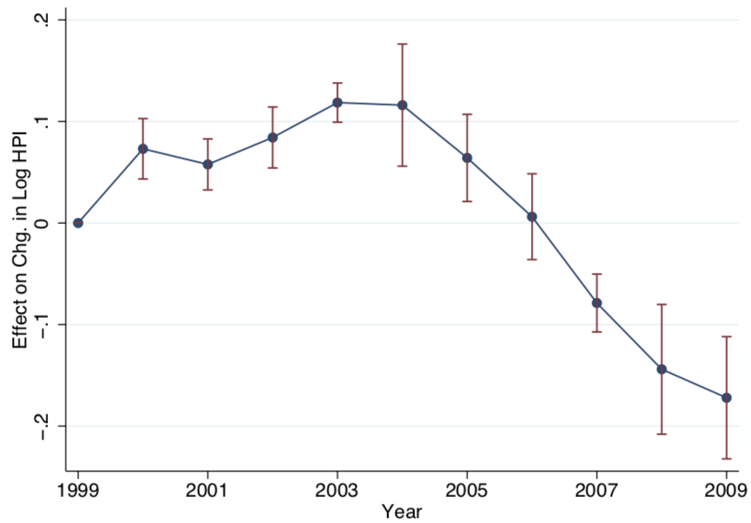


Table 3.8: IV Estimates of the Effect of BAH on Local House Prices

	(1)	(2)	(3)	(4)
Δ log(BAH) * Military Base			0.243***	0.287***
			[0.037]	[0.039]
Δ log(BAH)	0.522***	0.494***	0.459***	0.405***
	[0.045]	[0.048]	[0.032]	[0.041]
Military Base			-0.015***	-0.014***
			[0.005]	[0.005]
Sargan-Hansen J Stat.	22.099	22.615	27.460	28.692
Chi-Sq. P-val	0.011	0.006	0.115	0.073
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes
Controls		Yes		Yes
Military Interactions			Yes	Yes

This table presents instrumental variables estimates of the effect of a change in Log BAH on change in Log HPI. Baseline (1996) log Fair Market Rents, interacted with individual year dummies and military base dummies, are used to instrument for the endogenous regressors. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.9: IV Estimates of the Effect
of BAH on Local House Prices, including Supply Elasticity

	(1)	(2)	(3)	(4)
Δ log(BAH) * Supp. Elast.	-0.131***	-0.181***	-0.199***	-0.246***
	[0.015]	[0.016]	[0.017]	[0.017]
Δ log(BAH)	0.313***	0.356***	0.352***	0.360***
	[0.024]	[0.024]	[0.025]	[0.025]
Δ log(BAH) * Military Base * Supp. Elast.			0.081***	0.091***
			[0.014]	[0.014]
Δ log(BAH) * Military Base			0.049	0.111***
			[0.031]	[0.031]
Military Base			0.000	-0.002
			[0.003]	[0.003]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes
Controls		Yes		Yes
Military Interactions			Yes	Yes

This table presents instrumental variables estimates of the effect of a change in Log BAH on change in Log HPI. Baseline (1996) log Fair Market Rents, interacted with individual year dummies and military base dummies, are used to instrument for the endogenous regressors. Also interacted are estimates of local housing supply elasticity from Saiz (2010). Controls include county-level population and income dynamics.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

Chapter 1 Appendix

A.1 Judicial Multitask Model

Suppose that the judge chooses between two actions, a_1 and a_2, where the first action tends to expedite the case, and the second action tends to enhance procedural fairness. For example, a_1 might correspond to a pre-trial conference, and a_2 may correspond to granting additional time for discovery. Both actions are personally costly to the busy federal judge. The judge's cost function is

$$c(a_1, a_2) \tag{A.1.1}$$

where $\frac{\partial c}{\partial a_1} > 0$, $\frac{\partial^2 c}{\partial a_1^2} > 0$, and $\frac{\partial^2 c}{\partial a_2^2} > 0$. That is, the judge's private cost is increasing and convex in both actions.

These actions generate judicial output according to

$$x_1 = a_1 + \epsilon_1 \tag{A.1.2}$$

$$x_2 = a_2 + \epsilon_2, \tag{A.1.3}$$

where x_1 is inversely related to the judge's average motion processing time, and x_2 represents the substantive and the overall procedural fairness of her decisions. The individual judge's contribution to social welfare W is a function of both types of judicial output:

$$W = \phi_1 x_1 + \phi_2 x_2 \tag{A.1.4}$$

Among the most important features of the model is that, while x_1 is perfectly observable, x_2 is unobserved. That is, while the Congress and the Federal Judiciary can easily monitor a judge's average time-to-disposition as well as her disposition time on individual cases and motions, it is difficult to monitor her substantive or procedural fairness. The latter generally requires appellate review, which is both costly and subject to error in its own right.

Seeking to incentivize that which can be observed, judges are promoted with probability $p = \bar{p} + \beta x_1 + \nu$. That is, the probability of promotion increases linearly with the inverse of the judge's average motion processing time, and β represents the strength of the judge's incentives. For example, the introduction of the 6-month list, which tends to incentivize speed, would represent an increase the value of β.

The federal district judge chooses her actions a_1 and a_2 in order to maximize her private utility from promotion net of her private costs:

$$\max_{a_1,a_2} U(a_1, a_2) = u\left(p(a_1, a_2)\right) - c(a_1, a_2), \tag{A.1.5}$$

which yields the first order conditions:

$$[a_1] : \beta = \frac{\partial c(a_1, a_2)}{\partial a_1}$$
$$[a_2] : \frac{\partial c(a_1, a_2)}{\partial a_2} * a_2 = 0$$

If the cost of a_2 is always positive–that is, if $\frac{\partial c}{\partial a_2} > 0$–then the model yields a corner solution where the judge never expends any effort at procedural fairness. Instead, suppose that $\frac{\partial c(a_1, a_2=0)}{\partial a_2} \leq 0$, yielding an interior solution. That is, as long as efforts at fairness are costless at certain minimal levels, then the judge will expend some effort in that direction. Further suppose that actions a_1 and a_2 are substitutes, so that $\frac{\partial c^2(a_1, a_2)}{\partial a_1 \partial a_2}$. This seems like a reasonable assumption, given that actions tending to enhance procedural fairness will often tend to slow down an action and make speedy disposition more costly.

The key question is how the judge's behavior (namely, her choice of actions a_1 and a_2) responds to the strength of her incentives β. Differentiating her first order conditions with respect to β yields:

$$\frac{\partial a_1^*}{\partial \beta} > 0$$
$$\frac{\partial a_2^*}{\partial \beta} < 0$$

In other words, when x_1 is observable, x_2 is unobservable, and actions a_1 and a_2

are substitutes, high-powered incentives like the 6-month list will tend to increase investment in speed and decrease investment in procedural fairness.

A.1.1 Incorporating Judge Procrastination

The goal of this model is to evaluate how a present-biased responds to incentives similar to those generated by the six-month list. The model borrows much of its architecture from other models used to study the effects of final[1] or interim deadlines[2] on the behavior of present-biased agents. The six-month list, however, imposes a somewhat unique choice structure with similarities to both final and interim deadlines. The six-month list is similar to an interim deadline in the sense that it is non-binding—much like a student subject to an interim deadline for submitting a rough draft of a writing assignment, the judge is free to allocate her effort across the deadline, even if it triggers an appearance on the six-month list. However, if she chooses to discontinue her work in order to avoid an appearance on the six-month list, then her work becomes final, and it is too late to invest effort in order to improve it.

I will start by introducing a basic model of a judge subject to present-bias (i.e. procrastination). After establishing the framework, I will consider the likely effects of implementing a six-month list-style regime. Suppose a judge a required to enter an order disposing of a single motion. She has two periods $t \in \{1, 2\}$ during which to work on the order. At the end of period 1, she may choose to either continue working on the order during period 2, or she may discontinue her work

[1] *See, e.g.*, Ted O'Donoghue & Matthew Rabin, *Incentives for Procrastinators*, 114 Q. J. ECON. 769 (1999).

[2] *See, e.g.*, Fabian Herweg & Daniel Muller, *Performance of Procrastinators: on the Value of Deadlines*, 70 THEORY & DECISION 329 (2011); Ted O'Donoghue & Matthew Rabin, *Incentives and Self-Control* (2005) (unpublished working paper).

and enter the order immediately. For each period that she works on the order, she chooses an effort level $e_t \geq 0$ for which she incurs a cost of $c(e_t)$ where $c'(\cdot) > 0$ and $c''(\cdot) > 0$. The judge is rewarded for her efforts in period 3, where her probability of promotion $p\left(\sum_{t=1}^{2} e_t + \epsilon\right)$ is strictly increasing and in her total effort invested in the order ($p'(\cdot) > 0$; $p''(\cdot) < 0$). The noise term ϵ reflects the inherently imperfect observability of a judge's effort on any single motion. The judge's intertemporal preferences are given by a standard hyperbolic discounting utility function:

$$U_t(u_t, u_{t+1}, \ldots, u_T) = u_t + \beta \sum_{\tau=t+1}^{T} \delta^{\tau-t} u_\tau,$$

where u_t represents the judge's instantaneous utility in period t, $\delta \in [0, 1]$ represents a time-consistent (i.e. exponential) discount factor, and $\beta \in [0, 1]$ denotes the degree of the judge's time-inconsistent present bias. For convenience, we will assume that the judge's has a time-consistent discount factor of $\delta = 1$.

First we consider a regime without the six-month list. In the first period the judge chooses an actual first-period effort level e_1, decides whether to continue working in period 2, and conditional on choosing to continue, chooses a planned second-period effort level e_2. The judge's first-period intertemporal utility function is given by

$$U_1 = \max\left\{-c(e_1) + \beta p(e_1), -c(e_1) - \beta c(e_2) + \beta p(e_1 + e_2)\right\}. \tag{A.1.6}$$

The judge's second-period intertemporal utility function, which depends upon whether she chooses to continue working in period 2, is given by

$$U_2 = \begin{cases} \beta p(e_1) \text{ if judge discontinues work} \\ -c(e_2) + \beta p(e_1 + e_2) \text{ if judge continues work} \end{cases} \tag{A.1.7}$$

Time-Consistent Judge

First we consider a time-consistent judge. For a time-consistent agent, $\beta = 1$, which reflects an absence of present-bias. Since a time-consistent judge's preferences do not change over time, she is able to commit to whichever future course of action maximizes U_1. She continues working in the second period if $-c(e_1^*) - c(e_2^*) + \beta p(e_1^* + e_2^*) > -c(\tilde{e_1}) + p(\tilde{e_1})$, where $\{e_1^*, e_2^*\} = \arg\max_{e_1,e_2} -c(e_1) + p(e_1), -c(e_1) - c(e_2) + p(e_1 + e_2)$ and $\{\tilde{e_1}\} = \arg\max_{e_1} -c(e_1) + \beta p(e_1)$. Assuming that she continues working into the second period, the judge's optimal sequence of effort is characterized by the first-order conditions

$$c'(e_1) = c'(e_2) = p(e_1 + e_2). \tag{A.1.8}$$

That is, the judge invests the same in both periods. Moreover, due to the convexity of the cost curve, it can be shown that the judge will always prefer to continue working after the first period so that she may smooth her effort across two periods.

Present-Biased Judge

Next we consider a present-biased judge. We will assume for sake of simplicity that the judge is naive to her time-inconsistent preferences; the main results extend to the case of a sophisticated judge. The severity of the judge's present-bias is reflected by $\beta \in (0, 1]$.

In the first period, the naive agent chooses her actual first-period effort e_1^* and her planned second-period effort $\widehat{e_2^*}$ in order to maximize U_1. She continues working after the first period if $-c(e_1^*) - \beta c(\widehat{e_2^*}) + \beta p(e_1^* + \widehat{e_2^*}) > -c(\tilde{e_1}) + p(\tilde{e_1})$. The naive judge will always choose to continue working in the second period due to both the convexity of the cost function and the perceived lower cost of effort in the second period. Actual first-period effort e_1^* and planned second-period effort $\widehat{e_2^*}$ are characterized by the first order conditions

$$\begin{aligned} c'(e_1^*) &= \beta g'\left(e_1^* + \widehat{e_2^*}\right) \\ c'(\widehat{e_2^*}) &= p'\left(e_1^* + \widehat{e_2^*}\right). \end{aligned} \tag{A.1.9}$$

In the second period the judge is surprised to learn that her current effort is no less costly than it was in the previous period. The judge therefore re-optimizes in the second period, with her actual second-period effort e_2^* being characterized by

$$c'(e_2^*) = \beta p'(e_1^* + e_2^*). \tag{A.1.10}$$

Implementing the Six-Month List

Next I will modify my model to incorporate a policy like the six-month list. Before the imposition of the six-month list, a judge's probability of promotion depended only upon the effort she exerted plus a random noise term.

$$p(e_1, e_2) = \begin{cases} g(e_1 + \epsilon) \\ g(e_1 + e_2 + \epsilon) - B, \end{cases} \tag{A.1.11}$$

where $g(\cdot)$ is strictly increasing and concave in effort e and the constant B reflects a punishment for judges whose motions appear on the six-month list. In other words, a judge is free to continue working in the second period if she chooses, but the cost of doing so is a predictably lower probability of future promotion.

Proposition A.1.1. *For a naive or sophisticated present-biased judge, $\exists$ incentive B such that a non-complying judge (who continues working in the second period) becomes a complying judge (who concludes work in period one).*

Proposition A.1.2. *For a naive or sophisticated present-biased judge, total effort is weakly decreasing with compliance.*

Proposition A.1.3. *For a naive or sophisticated present-biased judge, for a given incentive B, compliance with the six-month list is increasing in the variance of $epsilon$.*

A.2 Additional Tables & Figures

Figure A-1: Excerpt from the CJRA six-month report for the period ending September 30, 2016

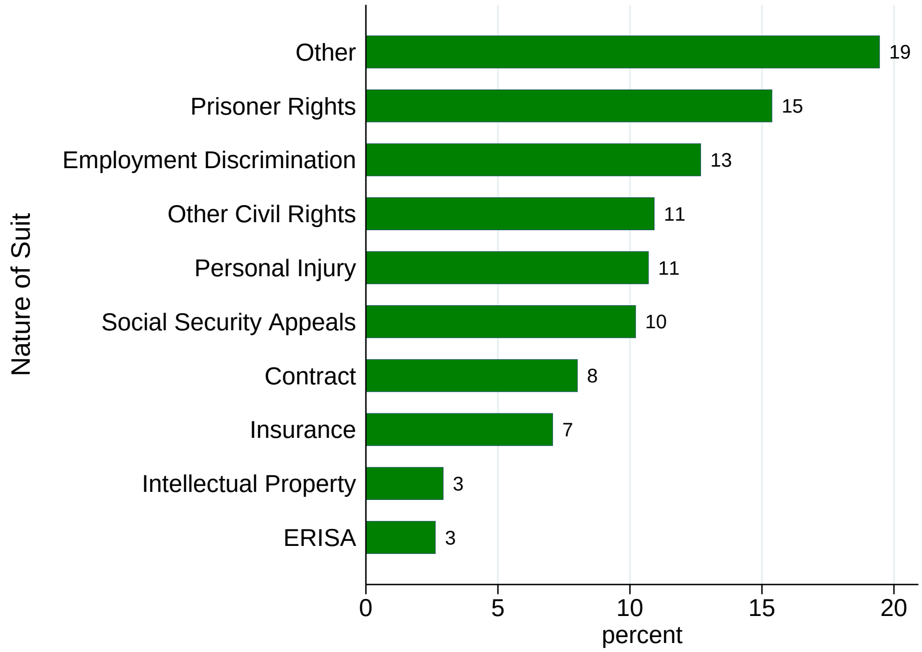
CJRA Table 8—Report Of Motions Pending Over Six Months
For Period Ending September 30, 2016

DC Circuit

District Judge BATES, JOHN D.

Office	Docket Number	NOS Code	Case Title	Motion Text	CJRA Deadline	Status	Status Description
1	15-cv-01945	360	OWENS et al v. BNP PARIBAS S.A. et al	MOTION to Dismiss	08/30/2016	B	Opinion/Decision in Draft
						Q	Complexity of Case
				MOTION for Summary Judgment	09/03/2016	B	Opinion/Decision in Draft
						Q	Complexity of Case

Figure A-2: Distribution of Case Types



Note: Category "Other" includes miscellaneous statutory claims, tax-related claims, certain employment rights claims, as well as a wide variety of other case types.

Table A.1: Effect of Reporting Time on Months Until Motion Disposition
Individual Reporting Month Dummies

	(1)	(2)	(3)	(4)
8-9 Months Reporting Time	0.213***	0.114***	0.137***	0.138***
	(0.029)	(0.042)	(0.041)	(0.049)
9-10 Months Reporting Time	0.365***	0.479***	0.487***	0.486***
	(0.029)	(0.043)	(0.043)	(0.049)
10-11 Months Reporting Time	0.483***	0.504***	0.517***	0.517***
	(0.029)	(0.032)	(0.032)	(0.032)
11-12 Months Reporting Time	0.639***	0.557***	0.573***	0.575***
	(0.030)	(0.044)	(0.044)	(0.051)
12-13 Months Reporting Time	0.633***	0.736***	0.726***	0.724***
	(0.030)	(0.042)	(0.042)	(0.049)
Observations	206,187	206,187	206,151	206,151
Case & Motion Controls	Yes	Yes	Yes	Yes
Calendar Trends		Yes	Yes	Yes
District*Year FEs			Yes	Yes
Day-of-Month FEs				Yes
Mean of Dep. Variable	5.36	5.36	5.36	5.36
Mean of Indep. Var	10.03	10.03	10.03	10.03

This table presents OLS estimates of the effect of additional reporting time on months until motion disposition. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

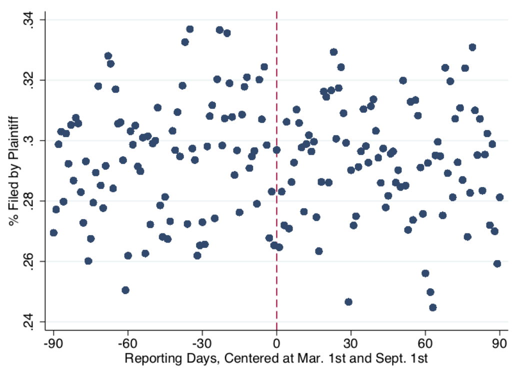
Table A.2: Effect of Reporting Time on Months Until Case Disposition Individual Reporting Month Dummies

	(1)	(2)	(3)	(4)
8-9 Months Reporting Time	0.050	-0.128	-0.115	0.061
	(0.109)	(0.146)	(0.145)	(0.170)
9-10 Months Reporting Time	0.027	0.372**	0.416***	0.244
	(0.107)	(0.149)	(0.148)	(0.167)
10-11 Months Reporting Time	0.116	0.289**	0.312***	0.301***
	(0.109)	(0.117)	(0.116)	(0.116)
11-12 Months Reporting Time	0.389***	0.367**	0.367**	0.522***
	(0.108)	(0.154)	(0.153)	(0.175)
12-13 Months Reporting Time	0.210**	0.407***	0.457***	0.280*
	(0.107)	(0.146)	(0.145)	(0.167)
Observations	183923	183923	183887	183887
Case & Motion Controls	Yes	Yes	Yes	Yes
Calendar Trends		Yes	Yes	Yes
District*Year FEs			Yes	Yes
Day-of-Month FEs				Yes
Mean of Dep. Variable	23.38	23.38	23.37	23.37
Mean of Indep. Var	10.04	10.04	10.04	10.04

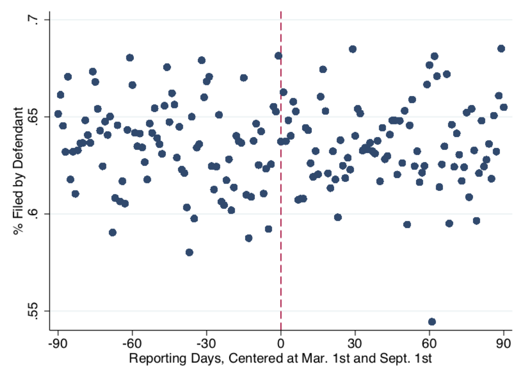
This table presents OLS estimates of the effect of additional motion reporting time on months until overall case disposition. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

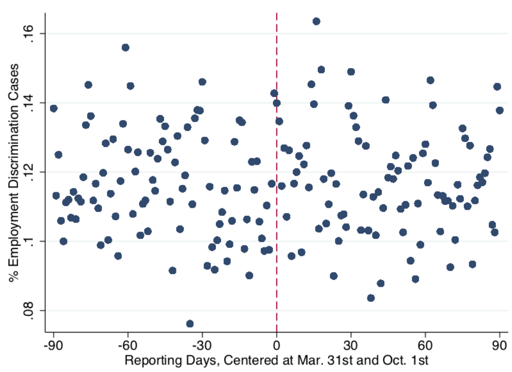
Figure A-3: Distribution of Covariates Across Filing Date Cutoffs



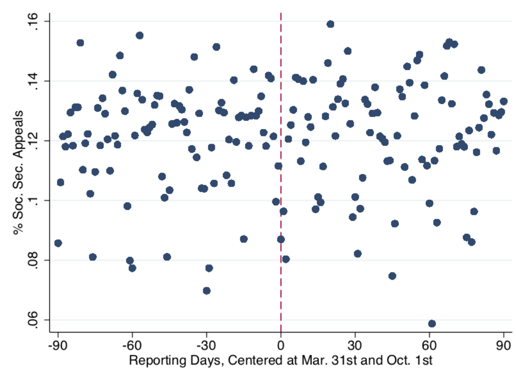
(a) % Motions filed by Plaintiff



(b) % Motions filed by Defendant

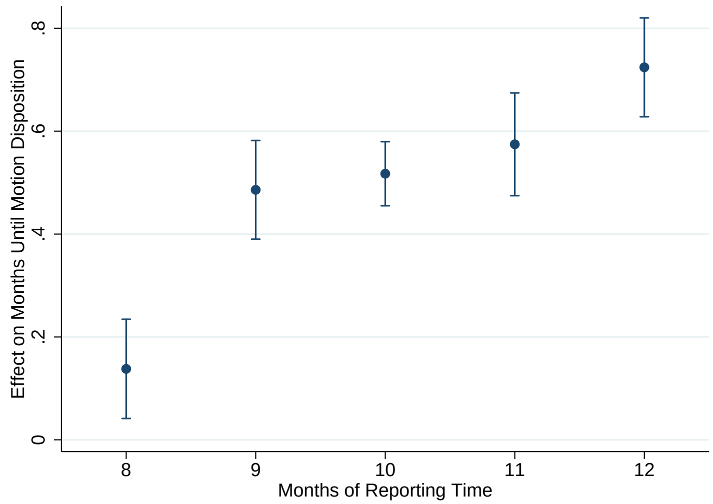


(c) % Employment Discrimination Cases

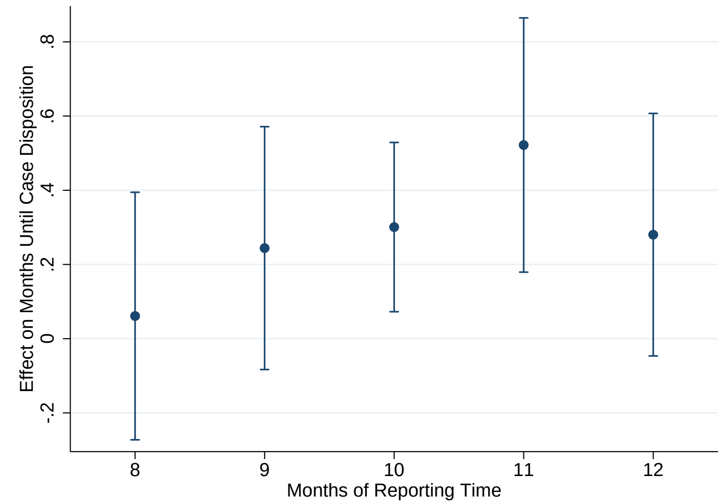


(d) % Social Security Cases

Figure A-4: Effect of Reporting Time on Months Until Disposition



(a) Motion Disposition



(b) Overall Case Disposition

Appendix Tables A.1 and A.2 report the coefficients and standard errors corresponding to Appendix Figures A-4a and A-4b.

Table A.3: Effect of Motion Reporting Time on Motion Outcomes

	(1)	(2)	(3)	(4)
		Motion Granted		
Months until Report	0.0009	0.0013**	0.0013**	0.0015**
	[0.0006]	[0.0006]	[0.0006]	[0.0006]
Observations	206,199	206,199	206,163	206,163
Mean of Dep. Variable	.48	.48	.48	.48
		Motion Denied		
Months until Report	-0.0002	-0.0004	-0.0005	-0.0003
	[0.0006]	[0.0006]	[0.0006]	[0.0006]
Observations	206,199	206,199	206,163	206,163
Mean of Dep. Variable	.36	.36	.36	.36
		Motion Granted in Part		
Months until Report	-0.0002	-0.0003	-0.0003	-0.0007
	[0.0004]	[0.0005]	[0.0005]	[0.0005]
Observations	206,199	206,199	206,163	206,163
Mean of Dep. Variable	.14	.14	.14	.14
		Motion Decided in Favor of Defendant		
Months until Report	0.0008	0.0011*	0.0012*	0.0015**
	[0.0006]	[0.0007]	[0.0007]	[0.0007]
Observations	191,635	191,635	191,614	191,614
Mean of Dep. Variable	.57	.57	.57	.57
		Motion Decided in Favor of Plaintiff		
Months until Report	-0.0003	-0.0004	-0.0006	-0.0005
	[0.0006]	[0.0006]	[0.0006]	[0.0006]
Observations	191,635	191,635	191,614	191,614
Mean of Dep. Variable	.28	.28	.28	.28
Case & Motion Controls	Yes	Yes	Yes	Yes
Calendar Trends		Yes	Yes	Yes
District*Year FEs			Yes	Yes
Day-of-Month FEs				Yes
Mean of Indep. Var	10.03	10.03	10.03	10.03

This table presents OLS estimates of the effect of additional reporting time on the probabilities of various motion outcomes. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

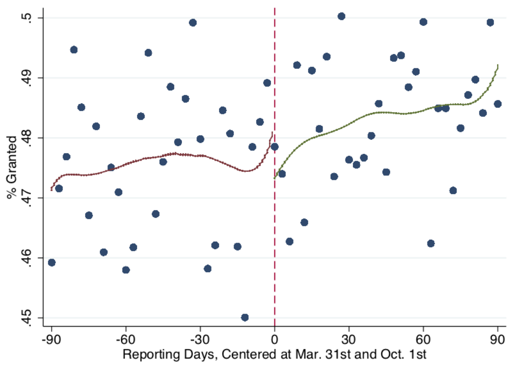
Table A.4: Effect of Motion Reporting Time on Appellate Outcomes

	(1)	(2)	(3)	(4)
	Appeal Filed			
Months until Report	0.0000	-0.0000	0.0005	0.0005
	[0.0003]	[0.0003]	[0.0006]	[0.0006]
Observations	480,534	480,534	206,163	206,163
Mean of Dep. Variable	.22	.22	.26	.26
	Affirmed on Appeal			
Months until Report	0.0011	0.0011	0.0026**	0.0028**
	[0.0009]	[0.0009]	[0.0013]	[0.0013]
Observations	106,202	106,202	53,755	53,755
Mean of Dep. Variable	.48	.48	.51	.51
	Reversed on Appeal			
Months until Report	-0.0010*	-0.0011*	-0.0007	-0.0008
	[0.0005]	[0.0005]	[0.0007]	[0.0007]
Observations	106,202	106,202	53,755	53,755
Mean of Dep. Variable	.10	.10	.08	.08
Case & Motion Controls	Yes	Yes	Yes	Yes
Calendar Trends		Yes	Yes	Yes
District*Year FEs			Yes	Yes
Day-of-Month FEs				Yes
Motion Outcome Dummies			Yes	Yes
Mean of Indep. Var	10.03	10.03	10.03	10.03

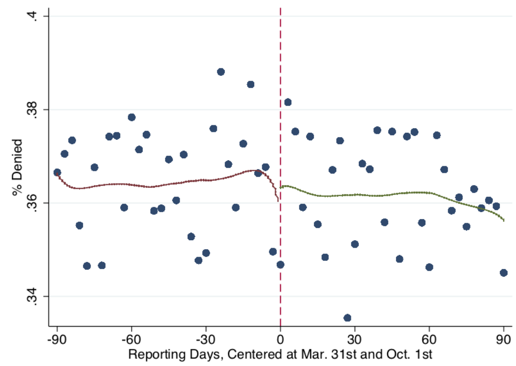
This table presents OLS estimates of the effect of additional reporting time on the probabilities of various appellate outcomes. Reporting time is measured in the number of months between the day on which a motion was filed and the earliest possible date on which it could appear on a CJRA 6-month report. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects. Columns (1) and (2) are unconditional on motion outcomes (i.e. whether the motion was granted, denied, granted in part, etc.), while columns (3) and (4) condition on motion outcome dummies. Robust standard errors are in parentheses.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

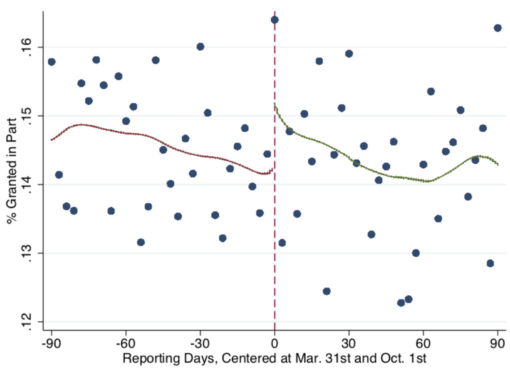
Figure A-5: Regression Discontinuity Plots of Motion and Appellate Outcomes



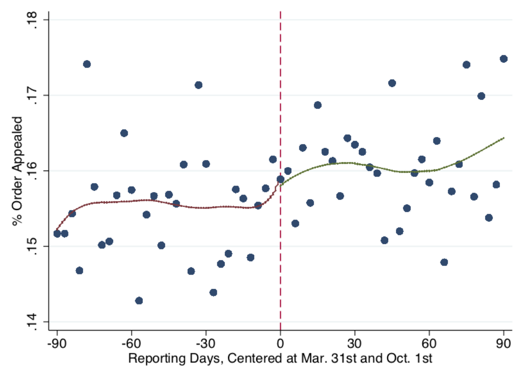
(a) % Granted



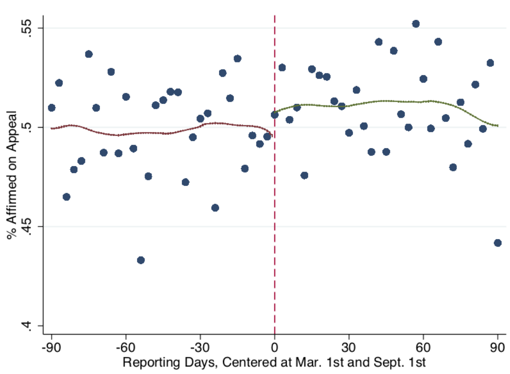
(b) % Denied



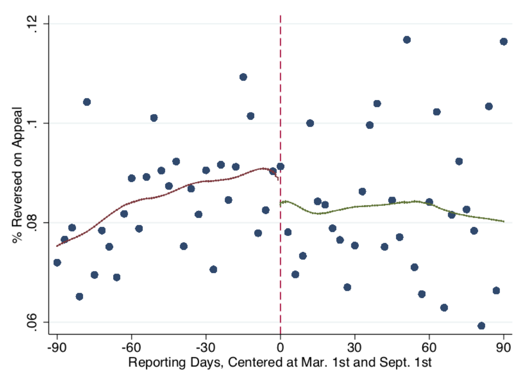
(c) % Granted in Part



(d) % Appealed



(e) % Affirmed



(f) % Reversed

Table A.5: Regression Discontinuity Estimates
Effect of Reporting Time on Baseline Motion- and Case-Level Characteristics

	Parametric			Non-Parametric (Local Linear)	
	(1) Linear	(2) Quadratic	(3) Cubic	(4) IK Bandwidth	(5) CCT Bandwidth
	Motion Filed by Defendant				
Filed After Cutoff	0.000***	-0.000***	-0.000***	-0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Mean of Dep. Variable	.64	.64	.64	.65	.63
Observations	204,137	204,137	204,137	6,826	35,818
	Motion Filed by Plaintiff				
Filed After Cutoff	0.001	0.001	0.003	0.013	0.010
	[0.002]	[0.003]	[0.004]	[0.018]	[0.006]
Mean of Dep. Variable	.29	.29	.29	.28	.29
Observations	204,137	204,137	204,137	4,208	38,323
	Motion Filed in Employment Discrimination Case				
Filed After Cutoff	0.000***	0.000***	-0.000***	-0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Mean of Dep. Variable	.12	.12	.12	.14	.12
Observations	204,137	204,137	204,137	4,208	51,247
	Motion Filed in Social Security Case				
Filed After Cutoff	-0.000***	-0.000***	0.000***	0.000	0.000
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Mean of Dep. Variable	.12	.12	.12	.10	.12
Observations	204,137	204,137	204,137	4,208	42,764

This table presents regression discontinuity (RD) estimates of the effect of additional reporting time on baseline motion- and case-level characteristics, including whether the motion filed by the plaintiff vs. defendant, and whether the motion was filed in an employment discrimination or social security-related case. The running variable represents the motion filing date relative to the six-month list eligibility cutoff. Motions filed just before the cutoff are eligible for the current six month list, whereas motions filed just after the cutoff have an additional six months before they might appear on a list. Columns (1)-(3) are estimated parimetrically with linear, quadratic, and cubic polynomials, respectives. Columns (4)-(5) are estimated nonparametrically with local linear regressions, using the IK and CCT methods of optimal bandwidth selection, respectively. All columns include basic case- and motion-level controls, including judge, district, and filing-year fixed effects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.6: Regression Discontinuity Estimates
Effect of Reporting Time on Motion-Level Outcomes

	Parametric			Non-Parametric (Local Linear)	
	(1) Linear	(2) Quadratic	(3) Cubic	(4) IK Bandwidth	(5) CCT Bandwidth
			Motion Granted		
Filed After Cutoff	0.002	-0.002	-0.000	-0.046	-0.016
	[0.005]	[0.007]	[0.009]	[0.037]	[0.014]
Mean of Dep. Variable	.48	.48	.48	.48	.47
Observations	204,137	204,137	204,137	6,826	35,818
			Motion Denied		
Filed After Cutoff	0.000	0.001	-0.002	0.030	0.002
	[0.004]	[0.006]	[0.008]	[0.033]	[0.013]
Mean of Dep. Variable	.36	.36	.36	.35	.36
Observations	204,137	204,137	204,137	6,826	44,864
			Motion Granted in Part		
Filed After Cutoff	0.001	0.007	0.006	0.009	0.011
	[0.003]	[0.005]	[0.007]	[0.032]	[0.010]
Mean of Dep. Variable	.14	.14	.14	.16	.14
Observations	204,137	204,137	204,137	4,208	57,360

This table presents regression discontinuity (RD) estimates of the effect of additional reporting time on motion-level outcomes, including whether the motion was granted, denied, or granted in part. The running variable represents the motion filing date relative to the six-month list eligibility cutoff. Motions filed just before the cutoff are eligible for the current six month list, whereas motions filed just after the cutoff have an additional six months before they might appear on a list. Columns (1)-(3) are estimated parimetrically with linear, quadratic, and cubic polynomials, respectives. Columns (4)-(5) are estimated nonparametrically with local linear regressions, using the IK and CCT methods of optimal bandwidth selection, respectively. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.7: Regression Discontinuity Estimates
Effect of Reporting Time on Appellate Outcomes

	Parametric			Non-Parametric (Local Linear)	
	(1) Linear	(2) Quadratic	(3) Cubic	(4) IK Bandwidth	(5) CCT Bandwidth
			Appeal Filed		
Filed After Cutoff	0.002	0.005	-0.004	-0.054*	-0.020*
	[0.004]	[0.006]	[0.008]	[0.029]	[0.011]
Mean of Dep. Variable	.26	.26	.26	.26	.26
Observations	204,137	204,137	204,137	6,826	53,262
			Affirmed on Appeal		
Filed After Cutoff	0.007	-0.004	0.027	-0.282***	0.007
	[0.009]	[0.013]	[0.018]	[0.093]	[0.026]
Mean of Dep. Variable	.51	.51	.51	.5	.51
Observations	53,370	53,370	53,370	1,752	15,120
			Reversed on Appeal		
Filed After Cutoff	-0.008	-0.001	-0.010	-0.049	-0.011
	[0.005]	[0.008]	[0.010]	[0.059]	[0.014]
Mean of Dep. Variable	.08	.08	.08	.09	.09
Observations	53,370	53,370	53,370	1,077	15,120

This table presents regression discontinuity (RD) estimates of the effect of additional reporting time on various appellate outcomes, including whether an appeal was filed subsequent to an order on the motion, whether the lower-court judgment was affirmed on appeal, and whether the lower-court judgment was reversed. The running variable represents the motion filing date relative to the six-month list eligibility cutoff. Motions filed just before the cutoff are eligible for the current six month list, whereas motions filed just after the cutoff have an additional six months before they might appear on a list. Columns (1)-(3) are estimated parimetrically with linear, quadratic, and cubic polynomials, respectives. Columns (4)-(5) are estimated nonparametrically with local linear regressions, using the IK and CCT methods of optimal bandwidth selection, respectively. All columns include basic case- and motion-level controls, including a dummy for the party (plaintiff or defendant) filing the motion and nature-of-suit, judge, district, and filing-year fixed effects.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A.8: Proportional Hazard Analysis: Effect of Reporting Time on Motion Survival

	(1)	(2)
8-9 Months until Report	0.980**	0.946***
	[0.008]	[0.010]
9-10 Months until Report	0.922***	0.900***
	[0.008]	[0.010]
10-11 Months until Report	0.913***	0.847***
	[0.007]	[0.009]
11-12 Months until Report	0.898***	0.798***
	[0.007]	[0.008]
12-13 Months until Report	0.894***	0.808***
	[0.007]	[0.008]
Observations	420,535	420,212
Survival Model	Cox	Cox
Stratified by NoS, Judge, District, & Filing-Year		Yes
Mean Months Motion Open	6.21	6.21
Mean Reporting Time (months)	10.05	10.05

This table presents hazard ratios for individual reporting month dummies (relative to a baseline hazard rate for motions with fewer than eight months of reporting time). All columns include basic case- and motion-level controls, including calendar day time trends, dummies for the moving party, and a dummy for whether previous summary judgment motions have been filed in the same case. Column (2) is also stratified to allow for independent baseline hazard rates by nature-of-suit, judge, district, and filing-year.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix B

Chapter 2 Appendix

B.1 Theoretical Framework for Public Sector Retention

The goal of this section is twofold. First, we demonstrate the importance of the parameter that we estimate, the differential sensitivity of soldiers to lump-sum bonuses by ability, for capturing how the quality of the military will change with various retention policies. Second, we show that in the simplest model of public sector retention, this key parameter is unambiguously positive – retention policies that increase the financial return should attract higher ability soldiers and increase the average quality of soldiers in the military. However, we show that away from that simple case, the theoretical predictions are ambiguous and depend on the underlying distribution of preferences across the population.

First, we relate the parameter that we estimate in Section 2.4 to the effect of retention policies on the average quality of the military $\overline{A}$, a parameter that analysis in the military is key for designing retention policies. Mechanically, the total quality of retained soldiers is

$$\overline{A} = \sum_{u} p_i(R) * a_i$$

where $p_i(R)$ is the probability that individual i reenlists and a_i is the ability of soldier i. The response of this average to a reenlistment bonus K is

$$\frac{d\overline{A}}{dK} = \sum_{i} \frac{dp_i(R)}{dK} * a_i = \sum_{i} \gamma_i a_i$$

where $\gamma_i = \frac{dp_i(R)}{dK}$. Using expectations, you can rewrite this as:

$$\frac{d\overline{A}}{dK} = \overline{\gamma}\,\overline{a} + cov(\gamma_i, a_i) = \overline{\gamma}\,\overline{a} + \beta Var(a_i)$$

where $\overline{\gamma}$ is the average response of soldiers to the bonus and $\overline{a}$ is average ability in the military. The key parameter that needs to be estimated to inform the effect of retention policies on average soldier quality is β, which is precisely the parameter we focus on estimating in Section 2.4.

Having established the importance of this parameter for the design of retention policies, we now explore a simple model of selection that underpins this parameter. Consider a soldier choosing whether to reenlist in the military for a fixed term. As discussed above, personnel management is notoriously rigid in the military. Although individual ability can indirectly influence compensation – for example, higher ability individuals might be promoted more quickly, entitling them to a steeper wage profile – at least in the short term, military compensation is largely independent of individual ability. Alternatively, in a competitive civilian labor market, higher ability individuals earn their full marginal product. Therefore, in our simplified model, military compensation is independent of individual ability, whereas civilian wages are increasing in ability.

We will write the individual's military payoff as:

$$U_i(military) = W^m(\mathbf{X}_i), \tag{B.1.1}$$

where W^m is the military wage function and $\mathbf{X}$ is a vector of individual characteristics affecting compensation (for example, rank, years of service, and military occupational specialty). Should she choose not to reenlist, the same individual earns a payoff of:

$$U_i(civilian) = W^c(\mathbf{X}_i, a_i), \tag{B.1.2}$$

where W^c is the civilian wage function, and a reflects individual ability, and $\frac{\partial W^c}{\partial a} \geq 0$.

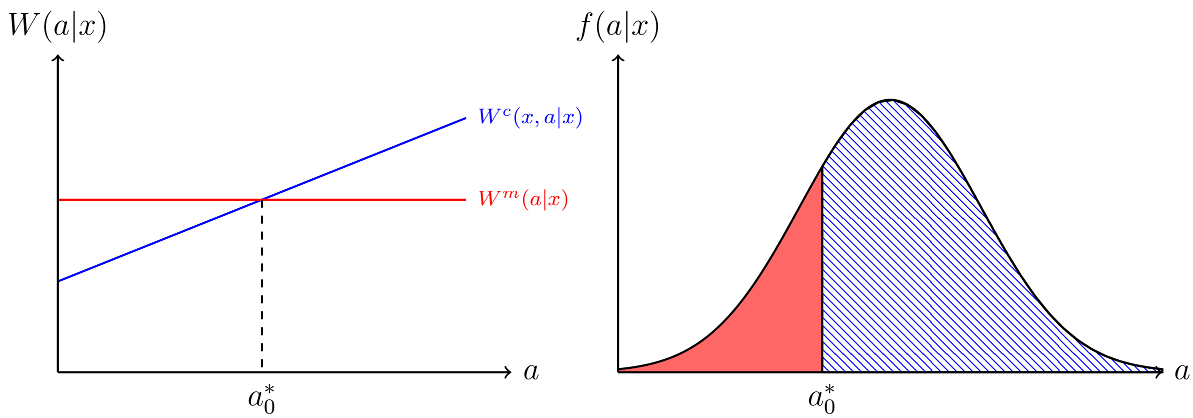
Figures B-1a and B-1b depict the military and civilian wage functions and the distribution of ability types, respectively. In this setting, there exists a threshold ability type a_0^*, such that soldiers of ability $a_i < a_0^*$ will always choose to reenlist, and soldiers of ability $a_i > a_0^*$ will always choose to separate from the military.

Now suppose that the military wants to attract more workers and therefore offers a lump-sum reenlistment bonus of K. The new military payoff is:

$$U_i(military) = W^m(\mathbf{X}_i) + K \tag{B.1.3}$$

Figure B-2a depicts the military and civilian wage functions subsequent to the level shift in military wage. As illustrated by the figure, a level shift in the military wage generates a corresponding increase in the threshold ability type, a_1^*.

Figure B-1: Simple Case



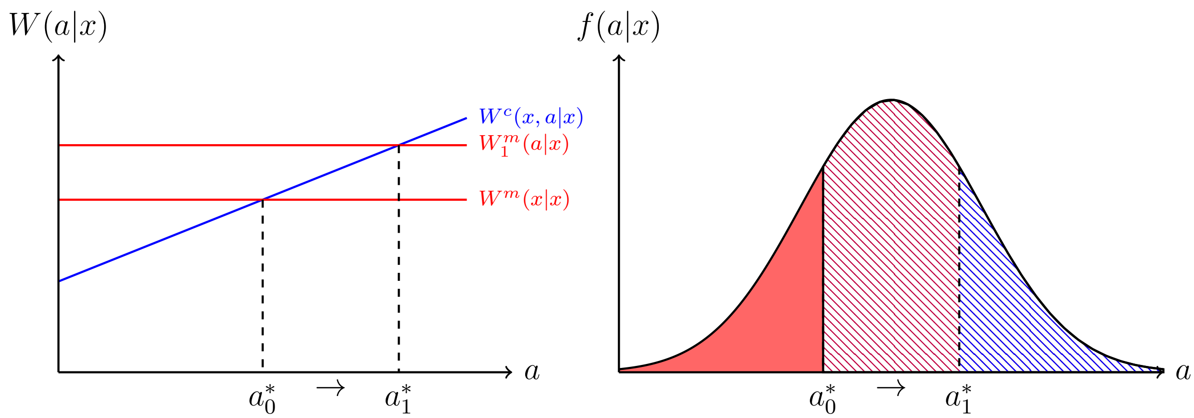
(a) Military and Civilian Wage Functions

(b) PDF of Soldier Ability

Intuitively, as military wages increase, the military will tend to retain more service members. Only the most productive soldiers will be able to command a comparable wage in the civilian labor market. Figure B-2b depicts the new cutoff rule. In this simple case, an increase to the relative military payoff generates an increase in the marginal ability type a^*, and implies that higher ability soldiers are more responsive to reenlistment bonuses than their lower-ability peers. It is only the higher-ability workers who are on the margin and thus affected by lump-sum bonuses. It also increases the average ability of the soldiers who the military retains, which is likely a key statistic that the policy-maker cares about.

While this simple model generates an unambiguous counterfactual prediction, a setting with richer soldier heterogeneity will produce theoretically ambiguous responses. Suppose that soldiers have heterogeneous "taste" for military service c_i drawn from a continuous distribution $F(\cdot)$. In particular, rewrite the military

Figure B-2: Exogenous Shift in Relative Military Compensation



(a) Military and Civilian Wage Functions

(b) PDF of Soldier Ability

payoff function as

$$U_i(military) = W^m(\mathbf{X}_i) + c_i, \tag{B.1.4}$$

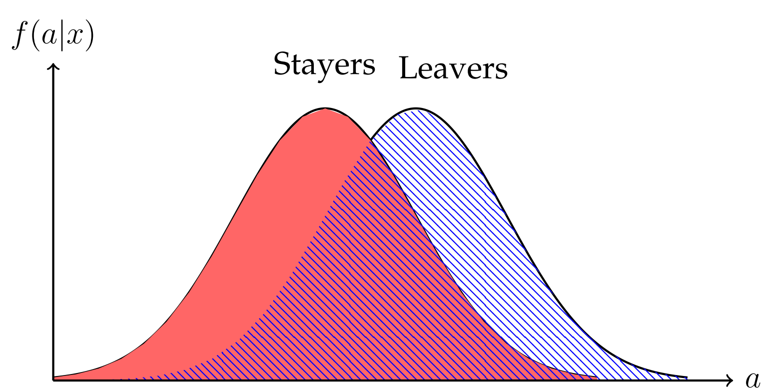
Given heterogenous taste for service, a soldier i reenlists if her military payoff exceeds her civilian payoff, or $W^m(\mathbf{X}_i) + c_i > W^c(\mathbf{X}_i, a)$. This yields a cutoff rule for the soldier's reenlistment decision with respect to ability type a_i. Namely, conditional on individual characteristics $\mathbf{X}$, a soldier reenlists if

$$a_i < g(c_i), \tag{B.1.5}$$

where $g(c_i) = W^{c^{-1}}\left(W^m(X_i) + c_i\right)$ and $g'(c_i) > 0$.

Figure B-3 depicts stylized baseline ability distributions of stayers and leavers in this continuous-type setting. As Equation (B.1.5) demonstrates, conditional on a soldier's taste for the military (c_i), the sorting of stayers and leavers looks identical to our simple case in Figure B-1b. However, in the continuous-type setting, we have to aggregate across values of taste-for-service types c_i in order to obtain the full distribution of ability types among either stayers or leavers. In other words, we obtain the "stayer" distribution in Figure B-3 by adding up the areas left of the cutoff value $g(c_i)$ for each taste-for-service type c_i. Consistent with the preliminary prediction that those who reenlist are of lower average ability than those who do not reenlist, we draw the PDFs so that the stayer ability distribution peaks to the left of the leaver ability distribution. In this more general case, there are many ability types for which soldiers will either reenlist *or* separate, depending upon their individual taste for service. Stayers on the far right-hand tail of their ability distribution – that is, those who reenlist despite highly marketable private-sector job skills – have a very high taste for military service. Conversely, leavers on the far left-hand tail of their ability distribution – that is, those who separate from the military despite relatively low private-sector job skills – have a very low taste for military service.

Figure B-3: Stayer and Leaver Ability Distributions,
Continuous Taste Types (c_i)

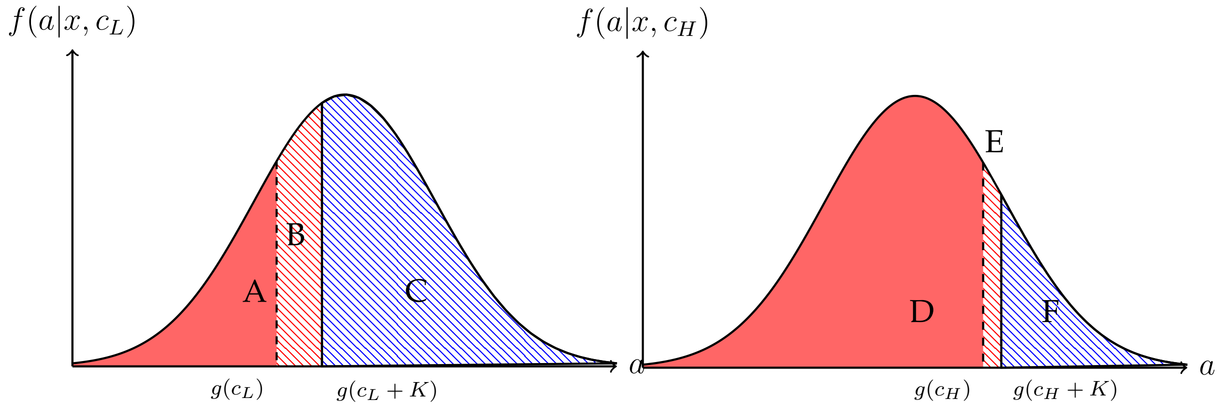


Now consider the introduction of lump-sum bonuses K, again in the form of a positive level shift in the military wage, so that the military payoff is $W^m(\mathbf{X}_i)+c_i+K$. Under the new cutoff rule, a soldier reenlists if $a_i < g(c_i + K)$. Conditional on taste for service, the stark predictions depicted in Figure B-2 from the simple case still hold. That is, for each value of c_i, an increase to the relative military payoff generates an increase in the marginal ability type a^* and increases in the average abilities of those who chose to reenlist. However, in aggregating the changes across soldier types, the predictions for how soldiers of different abilities respond to the bonus become ambiguous. What the differential elasticity to bonuses by ability will be will depend upon at least three factors: 1) the shape of the function $g(\cdot)$ (which incorporates both how individuals trade off taste for military service with other types of compensation and how civilian employers reward ability), 2) the density of the ability distribution around cutoff values and 3) the correlation between ability a and taste for service c.[1]

To fix intuitions, suppose there are just two types of taste for military service, $c_i \in \{c_L, c_H\}$, denoting either a low or high taste for military service. Figure B-4a shows the new cutoff rule after the bonus K for individuals with a low taste for service c_L, and Figure B-4b shows the new cutoff rule for individuals with a high taste for service c_H. Soldiers in areas A and D were always going to reenlist in the military, and soldiers in areas C and F were never going to reenlist. Areas B and E, on the other hand, correspond to soldiers who were induced to stay in the military due to the change in the compensation policy. The estimated differential response to the bonuses by ability will depends on the size and placement of these

[1] In the dynamic version of this static problem where soldiers consider the expected future stream of compensation, this would also depend on the correlation between discount factors and ability a.

Figure B-4: Change in Relative Return to Military Service, Two-Type Case



(a) Low Taste for Military Service (c_L)

(b) High Taste for Military Service (c_H)

two areas. Specifically, the size of area B and E is going to depend on the distance between $g(c_L)$ and $g(c_L + K)$ or between $g(c_H)$ and $g(c_H + K)$. This is determined by the shape of the g function. The size of area B and E is also going to depend on the density of soldiers around these cutoffs (i.e the height of the distribution). Affecting parts of the ability distribution where there are more soldiers will have a bigger effect on the average quality of the group. Even in this simple two-type case, without further assumptions, there is no clear prediction for whether higher or lower skill soldiers will be more responsive to reenlistment bonuses. In this simple model, our empirical finding that lower ability soldiers are more responsive to these lump-sum bonuses corresponds to the case where B is larger than E.

B.2 Data Appendix

B.2.1 Data Details

Reenlistment Data

The data for this analysis comes from the U.S. Army's Total Army Personnel Database (TAPDB), from which we have constructed a panel of enlistment spells between 1992 and 2016. We exclude from the analysis all current spells. For our analysis, the date of entry into the military is identified for each soldier according to the first month in which they received payments. This captures military service that the soldier may have preformed in the past either in nonconsecutive spells or in other branches of the military. We drop all observations where we observe only 1 spell for the soldier that is less than 3 months. These spells are likely soldiers who did not complete basic training. We also drop spells that are are the end of the soldier's tenure, are less than 3 months, and result in the soldier entering officer training. We code that soldier as reenlisting in our analysis.

In addition to making the choice of whether to reenlist at the end of their spell, some soldiers have the option of extending their contract by up to a year. We identify spells as extension if the entry date of the spell is the same as the extension date of the previous spell. Since we are interested in major reenlistment decisions, we absorb all extensions into the previous spell. For example, if a soldier served for 3 years and extended their spell for 1 year, but then left the military, we code the soldier as having 1 four year spell and then choose not to reenlist. The left panel of Figure B-5 shows the distribution of spell length in the resulting sample, and the right panel of Figure B-5 shows the distribution of enlistment terms in our sample.

In addition to knowing the date at which the soldier decided to reenlist and the date at which the term of service was due to end, we need to identify the date at which the soldier entered the reenlistment window. We use this date to assign the unemployment rate and SRB offer that the soldier faces. When in the reenlistment window the soldier decided to reenlist is the soldier's choice, and we want to abstract from variation in the relative military wage that are the result of strategic timing of the market. For each fiscal year, the Army announces in MILPER messages the date at which the soldier is eligible to enter their reenlistment window. Before fiscal year 2007, soldiers entered their reenlistment window 12 months before the end of their contracted service. However, for 2007, 2008 and 2009, the army extended this to 24 months. In the following years, all soldiers with terms expiring in the following year became eligible for reenlistment window on a given date. Figure B-6 plots the distribution of the number of months in advance the end of service (ETS) date that the soldier enters their reenlistment window. Most soldiers enter 12 months in advance, with additional masses at 15 and 24 months. Most soldiers also reenlist at some point in that window.

We use two main measures of soldier quality throughout our analysis: the soldier's AFQT score at entry and the number of months in their first term that the soldier spends below Sergeant (E-5). Table B.1 shows estimates from Wigdor and Green (1991) showing that AFQT score are highly correlated with within-military hands on performance metrics. Figure B-7 also shows that AFQT scores are highly predictive of being promoted quickly within the military. We chose the number of months below sergeant as our measure of military performance because it is highly correlated with future performance in the military. Table B.2 shows the pairwise correlations for the number of months that it takes soldiers to get to each rank. The speed of promotion to E-3 or E-4 is not highly correlated with strong performance later in the soldier's career, as those promotions are more defaulted, so we use the speed of promotion to E-5.

Credit, GI Bill, and Thrift Savings Plan (TSP) Data

Credit data were obtained from a major credit reporting agency, which we then matched with the TAPDB enlistment database. Credit data consists of a panel of twice-annual observations for soldiers with service between April 2007 and March 2015. Among soldiers who were eligible for at least one reenlistment during that time period, we are able to match nearly 90% to credit reporting data. For each soldier facing a reenlistment choice, we match the soldier to her credit report that is closest in time to the beginning of her reenlistment window. In addition to individual credit scores, we observe open lines of credit, balances, and delinquencies, grouped by major lending categories. For simplicity, we focus our analysis on credit scores, but we have confirmed that our results are largely robust to proxying for credit constraints with past delinquencies.

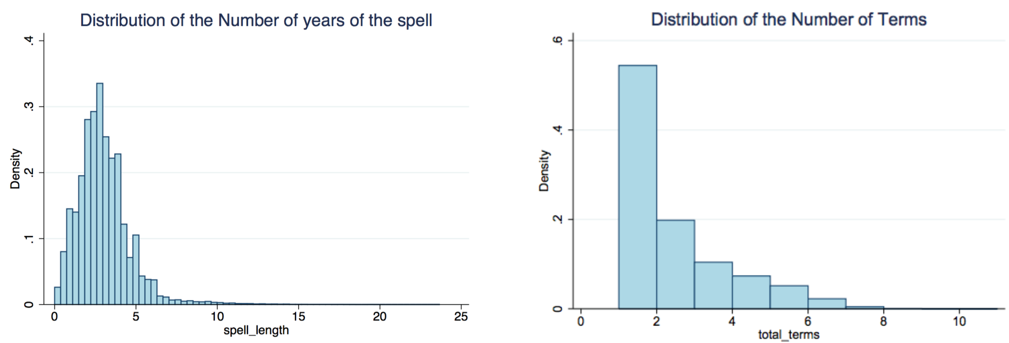
GI Bill data are directly observable within the TAPDB enlistment database. Immediately upon enlistment, soldiers who meet minimum eligibility requirements are offered the opportunity to enroll in the Montgomery GI Bill (MGIB) benefits package. In order to enroll, a soldier must consent to having $1,200 deducted from her military pay, usually in equal $100 deductions from her first twelve monthly pay checks. Under 2016 rates, soldiers who enrolled in the basic MGIB package were eligible to receive up to $66,852 in total educational benefits (up to $1,857 per month for 36 total months of higher education). Soldiers who enroll in the MGIB are given the further opportunity to participate in the MGIB "buy-up" by consenting to an additional deduction of between $20 and $600. Soldiers who participate in the full $600 buy-up become eligible to receive up to $5,400 in total MGIB educational benefits ($150 per month for 36 months). In our data we observe whether a soldier is eligible to enroll in the standard MGIB benefits package and whether she actually enrolls, as well as the amount of her total accrued MGIB contributions. We code soldiers as having participated in the buy-up when they have contributed a total of $1,800 towards the MGIB (the basic $1,200 contribution plus the full $600 buy-up contribution). Among our soldiers in our baseline sample, more than 93.3% enrolled in the basic MGIB, and among those, 3.3% participated in the full $600 buy-up.

Thrift Savings Plan (TSP) contribution data are also directly observable within the TAPDB enlistment database. The TSP is a 401(k)-like retirement savings plan available to many federal workers. First established for civilian workers in 1986, members of the military became eligible for the TSP in 2001. For each spell, we observe the soldier's total contribution to her TSP account. We also observe her total base military pay over the course of her spell, which we use to calculate her TSP contribution as a share of her total basepay. We also create an indicator variable for whether a soldier has made any contribution greater than zero to her TSP account over the course of her spell. Among enlistment spells since 2001, approximately 32% of soldiers make some positive contribution to the TSP, and the average contribution (as a share of total spell base military pay) is approximately 2.2%.

Appendix Table B.6 shows pairwise correlations between credit score, basic MGIB enrollment, participation in the MGIB buy-up, participation in the TSP (i.e., any contribution), and total TSP contributions as a share of the servicemember's military pay. Credit scores are positively correlated with ability measures, as are participation in the MGIB buy-up and participation in the TSP. Enrollment in the basic MGIB is slightly negatively correlated with both of our ability measures.

Data Appendix Tables and Figures

Figure B-5: Distribution of the Number of Terms amoung Enlisted Soldiers (1992-2017)



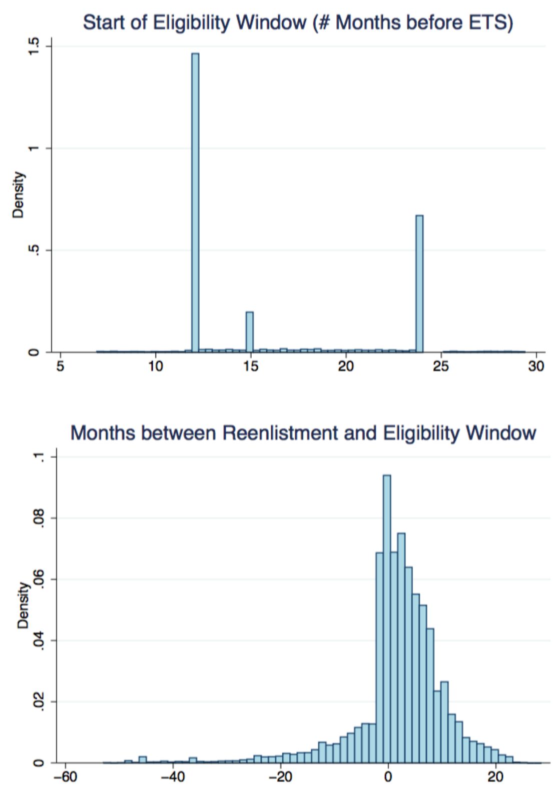
Notes: Sample includes all enlisted soldiers from 1992-2016 and excludes soldiers currently serving in the Army.

Table B.1: Correlations of Armed Forces Qualifications Test (AFQT) and Job-Specific Hands-On Performance Measure

Specialty	AFQT w/ Performance
Administrative specialist	0.35
Air traffic control operator	0.10
Rifleman	0.40
Machinegunner	0.49
Mortarman	0.33
Motor transport operator	0.24
Radio operator	0.22
Median Correlation	0.26

Source: Wigdor and Green (1991), Table 8-10.

Figure B-6: The Timing of Reenlistment Decisions and the Eligibility Window



Notes: Sample includes all enlisted soldiers from 1992-2016 and excludes soldiers currently serving in the Army. The left panel plots the distribution of the time between the beginning of the reenlistment window and the end of the soldier's term The right panel plots the distribution of the difference between the start of the reenlistment window and the date that the soldier actually reenlists.

Figure B-7: The Correlation of AFQT scores and speed of promotion

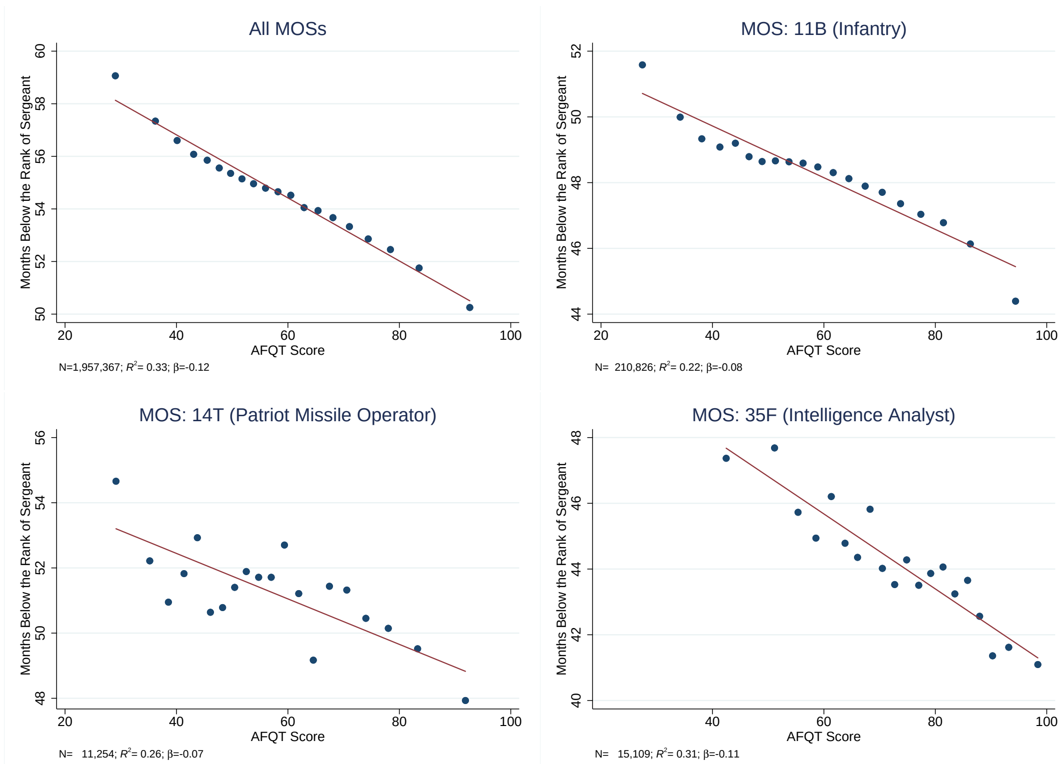


Table B.2: Correlation of Promotion Speeds Across Ranks

	(1)						
	Time to E-2	Time to E-3	Time to E-4	Time to E-5	Time to E-6	Time to E-7	Time to E-8
Time to E-2	1						
Time to E-3	0.758***	1					
Time to E-4	0.598***	0.686***	1				
Time to E-5	0.0764***	0.128***	0.298***	1			
Time to E-6	0.0526***	0.0876***	0.213***	0.620***	1		
Time to E-7	0.0812***	0.112***	0.241***	0.565***	0.803***	1	
Time to E-8	0.112***	0.144***	0.256***	0.505***	0.653***	0.774***	1

Notes: Sample includes all enlisted soldiers from 1992-2016. Correlations are pairwise.* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B.3: Correlation of SRB offers Across Chosen Reenlistment Term

	(1) 4 Year Term
2 Year Term	0.593***
3 Year Term	0.986***
5 Year Term	0.988***
6 Year Term	0.964***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Notes: Sample includes all SRB offers from 1997-2016. Correlations are pairwise.

Table B.4: Correlations of Unconditional and Conditional (Location-Specific) SRB Offers (4-year terms)

	(1) Regular Offer
Continental US 1	0.372***
Continental US 2	0.510***
Continental US 3	0.585***
Continental US 4	0.698***
Continental US 5	0.722***
Continental US 6	0.846***
Continental US 7	0.831***
Non-continental 1	0.586***
Non-continental 2	0.608***

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$
Notes: Sample includes all SRB offers from 1997-2016. Correlations are pairwise.

Figure B-8: Continuation Profiles by AFQT Score

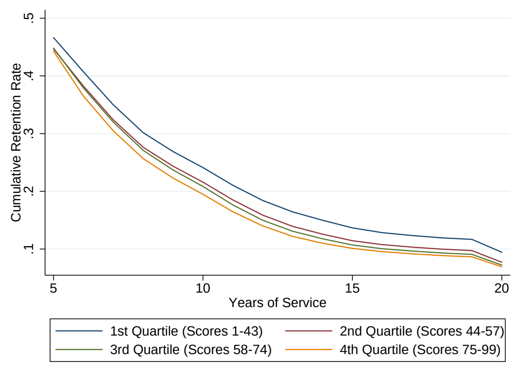


Table B.5: Eligibility for Early Retirement Programs

	Panel A: TERA Program		
	All	15+ Years	Within 1 year
	Soldiers	of Service	of cutoff
Total Soldiers	259,998.00	25,441.00	3,114.00
Eligible Soldiers	1,731.00	1,731.00	1,731.00
Fraction Eligible for TERA	0.67	6.80	55.59
	Panel B: VSI/SSB Program		
	All Soldiers	6+ YOS	–
Total Soldiers	194,017.00	62,420.00	–
Eligible Soldiers	7,326.00	7,326.00	
Fraction Eligible for VSI	3.78	11.74	

Notes: In Panel A, Column 1 includes sample is all enlisted solders serving in the military on August 31, 1994, the start date for the TERA program. In Panel A Column 2, the sample is restricted to those with at least 15 years of service. In Column 3, the sample is restricted to those in eligible occupations and ranks with service that puts them within 1 year of eligibility. In Panel B, Column 1 includes all enlisted soldiers serving in August 1, 1993, the start date of the VSI program. Column 2 further restricts the sample to those soldiers with at least 6 years of service.

Table B.6: Pairwise Correlations Between Ability Measures (AFQT and Months Below Sergeant)
and Credit Score, MGIB Participation, and TSP Participation

Specialty	AFQT	Months Below Sergeant
Credit Score	0.21	-0.14
MGIB Enrollment	-0.06	0.01
MGIB Buy-up	0.07	-0.04
Any TSP Contribution	0.09	-0.03
% TSP Contribution	0.12	-0.04

B.2.2 Case Studies: Time Series Variation in SRBs

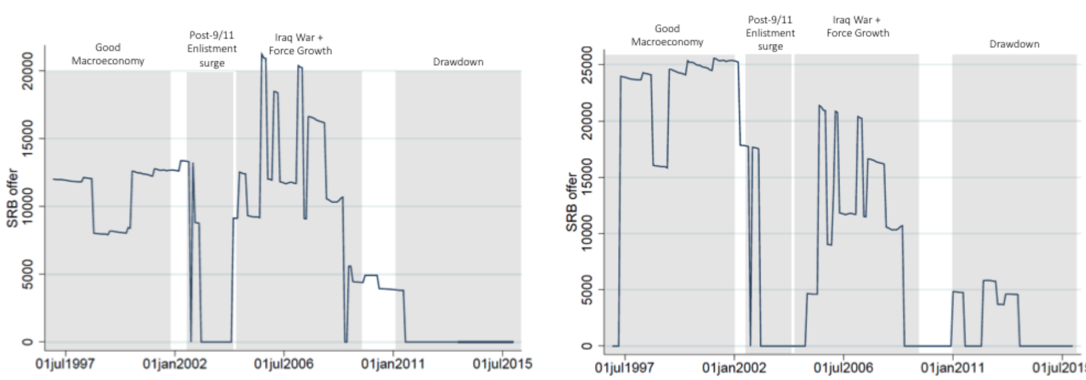
While it is difficult to know precisely what drives the high-frequency variation in SRB offers, anecdotal and observational evidence suggests that variation in SRBs is driven largely by a combination of "inside" factors – namely, the military's operational and strategic requirements – and "outside factors" – namely, labor market conditions and other economic trends affecting civilian labor market opportunities.[2] We study how these factors may have driven time-series variation in SRB offers across two separate MOSs in Figure B-9. The left-most panel plots the time series of SRB offers for infantrymen. This MOS is the largest in the Army (11% of our sample) and is the most representative of the Army as a whole. Infantry SRBs remained moderately high throughout the period preceding the September 11, 2001 attacks. Although operational requirements were relatively minimal during this period, pre-war SRBs might reflect positive macroeconomic conditions, which forced the military to compete with civilian employers for qualified workers. Infantry SRBs dipped dramatically in early 2002 and remained low throughout much of the 2002-2004 period. This was a period of surging enlistment, which many attribute to heightened patriotism in the aftermath of the 9/11 attacks. However, SRBs increased again in 2004, and despite considerable volatility, they remained high through approximately 2008, reflecting the military's growing operational requirements in Iraq and Afghanistan. Though we might be concerned that this period also had higher casualties than other periods (a negative job amenity), we control for month fixed effects in all regressions and occupation by month fixed effects in others. Infantry SRBs have remained low

[2] While the Army process does not directly measure civilian economic opportunities, they do track the personnel inventories and adjust SRBs accordingly. So current labor market conditions may affect individual choices regarding reenlistment, which then affect the *future* SRBs offered to service members to maintain desired personnel inventories.

since approximately 2011, likely reflecting the military's gradual exit from Iraq and its overall drawdown of personnel.

In contrast to infantry SRBs, SRB offers for Patriot missile operators, plotted in the right panel of Figure B-9, appear to be largely driven by operational requirements and large-scale changes to the Army's overall force structure. SRB offers to Patriot missile operators were highest between 1997 and 2002 – precisely the period during which the Army was expanding its number of Patriot missile battalions from 13 to 15. The Army's focus on Patriot missiles was likely influenced by a period of perceived threat by Iraqi Scud missiles, against which Patriot missiles were intended to defend. The Patriot missile operator SRBs illustrate how exogenous changes in Army force structure – due to the standing-up of a new unit or perhaps the introduction of new military technology – can be an important driver of variation in SRBs over time.

Figure B-9: Selective Reenlistment Bonus (SRB) Case Studies
SRB offers by MOS (E-4), 1997-2015

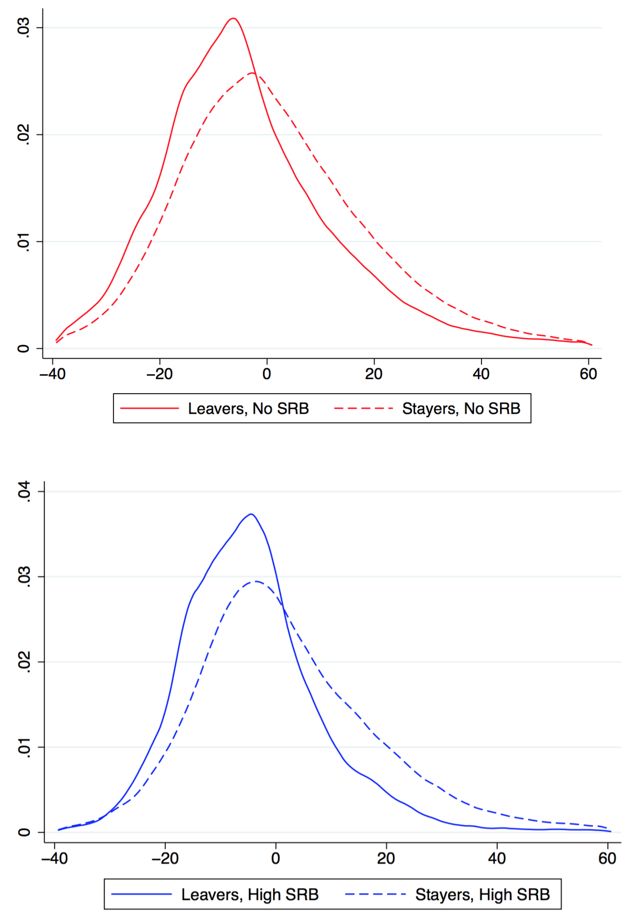


(a) MOS: 11B (Infantry)

(b) MOS: 14T (Patriot Missile Operator)

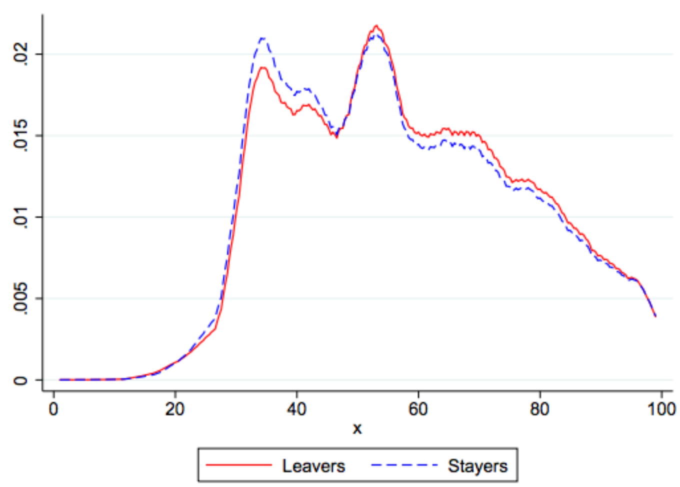
B.3 Robustnes of Empirical Results

Figure B-10: The distribution of first term promotion speeds, split by reenlistment decisions.



Notes: The figure plots the residuals of a regression of the number of months the soldier spent below sergent (rank E4 of below) on MOS*rank*YOS dummies as well as date dummies. The sample includes those soldiers who have a choice to reenlist. The left panel plots the distributions for the set of soldiers who do not have a SRB available at the start of their reenlistment window. The right panel shows the distributions for the set of soldiers who have an offered SRB of at least $8,000. The left figure includes 1.7 million observations (75% of the sample) while the right panel includes 300,000 observations (13% of the sample). Each distribution is truncated at the top and bottom 1%.

Figure B-11: The raw distribution of AFQT scores for soldiers, split by reenlistment decisions.



Notes: The figure plots the raw AFQT score distribution for soldiers by their reenlistment decision. The sample includes those soldiers who have a choice to reenlist.

Table B.7: Soldier's Reenlistment Probabilities by AFQT Score and Offered Bonuses (SRBs): Alternate Specifications

*Dependent Variable: Indicator for Reenlisting*100*						
	(1)	(2)	(3)	(4)	(5)	(6)
			Subsamples			
			Main MOS	High-Corr.	No Surge	Positive
	Baseline	SRB in logs	only	MOS only	Years	SRB Offer
SRB	0.615***		0.465**	0.600***	0.527***	0.216**
	(0.078)		(0.207)	(0.221)	(0.076)	(0.109)
SRB*AFQT	-0.710***		-0.646*	-0.574	-0.648***	-0.224*
	(0.116)		(0.335)	(0.366)	(0.113)	(0.117)
AFQT	-9.347***		-11.889***	-10.195***	-9.201***	-17.428***
	(0.868)		(1.950)	(2.765)	(0.938)	(1.669)
log(SRB)		0.752***				
		(0.098)				
log(SRB)*AFQT		-0.850***				
		(0.184)				
R-squared	0.157	0.157	0.127	0.142	0.155	0.114
Observations	1761615	1761615	627775	382301	1457868	516754
Year * Month FE	x	x	x	x	x	x
MOS*Rank*YOS FE	x	x	x	x	x	x
Demographic Controls	x	x	x	x	x	x
Average Dep. Var	65.1	65.1	63.92	63.25	63.92	66.35
Average SRB	2.89	2.66	2.72	3.5	2.72	9.86

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015.Demographic controls include gender, age, marital status, race, and special skill dummies. SRBs are in $1000s of 2015 dollars and AFQT is on a scale from 0-1. The ``main MOS only" column restricts to the 10 largest occupations in our sample. The ``high corr. mos" column restricts to MOSs identified by Wigdor and Green (1991) as exhibiting a high correlation between AFQT score and hands-on job performance. The ``no surge years" specification excludes soldiers entering their reenlistment window during the Iraq qurve years (2007-2009). The ``positive SRB offer" column incldues only soldiers who were offered a positive SRB.

Table B.8: Soldier's Reenlistment Probabilities by Months E-4 or Below and Offered Bonuses (SRBs): Alternate Specifications

*Dependent Variable: Indicator for Reenlisting*100*					
	(1)	(2)		(3)	(4)
		Subsamples			
	Baseline	Main MOS only	High-Corr. MOS only	No Surge Years	Positive SRB Offer
SRB	-0.607***	-0.877***	-0.952***	-0.672***	0.461***
	(0.108)	(0.295)	(0.326)	(0.116)	(0.111)
SRB*Months E4 or Below	0.015***	0.019***	0.024***	0.016***	-0.009***
	(0.002)	(0.005)	(0.005)	(0.002)	(0.002)
Months E4 or Below	0.309***	0.334***	0.273***	0.288***	0.599***
	(0.024)	(0.054)	(0.067)	(0.024)	(0.030)
log(SRB)					
log(SRB)*AFQT					
R-squared	0.171	0.150	0.155	0.167	0.150
Observations	1708425	619066	376659	1403790	522354
Year * Month FE	x	x	x	x	x
MOS*Rank*YOS FE	x	x	x	x	x
Demographic Controls	x	x	x	x	x
Average Dep. Var	66.3	65.24	63.92	65.24	66.4
Average SRB	3.02	2.86	3.59	2.86	9.86

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. Demographic controls include gender, age, marital status, race, and special skill dummies. SRBs are in $1000s of 2015 dollars. ``Months E4 or Below" is defined as the number of months spent in a rank below Sergeant during the soldier's first enlistment. The ``main MOS only" column restricts to the 10 largest occupations in our sample. The ``high corr. mos" column restricts to MOSs identified by Wigdor and Green (1991) as exhibiting a high correlation between AFQT score and hands-on job performance. The ``no surge years" specification excludes soldiers entering their reenlistment window during the Iraq qurve years (2007-2009). The ``positive SRB offer" column incldues only soldiers who were offered a positive SRB.

Table B.9: Selective Reenlistment Bonuses (SRBs) and Average AFQT: Alternate Specifications

Dependent Variable: AFQT Score Percentile									
	(1)	(2)	(3)	(4)	(5)		(6)	(7)	(8)
					Subsamples				IV Spec
		CZ	MOS		Main MOS	High-Corr.	No Surge	Positive SRB	
	Baseline	Trends	Trends	SRB in Logs	only	MOS only	Years	Offers	Actual SRBs
SRB*Stay	-0.048***	-0.056***	-0.018		-0.166***	-0.059	-0.042**	-0.020	-0.038
	(0.015)	(0.016)	(0.012)		(0.035)	(0.056)	(0.019)	(0.020)	(0.065)
SRB*Leave	0.066***	-0.000	0.108***		-0.047	0.074	0.061**	0.024	
	(0.022)	(0.023)	(0.016)		(0.063)	(0.071)	(0.024)	(0.020)	
log(SRB)*Stay				-0.064***					
				(0.016)					
log(SRB)*Leave				0.087***					
				(0.028)					
Stay	-1.216***	-1.817***	-1.183***	-1.132***	-1.544***	-1.250***	-1.173***	-2.195***	
	(0.118)	(0.089)	(0.121)	(0.124)	(0.248)	(0.381)	(0.126)	(0.279)	
R-squared	0.304	0.351	0.326	0.304	0.251	0.226	0.302	0.313	0.290
Observations	1761615	1422783	1757584	1761615	627775	382301	1457868	516754	913070
Year * Month FE	x			x	x	x	x	x	x
Year * Month * CZ FE		x							
Year * Month * MOS FE			x						x
MOSxRankxYOS FE	x	x	x	x	x	x	x	x	x
Demographic Controls	x	x	x	x	x	x	x	x	x
Mean Dep. Var	58.26	59.08	58.25	58.26	54.83	59.83	58.17	61.17	56.61
Mean SRB	2.89	3.26	2.9	2.66	2.96	3.5	2.72	9.86	3.36

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. SRBs are in $1000 of 2015 dollars. Demographic controls include gender, age, marital status, race, and special skill dummies. The dependent variable is a soldier's AFQT score. AFQT is on a scale from 0-100. The ``main MOS only'' column restricts to the 10 largest occupations. The ``high corr. mos'' column restricts to MOSs identified by Wigdor and Green (1991) as exhibiting a high correlation between AFQT score and hands-on job performance. The ``no surge years'' specification excludes soldiers entering their reenlistment window during the Iraq qurve years (2007-2009). The ``positive SRB offer'' column incldues only soldiers who were offered a positive SRB. The ``IV Specification'' restricts to only those who chose to reenlist and uses the offered SRB as an instrument for the actual SRB offer that the soldier receives. The first stage F-statistic for the IV regression is 460.

Table B.10: Selective Reenlistment Bonuses (SRBs) and Average Months Below Sergeant: Alternate Specifications

Dependent Variable: Months E4 or Below	(1)	(2)	(3)	(4)		(5)	(6)	(7)
				Subsamples				IV Spec
	Baseline	CZ Trends	MOS Trends	Main MOS only	High-Corr. MOS only	No Surge Years	Positive SRB Offers	Actual SRBs
SRB*Stay	0.022	-0.007	0.051*	0.073	0.054	0.021	0.013	0.076
	(0.030)	(0.031)	(0.030)	(0.091)	(0.106)	(0.032)	(0.029)	(0.082)
SRB*Leave	-0.076***	0.037	-0.055	-0.079	-0.134	-0.116***	0.428***	
	(0.027)	(0.023)	(0.043)	(0.061)	(0.086)	(0.028)	(0.052)	
log(SRB)*Stay								
log(SRB)*Leave								
Stay	6.693***	8.416***	6.652***	7.115***	5.396***	5.974***	13.041***	
	(0.505)	(0.451)	(0.529)	(1.074)	(1.132)	(0.492)	(0.962)	
R-squared	0.342	0.391	0.361	0.305	0.316	0.336	0.327	0.343
Observations	1708425	1433249	1704497	619066	376659	1403790	522354	897384
Year * Month FE	x			x	x	x	x	x
Year * Month * CZ FE		x						
Year * Month * MOS FE			x					x
MOSxRankxYOS FE	x	x	x	x	x	x	x	x
Demographic Controls	x	x	x	x	x	x	x	x
Mean Dep. Var	54.4	53.1	54.4	53.57	52.16	54.41	51.93	58.77
Mean SRB	3.02	3.27	3.02	3.04	3.59	2.86	9.86	3.46

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. SRBs are in $1000 of 2015 dollars. Demographic controls include gender, age, marital status, race, and special skill dummies. The dependent variable ``Months E4 or Below" is defined as the number of months spent in a rank below Sergeant during the soldier's first enlistment. The ``main MOS only" column restricts to the 10 largest occupations. The ``high corr. mos" column restricts to MOSs identified by Wigdor and Green (1991) as exhibiting a high correlation between AFQT score and hands-on job performance. The ``no surge years" specification excludes soldiers entering their reenlistment window during the Iraq qurve years (2007-2009). The ``positive SRB offer" column incldues only soldiers who were offered a positive SRB. The ``IV Specification" restricts to only those who chose to reenlist and uses the offered SRB as an instrument for the actual SRB offer that the soldier receives. The first stage F-statistic for the IV regression is 460.

Table B.11: Selective Reenlistment Bonuses (SRBs) and Average AFQT: Alternative SRB Offer Windows

Dependent Variable: AFQT Score Percentile						
	(1)	(2)	(3)	(4)		(5)
		Alternative SRB Offer Windows				
	Baseline	6-mo. Avg. SRB	12-mo. Avg. SRB	6-mo. Max. SRB	12-mo. Max. SRB	Final SRB Offer
SRB*Stay	-0.048***	-0.061***	-0.072***	-0.055***	-0.059***	-0.063***
	(0.015)	(0.017)	(0.018)	(0.014)	(0.014)	(0.011)
SRB*Leave	0.066***	0.055**	0.044*	0.055***	0.050**	-0.001
	(0.022)	(0.024)	(0.025)	(0.021)	(0.021)	(0.015)
Stay	-1.216***	-1.227***	-1.247***	-1.189***	-1.163***	-1.530***
	(0.118)	(0.118)	(0.118)	(0.119)	(0.119)	(0.117)
R-squared	0.304	0.304	0.304	0.304	0.304	0.304
Observations	1761615	1761615	1761615	1761615	1761615	1761615
Year * Month FE	x	x	x	x	x	x
MOS*Rank*YOS FE	x	x	x	x	x	x
Demographic Controls	x	x	x	x	x	x
Average Dep. Var	58.26	58.26	58.26	58.26	58.26	58.26
Average SRB	2.89	2.71	2.53	3.21	3.45	.4

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015.Demographic controls include gender, age, marital status, race, and special skill dummies. SRBs are in $1000s of 2015 dollars and AFQT is on a scale from 0-100. The ``Baseline'' column uses soldiers' highest SRB offer on the first day of their reenlistment eligibility window. The ``6-mo. Avg.'' column uses the average of the high SRB offer on the first day of the first six months of a soldier's reenlistment eligibility window. The ``12-mo. Avg.'' column averages the high SRB offers across the first 12 months of the soldier's reenlistment eligibility window. The ``6-mo. Max.'' column uses the highest SRB offer from the first six months of the reenlistment eligibility window. The ``12-mo. Max.'' column uses the highest SRB offer from the first 12 months of the reenlistment eligibility window. The ``Final SRB Offer'' uses the highest SRB offer available on the last day of a soldier's reenlistment eligibility window, which is generally 90 days prior to the end of the soldier's current enlistment.

Table B.12: Selective Reenlistment Bonuses (SRBs) and Average Months Below Sergeant: Alternative SRB Offer Windows

Dependent Variable: AFQT Score Percentile						
	(1)	(2)	(3)	(4)		(5)
		Alternative SRB Offer Windows				
	Baseline	6-mo. Avg. SRB	12-mo. Avg. SRB	6-mo. Max. SRB	12-mo. Max. SRB	Final SRB Offer
SRB*Stay	0.022	0.012	0.003	0.019	0.017	0.010
	(0.030)	(0.031)	(0.031)	(0.028)	(0.026)	(0.016)
SRB*Leave	-0.076***	-0.083***	-0.089***	-0.077***	-0.080***	-0.016
	(0.027)	(0.030)	(0.033)	(0.027)	(0.027)	(0.022)
Stay	6.693***	6.718***	6.742***	6.668***	6.637***	6.987***
	(0.505)	(0.506)	(0.507)	(0.508)	(0.512)	(0.475)
R-squared	0.342	0.342	0.342	0.342	0.342	0.342
Observations	1708425	1708425	1708425	1708425	1708425	1708425
Year * Month FE	x	x	x	x	x	x
MOS*Rank*YOS FE	x	x	x	x	x	x
Demographic Controls	x	x	x	x	x	x
Average Dep. Var	54.4	54.4	54.4	54.4	54.4	54.4
Average SRB	3.02	2.83	2.64	3.35	3.6	.42

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015.Demographic controls include gender, age, marital status, race, and special skill dummies. SRBs are in $1000s of 2015 dollars, and the dependent variable ``Months E4 or Below'' is defined as the number of months spent in a rank below Sergeant during the soldier's first enlistment. The ``Baseline'' column uses soldiers' highest SRB offer on the first day of their reenlistment eligibility window. The ``6-mo. Avg.'' column uses the average of the high SRB offer on the first day of the first six months of a soldier's reenlistment eligibility window. The ``12-mo. Avg.'' column averages the high SRB offers across the first 12 months of the soldier's reenlistment eligibility window. The ``6-mo. Max.'' column uses the highest SRB offer from the first six months of the reenlistment eligibility window. The ``12-mo. Max.'' column uses the highest SRB offer from the first 12 months of the reenlistment eligibility window. The ``Final SRB Offer'' uses the highest SRB offer available on the last day of a soldier's reenlistment eligibility window, which is generally 90 days prior to the end of the soldier's current enlistment.

Table B.13: Soldier's Survival Probabilities by Soldier Quality and VSI Program Eligibility

Dependent Variable: Indicator for Remaining in Military through VSI Period								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality Measure:	AFQT Score Percentile				Months below Sergent in first term			
	All Soldiers		6+ Years of Service		All Soldiers		6+ Years of Service	
VSI/SSB Eligibility	-0.099***	-0.196***	-0.097***	-0.151***	-0.198***	0.411***	-0.174***	0.047*
	(0.014)	(0.032)	(0.016)	(0.032)	(0.014)	(0.031)	(0.012)	(0.026)
VSI/SSB*Quality		0.193***		0.106***		-0.006***		-0.002***
		(0.030)		(0.029)		(0.000)		(0.000)
Quality	-0.099***	-0.107***	-0.022***	-0.034***	0.005***	0.005***	-0.001***	-0.000
	(0.011)	(0.011)	(0.008)	(0.008)	(0.000)	(0.000)	(0.000)	(0.000)
R-squared	0.154	0.155	0.168	0.168	0.230	0.240	0.176	0.182
Observations	189243	189243	60678	60678	161364	161364	32356	32356
Average Dep. Var	0.83	0.83	0.84	0.84	.84	.84	.85	.85
Fraction Eligible	.04	.04	.12	.12	.03	.03	.17	.17

Note: Standard errors are reported in parentheses. They are clustered at the MOS*Rank*YOS. Sample in column 1, 2, 5 and 6 is restricted to all soldiers serving on August 31, 1994 (the start of the sample period). Sample in Column 3, 4,79 and 8 is further restricted to those soldiers with between 6 an 20 years of service as of August 31, 1994. All regressions include occupation and rank fixed effects, a control for the years of service as of August 31, 1994, as well as controls for gender, age, marital status, and race. ``Ability'' is defined as AFQT score for columns (1)-(4) and months below Sergeant for columns (5)-(8). AFQT is on a scale from 0-1.

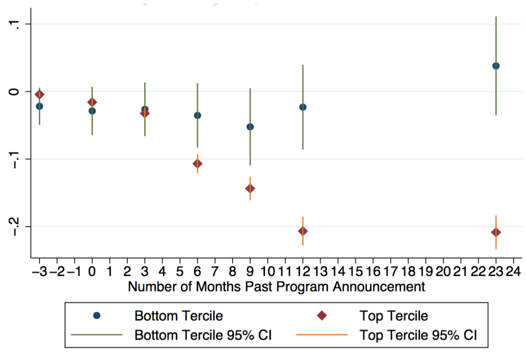
Table B.14: Soldier′s Survival Probabilities by Soldier Quality and TERA Program Eligibility

Dependent Variable: Indicator for Remaining in Military through TERA Period								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Quality Measure:	AFQT Score Percentile						Months below Sergent in first term	
	All Soldiers		15+ Years of Service		Around Cutoff		All Soldiers	
TERA Eligibility	-0.046***	-0.088***	-0.021	-0.053	-0.024	-0.089**	-0.145***	0.264*
	(0.014)	(0.032)	(0.016)	(0.032)	(0.021)	(0.040)	(0.032)	(0.152)
TERA*Ability		0.078		0.061		0.122*		-0.003***
		(0.056)		(0.057)		(0.067)		(0.001)
Ability	-0.060***	-0.060***	-0.022*	-0.026**	-0.023	-0.066*	0.003***	0.003***
	(0.008)	(0.008)	(0.012)	(0.012)	(0.035)	(0.037)	(0.000)	(0.000)
R-squared	0.107	0.107	0.115	0.115	0.078	0.079	0.148	0.148
Observations	254274	254274	24589	24589	4387	4387	219156	219156
Average Dep. Var	.91	.91	.87	.87	.84	.84	.92	.92
Fraction Eligible	.01	.01	.07	.07	.33	.33	<.01	<.01

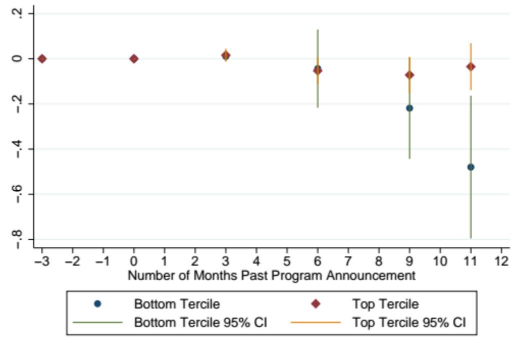
Note: Standard errors are reported in parentheses. They are clustered at the MOS*Rank*YOS. Sample in column 1, 2, 7 and 8 is restricted to all soldiers serving on August 31, 1994 (the start of the sample period). Sample in Column 3 and 4 is further restricted to those soldiers with between 15 an 20 years of service as of August 31, 1994. Columns 5 and 6 restrict the sample to those soldiers in an eligible occupation/rank but within 2 years (above or below) the mininmum years of service for program eligibility. All regressions include occupation and rank fixed effects, a control for the years of service as of August 31, 1994, as well as controls for gender, age, marital status, and race. ``Ability'' is defined as AFQT score for columns (1)-(6) and months below Sergeant for columns (7) and (8). AFQT is on a scale from 0-1.

Figure B-12: The Effect of Early Retirement Programs on Retention by Soldier Promotion Speeds

(a) The probability of surviving by VSI/SSB Eligibility



(b) The probability of surviving by TERA Eligibility



Notes: The left panel shows the probability of remaining in the Army for each month relative to August 1, 1993, the start of the VSI/SSB program, split by the soldier's promotion speed in his first term. We split soldiers into terciles of the months spent below sergeant in their first term. In each time period, we run a regression of program eligibility interacted with the soldier's promotion tercile on the probability of remaining in the military in period t. Each regression also includes occupation and rank fixed effects, a control for the soldier's tenure as of the program start date, dummies for the soldier's promotion speed tercile, and demographic controls (age, marital status, gender and race). Blue circle plot the coefficient on program eligibility interacted with the top tercile, and red triangles plot plot the coefficient on program eligibility interacted with the bottom tercile. The middle tercile was also included in the regression but is not plotted here. Lines show the 95% confidence intervals, with standard errors clustered at the occupation*rank*year of service bin. The sample includes the set of soldiers in the military on February 1, 1993, 6 months prior to the VSI program. The right panel shows similar specifications, but defines the sample and the time period relative to August 31, 1994, the day the TERA program was introduced. The right panel further restricts the sample to include only soldiers in the affected ranks and occupations, who are within 1 year of being eligible.

Table B.15: Relationship Between Soldier Ability and Take-Up of SSB vs. VSI

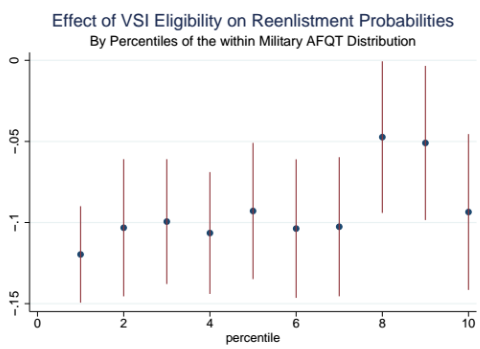
	(1)	(2)	(3)	(4)	(5)	(6)
	Ind. Var.: AFQT			Ind. Var.: Months E-4 or Below		
AFQT	-0.154***	-0.094***	-0.065**			
	(0.021)	(0.023)	(0.026)			
Months E-4 or below				0.082***	0.028	0.042**
				(0.016)	(0.017)	(0.019)
R^2	0.012	0.085	0.096	0.006	0.087	0.101
MOS FE	N	Y	Y	N	Y	Y
Rank FE	N	Y	Y	N	Y	Y
Demographic Controls	N	N	Y	N	N	Y
Dep. mean	.91	.91	.91	.92	.92	.92
Ind. Mean	53.81	53.78	53.94	88.23	88.35	87.57
Observations	5,620	5,573	5,323	4,970	4,928	4,753

Standard errors are reported in parentheses. Sample is restricted to the soldiers who were eligible for the second wave of the VSI/SSB programs and who chose to separate under one of the two programs. Demographic controls include gender, age, marital status, race, and special skill dummies. AFQT is on a scale from 0-1.

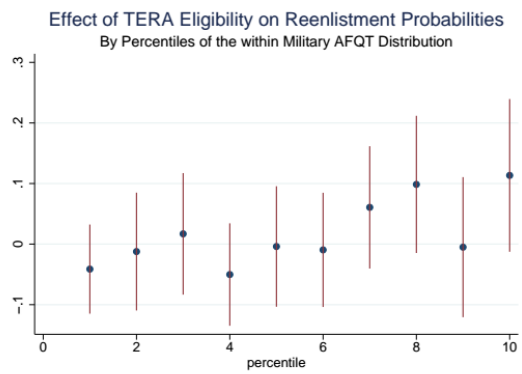
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure B-13: The Effect of Early Retirement Programs on Soldier Retention by Soldier Quality: Nonlinear Specifications

(a) The probability of remaining in the military by VSI/SSB Eligibility



(b) The probability of remaining in the military by TERA Eligibility



Notes: Each blue dot shows the estimate of program eligibility interacted with the soldier's AFQT score percentile from a regression where the dependent variable is an indicator for the solider still being in the military at the end of the program period. The regression also includes occupation and rank fixed effects, a control for the year of service, dummies for the soldier's AFQT score percentile, and demographic controls (age, marital status, gender and race). Standard errors are clustered at the occupation*rank*year of service bin. The left panel includes the sample of soldiers who were serving on August 1, 1993, the start of the VSI/SSB period, and the right panel includes the set of soldiers who were serving on August 31, 1994, the start of the TERA program. Additioally, the left panel also restricts the sample to those soldiers with at least 6 years of experience. The right panel restricts the sample to include only soldiers in the affected ranks and occupations, who have tenures that put them within 1 year of being eligible.

Table B.16: The Effect of SRBs on Soldier Retention, by AFQT Robustness Specifications Including Credit Score, Montgomery GI Bill, and Thrift Saving Program Interactions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Prime Credit Score			GI Bill Enrollment			Any TSP Contribution		
SRB	0.477***	0.481***	0.631***	0.282**	0.292**	0.562***	0.365***	0.360***	0.346***
	(0.145)	(0.145)	(0.152)	(0.115)	(0.115)	(0.112)	(0.093)	(0.094)	(0.094)
SRB * AFQT	-0.847***	-0.851***	-0.756***	-0.567***	-0.573***	-0.608***	-0.708***	-0.695***	-0.699***
	(0.188)	(0.187)	(0.181)	(0.126)	(0.126)	(0.125)	(0.132)	(0.132)	(0.135)
AFQT	-9.652***	-8.903***	-7.309***	-17.463***	-17.598***	-24.432***	-10.171***	-10.999***	-11.306***
	(0.955)	(0.912)	(1.050)	(0.849)	(0.847)	(1.678)	(0.907)	(0.906)	(0.903)
Mechanism Var.		-3.443***	-1.401**		-3.628***	-7.483***		4.870***	3.943***
		(0.290)	(0.624)		(0.456)	(1.066)		(0.181)	(0.369)
SRB * Mechanism Var.			-0.291***			-0.267***			0.080**
			(0.039)			(0.057)			(0.040)
AFQT * Mechanism Var.			-2.610***			7.558***			1.231**
			(0.882)			(1.414)			(0.564)
R^2	0.207	0.209	0.209	0.222	0.222	0.223	0.232	0.233	0.233
Year * Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
MOS * Rank * YOS FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year * Month * MOS FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Demographic Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
Avg. Reenlistment Rate	68.28	68.42	68.42	52.38	52.38	52.38	64.62	64.62	64.62
Avg. SRB	2.06	2.06	2.06	3.29	3.29	3.29	2.70	2.70	2.70
Observations	606,350	600,688	600,688	1,078,808	1,078,808	1,078,808	1,168,621	1,168,621	1,168,621

Standard errors are reported in parentheses. They are two-way clustered at the MOS*Rank*YOS and individual level. Sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. Samples for columns (1)-(3) are further restricted to soldiers with non-missing credit scores. Samples for columns (4)-(6) are restricted to soldiers with non-missing GI Bill participation data. Samples for columns (7)-(9) are restricted to soldiers with non-missing TSP contribution data. Prime credit score is a dummy variable for whether the soldier has a credit score of 680 or greater. GI Bill Enrollment is defined as a dummy variable for whether the soldier enrolls in the GI Bill at all. SRBs are in $1000s of 2015 dollars. Demographic controls include gender, age, marital status, race, and special skill dummies. AFQT is on a scale from 0-1.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

B.4 Selection and Average Ability Levels

In this section, we present empirical specifications and results demonstrating how the offer of either reenlistment bonuses or early retirement benefits affects the average quality of soldiers who are retained. The results in Section 2.4 showed that soldiers of higher ability are both less likely to reenlist in the military on average and are less responsive to both SRB offers and a pair of early early-retirement programs. Appendix Section B.1 further demonstrates that the effect this has on the average quality of retained soldiers is ambiguous and depends on the magnitude of the selection on ability. In this section, we show that our individual-level effects are large enough to generate changes in average soldiers ability-levels. This second analysis also enables us to characterize the quality of the marginal soldiers, i.e. the soldiers who were induced to reenlist when offered higher compensation.

Starting with the Army's SRBs, we estimate the change in the *average* quality of the "stayers" and the "leavers" using the following specification:

$$\begin{aligned}\text{AFQT}_i = \alpha_0 + \alpha_1 \text{SRB}_{it} * \text{Stay}_{it} + \alpha_2 \text{SRB}_{it} * \text{Leave}_{it} + \alpha_3 \text{Stay}_{it} + \\ \gamma_{MOS,rank,yos} + \mu_t + \delta \mathbf{X}_{it} + \epsilon_{it}, \quad \text{(B.4.1)}\end{aligned}$$

The coefficients of interest are α_1 and α_2, which estimate the effect of higher reenlistment bonus offers on the average ability of stayers or leavers, respectively. A positive value on α_1 would indicate that higher bonus offers tend to retain soldiers of higher average ability. As discussed in Section B.1, our basic conceptual framework offers ambiguous predictions regarding the effect of a change in relative military compensation on the average ability of either stayers or leavers. As in Equation 2.4.1, we include MOS×rank×years-of-service fixed effects.

Table B.17 shows estimates from Equation B.4.1, showing how the average ability of soldiers who chose to stay varies with the offered bonus. The identifying assumption underlying this analysis is that SRB offers are not systematically offered to cohorts of soldiers that are of higher quality. If this were the case, then we would observe that higher SRB offers are associated with higher quality reenlisted soldiers, but it would not reflect soldier selection.[3] The first column shows that this assumption is indeed satisfied – once we control for the set of fixed effects that determine the SRB offer, there is no correlation between the average ability of the soldiers eligible for reenlistment and their SRB offer. Columns 2 and Column 3 then split the sample by the soldier's reenlistment decision. Column 2 shows that when the SRB offer is \$10,000 dollars, the average ability of those soldiers who endogenously chose to stay in the military is 0.2 percentage points lower, although the estimate is noisy. As with the results in Table 2.2 and Figure 2-3, this shows that lower ability soldiers are more responsive to SRB offers, and enough so that they bring down average soldier quality. Column 3 shows, conversely, that when the SRB is higher, the average ability of those who leave the military is higher, although the estimate is also noisy. Column 4 pools the two samples and jointly estimates how the quality of the two groups endogenously changes as the bonus offer changes. The only difference between this specification and the split-sample specification in columns 2 and 3 is that the fixed effects are restricted to be the same, which gives us more power. When we do this, the results are qualitatively similar but even stronger – when an SRB of \$10,000 i is offered, the average AFQT

[3] Note that on average, in the raw data, soldiers of higher ability are offered higher bonus offers. This reflects the fact that soldiers of higher ability tend to be in higher skill occupations with more outside options. However, once we control for the soldiers occupation, tenure and rank, this positive correlation goes away.

score of the soldiers who reenlist is 0.48 percentage points lower and the average AFQT score of those who exit the military is 0.66 percentage points higher.

While at first glance these magnitudes look small, these are in fact quantitatively large effects. The average difference in quality between the stayers and the leavers is 1.2 percentage points. A $10,000 SRB bonus increases the difference between the two groups by an additional 1.1 percentage points, a 92 percent increase over the average difference between the two groups. Additionally, this reflects a difference in the *average* quality of the two groups. We can also examine the effect of SRBs on the quality of the *marginal* soldier – the soldier who would not have reenlisted but for the bonus offer. We can benchmark this with a simple back of the envelope calculation.[4] Column 1 of Table 2.2 shows that an SRB offer of $10,000 makes soldiers 1.5 percentage points more likely to reenlist. On average, 22,000 soldiers are eligible to reenlist each period, meaning that this SRB retained 330 additional soldiers. These marginally retained soldiers compose 2 percent of the reenlisted soldiers. Thus, in order for them to bring down the average of the reenlisted soldiers by 0.48 percentage points, the average AFQT score of the marginal soldiers must have been around the 32nd percentile. This would put the marginal soldier around the enlistment cutoff for AFQT scores, the lowest scores at which a person is eligible to join the Army.

The last two columns of Table B.17 repeat the analysis using our within-military measures of soldier quality. We see results here that are largely consistent with the AFQT results – when the SRB if higher, the average quality of the leavers is higher, in that they spent on average 4 more days as Sergeant in their first term when the SRB offered is $10,000 higher. While the selection along this dimension goes in the same direction as the selection across AFQT scores, the magnitude of the differ-

[4] We also plan to characterize this more formally following Gruber, Levine and Staiger (1999).

ence is smaller. For this measure of soldier quality, there is only an increase in the difference between the stayers and the leavers of 1.5 percent. Appendix Tables B.9 and B.10 show that these patterns are largely robust to alternative specifications and sample restrictions, including when we instrument for *actual* reenlistment bonuses with SRB offers.[5]

[5]Because the actual SRB offer is only observed for the set of people who reenlist, we restrict the sample to the stayers only. The actual SRB and the offered SRB can vary for several reasons—for example, the soldier may decide to reenlist for a term that is longer or shorter than 4 years, she may wait to reenlist until later in her enlistment window when the initial SRB offer is no longer available, or she may choose to switch occupations, thereby becoming eligible for an alternative SRB offer. Even so, the SRB offer available at the beginning of a soldier's reenlistment window is highly predictive of the actual SRB offer received. The IV estimates are noisier but similar in magnitude to the OLS regressions.

Table B.17: Selective Reenlistment Bonuses (SRBs) and Average Soldier Ability

	(1)	(2)	(3)	(4)	(5)	(6)
					Months below Sergent in first term	
Dependent Variable:	AFQT Score Percentile					
	Full Sample	*Stayers Only*	*Leavers Only*	*Full Sample*		
SRB	-0.015				0.004	
	(0.015)				(0.020)	
SRB*Stay		-0.021		-0.048***		0.022
		(0.015)		(0.015)		(0.030)
SRB*Leave			0.014	0.066***		-0.076***
			(0.019)	(0.022)		(0.027)
Stay				-1.216***		6.693***
				(0.118)		(0.505)
R-squared	0.302	0.313	0.293	0.304	0.326	0.342
Observations	1761615	1146584	614559	1761615	1708425	1708425
Year * Month FE	x	x	x	x	x	x
MOSxRankxYOS FE	x	x	x	x	x	x
Demographic Controls	x	x	x	x	x	x
Mean Dep. Var	58.26	57.5	59.67	58.26	54.4	54.4
Mean SRB	2.89	2.98	2.73	2.89	3.02	3.02

Note: Standard errors are reported in parentheses. They are twoway clustered at the MOS*Rank*YOS and individual level. The full sample is restricted to the soldiers who are eligible to reenlist in spells ending between 1997-2015. Column 2 restricts to the spells in which the soldier decides to renlist in the Army Column 3 restricts to the enlistment spells where the soldier decides to leave the Army. SRBs are in $1000 of 2015 dollars. Demographic controls include gender, age, marital status, race, and special skill dummies. The dependent variable is defined as AFQT score for columns (1)-(4) and months below Sergeant for columns (5)-(6).AFQT is on a scale from 0-100.

As before, we also examine the effect of these programs on average quality of retained soldiers by running the regression described in Equation B.4.1.

$$\text{AFQT}_i = \alpha_0 + \alpha_1 \text{ELIG}_i * stay_{i,t_T} + \alpha_2 * \text{ELIG}_i * leave_{i,t_T} + \alpha_3 stay_{i,t_T} + \gamma_{MOS,rank} + \delta \mathbf{X}_i + \epsilon_i, \quad \text{(B.4.2)}$$

The coefficients of interest from Equation B.4.2 are α_1 and α_2, which estimate the effect of drawdown program eligibility on the average ability among either stayers or leavers, respectively. Stayers are those who remain in the military at the end of the program eligibility window (t_T), and leavers are those who separate from the military at any point during the program eligibility window.

Table B.18 presents estimates from Equation B.4.2, showing how the average ability of those who chose to stay in the Army at the end of the program and those who chose to leave the Army varies with eligibility for the program. The first column shows that even after controlling for soldier rank, occupation, tenure and demographics, the average AFQT score of VSI/SSB-eligible soldiers is lower than that of ineligible soldiers. This is not a problem for identification, but it means that the coefficients in Column 2, which show the relative ability of the stayers and the leavers by the end of the VSI sample period, must be interpreted in relation to the coefficient on VSI/SSB eligibility in Column 1, rather than relative to 0 as in the earlier analysis.

Column 2 shows that by the end of the VSI period, the average AFQT score of the eligible stayers is about 1.2 percentage points higher and the average AFQT score of the eligible leavers is 1.6 percentage points lower than the average for the eligible population, shown in Column 1. Columns 3 and 4 show similar results on a more restricted sample of soldiers (namely, those with enough tenure to be among the general group of soldiers targeted by the early retirement program). Finally, Columns 5 through 8 show that the patterns are similar when considering the soldier's speed of promotion – by the end of the VSI period, the average ability of the soldiers still in the Army increased with program eligibility and the average ability of those outside the Army decreased with eligibility. Stayers spent 14.4 fewer months below the rank of sergeant than leavers – a large difference, equivalent to 18.7 percent of the average in the population. Appendix Table B.19 shows comparable results for the TERA program, which are qualitatively similar but statistically weaker.

Table B.18: Average Soldier Ability and VSI/SSB Eligibility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	AFQT Score Percentile				Months below Sergent in first term			
	All Soldiers		6+ Years of Service		All Soldiers		6+ Years of Service	
VSI/SSB Eligibility	-1.705***		-1.760***		4.390***		-3.903***	
	(0.343)		(0.365)		(0.816)		(0.528)	
VSI/SSB Eligibility*Stay		-0.529		-0.616		-9.812***		-10.653***
		(0.444)		(0.438)		(0.833)		(0.730)
VSI/SSB Eligibility*Leave		-3.346***		-3.134***		23.914***		3.768***
		(0.445)		(0.462)		(0.927)		(0.628)
Stay		-2.433***		-0.901***		20.752***		0.854**
		(0.248)		(0.224)		(0.584)		(0.361)
R-squared	0.281	0.283	0.320	0.321	0.370	0.439	0.641	0.650
Observations	189243	189243	60678	60678	161364	161364	32356	32356
Mean Dep. Var	58.57	58.57	54.74	54.74	59.24	59.24	81.06	81.06
Fraction Eligible	.04	.04	.12	.12	.03	.03	.17	.17

Notes: Sample in Column 1, 2, 5 and 6 is restricted to all soldiers serving on August 31, 1994 (the start of the sample period). Sample in Column 3, 4, 7 and 8 is further restricted to those soldiers with between 6 and 20 years of service as of August 31, 1994. All regressions include occupation and rank fixed effects, a control for the years of service as of August 31, 1994, as well as controls for gender, age, marital status, and race. Stay is defined as being in the Army at the end of the VSI/SSB period. The dependent variable is defined as AFQT score for columns (1)-(4) and months below Sergeant for columns (5)-(8). AFQT is on a scale from 0-100.

Table B.19: Average Soldier Quality and TERA Eligibility

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent Variable:	AFQT Score Percentile						Months below Sergent in first term	
	All Soldiers		15+ Years of Service		Around Cutoff		All Soldiers	
TERA Eligibility	1.450***		-0.211		0.529		31.700***	
	(0.528)		(0.559)		(1.127)		(1.945)	
TERA Eligibility*Stay		2.518***		1.436		0.153		14.569***
		(0.885)		(0.879)		(1.441)		(3.802)
TERA Eligibility*Leave		0.834		-0.922		0.310		44.929***
		(0.611)		(0.622)		(1.110)		(1.645)
Stay		-2.296***		-0.681**		0.440		18.480***
		(0.271)		(0.334)		(0.816)		(0.639)
R-squared	0.277	0.278	0.347	0.347	0.336	0.361	0.334	0.367
Observations	254274	254274	24589	24589	4387	4377	219156	219156
Mean Dep. Var	58.62	58.62	53.75	53.75	52.15	52.13	59.15	59.15
Fraction Eligible	.01	.01	.07	.07	.33	.33	<.01	<.01

Notes: Sample in Column 1, 2, 7 and 8 is restricted to all soldiers serving on August 31, 1994 (the start of the sample period). Sample in Column 3 and 4 is further restricted to those soldiers with between 15 an 20 years of service as of August 31, 1994. Columns 5 and 6 restrict the sample to those soldiers in an eligible occupation/rank but within 2 years (above or below) the mininmum years of service for program eligibility. All regressions include occupation and rank fixed effects, a control for the years of service as of August 31, 1994, as well as controls for gender, age, marital status, and race. Stay is defined as being in the Army at the end of the VSI/SSB period. The dependent variable is defined as AFQT score for columns (1)-(6) and months below Sergeant for columns (7) and (8). AFQT is on a scale from 0-100.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C

Chapter 3 Appendix

C.1 Additional Tables and Figures

Table C.1: Sample Average BAH Rates for 2016

Rank	Basic Monthly Pay[1]	BAH w/ Dependent	BAH w/o Dependents
E-3	$2,082.00	$1,042	$1,353
E-6	$3,612.30	$1,292	$ 1,663
O-5	$7,356.00	$1,785	$2,243

[1]Basic pay based on 3 years of service for E-3 and 12 years of service for E-6 and O-5

Table C.2: Mean Comparisons, by Quartile of Baseline (1996) Fair Market Rent

	1st Quartile	2nd Quartile	3nd Quartile	4nd Quartile	F-Stat
County Population (1998)	190,102.0	351,579.0	645,943.0	1,790,270.0	967.6
	(251,077.0)	(398,542.0)	(684,191.0)	(2,502,097.0)	
Population Growth (1998-2009)	8.9	14.8	24.0	13.2	498.0
	(13.6)	(15.6)	(20.9)	(13.5)	
County Median Income (1998)	35,812.0	41,052.0	43,943.0	49,186.0	1,459.7
	(5,610.0)	(7,626.0)	(8,940.0)	(10,276.0)	
Median Income Growth(1998-2009)	20.5	20.3	27.4	36.7	1,823.0
	(9.4)	(11.6)	(10.3)	(9.5)	
Military Share of Labor Force (%)	0.4	0.5	0.4	0.5	2.0
	(2.0)	(2.3)	(1.7)	(2.3)	
Distance to Nearest Base (mi)	69.0	56.0	35.0	25.0	950.0
	(42.0)	(41.0)	(35.0)	(25.0)	

This table compares covariate means across zip codes in the four quartiles of baseline (1996) Fair Market Rent. County median income is shown in 1998 U.S. $. Also shown is an F-statistic corresponding to the null that all four means are jointly equal.

Table C.3: OLS Estimates
of the Effect of BAH on Local Rental Prices, Including Lags

	(1)	(2)	(3)	(4)
Δ log(BAH) * Military Base	-0.043	-0.044*	-0.041*	-0.041*
	[0.026]	[0.026]	[0.023]	[0.023]
L.Δ log(BAH) * Military Base	0.059***	0.059***	0.068***	0.070***
	[0.021]	[0.021]	[0.022]	[0.021]
L2.Δ log(BAH) * Military Base	0.038**	0.036**	0.039**	0.037**
	[0.018]	[0.018]	[0.017]	[0.017]
L3.Δ log(BAH) * Military Base	0.030**	0.029**	0.031**	0.033**
	[0.014]	[0.014]	[0.014]	[0.014]
Δ log(BAH)	0.218***	0.217***	0.206***	0.205***
	[0.033]	[0.033]	[0.031]	[0.031]
L.Δ log(BAH)	0.062***	0.062***	0.052***	0.051***
	[0.016]	[0.016]	[0.016]	[0.016]
L2.Δ log(BAH)	0.085***	0.085***	0.081***	0.079***
	[0.024]	[0.024]	[0.022]	[0.022]
L3.Δ log(BAH)	0.038**	0.038**	0.040**	0.037**
	[0.017]	[0.017]	[0.016]	[0.015]
Military Base	0.008	0.008	0.001	0.000
	[0.014]	[0.013]	[0.016]	[0.015]
L.Military Base	-0.023	-0.022	-0.023	-0.023
	[0.015]	[0.015]	[0.015]	[0.015]
L2.Military Base	0.003	0.002	0.003	0.002
	[0.010]	[0.009]	[0.011]	[0.011]
L3.Military Base	0.004	0.006	0.007	0.008
	[0.012]	[0.012]	[0.012]	[0.011]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs			Yes	Yes
Controls		Yes		Yes

This table presents OLS estimates of the effect of change in log BAH on change in log Fair Market Rents. Changes in BAH rates are interacted with a dummy variable the presence of a military base in the same 3-digit zipcode. Explanatory variables are lagged up to three periods. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table C.4: Effect of Cohen Initiative Instruments on Fair Market Rents (Reduced Form)

	(1)	(2)	(3)	(4)
1996 FMR * Year 2000	0.036**	0.039***	0.030**	0.033**
	[0.015]	[0.013]	[0.014]	[0.013]
1996 FMR * Military Base * Year 2000			0.026**	0.029**
			[0.012]	[0.013]
1996 FMR * Year 2001	0.077***	0.080***	0.073***	0.074***
	[0.020]	[0.020]	[0.021]	[0.021]
1996 FMR * Military Base * Year 2001			0.007	0.014
			[0.024]	[0.025]
1996 FMR * Year 2002	0.045**	0.047**	0.037*	0.038*
	[0.020]	[0.019]	[0.020]	[0.019]
1996 FMR * Military Base * Year 2002			0.033**	0.039**
			[0.015]	[0.016]
1996 FMR * Year 2003	0.093***	0.095***	0.084***	0.084***
	[0.029]	[0.028]	[0.029]	[0.028]
1996 FMR * Military Base * Year 2003			0.034	0.041
			[0.027]	[0.027]
1996 FMR * Year 2004	0.064***	0.065***	0.055***	0.054***
	[0.007]	[0.007]	[0.007]	[0.007]
1996 FMR * Military Base * Year 2004			0.039***	0.045***
			[0.012]	[0.013]
1996 FMR * Year 2005	-0.131***	-0.134***	-0.152***	-0.155***
	[0.039]	[0.037]	[0.038]	[0.037]
1996 FMR * Military Base * Year 2005			0.088**	0.095**

Continued on next page

Table C.4 – continued from previous page

	(1)	(2)	(3)	(4)
			[0.042]	[0.044]
1996 FMR * Year 2006	-0.012	-0.013	-0.024	-0.026
	[0.024]	[0.023]	[0.026]	[0.025]
1996 FMR * Military Base * Year 2006			0.065*	0.072*
			[0.039]	[0.038]
1996 FMR * Year 2007	0.035***	0.035***	0.028***	0.027***
	[0.010]	[0.010]	[0.009]	[0.009]
1996 FMR * Military Base * Year 2007			0.040***	0.045***
			[0.012]	[0.015]
1996 FMR * Year 2008	0.003	0.003	-0.001	-0.002
	[0.014]	[0.014]	[0.013]	[0.012]
1996 FMR * Military Base * Year 2008			0.027	0.033
			[0.029]	[0.031]
1996 FMR * Year 2009	0.005	0.006	0.005	0.005
	[0.012]	[0.012]	[0.010]	[0.010]
1996 FMR * Military Base * Year 2009			0.006	0.012
			[0.020]	[0.022]
		[0.020]	[0.022]	
1996 FMR * Military Base			-0.000	-0.000
			[0.001]	[0.001]
Year FEs	Yes	Yes	Yes	Yes
Zipcode FEs	Yes	Yes	Yes	Yes
Population & Income Controls		Yes		Yes

Continued on next page

Table C.4 – continued from previous page

	(1)	(2)	(3)	(4)

This table presents Reduced Form estimates of the effect of baseline log Fair Market Rent on change in log FMR, where the baseline is measured in 1996. Controls include county-level population and income dynamics. Standard errors are clustered at the state level.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$