From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

---

Doctoral Dissertations, 2011-2026 Economics

2018

Essays in Debt, Income, and Inequality

Herbst, Daniel

Princeton, NJ : Princeton University

---

http://arks.princeton.edu/ark:/88435/dsp012b88qf882

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

Essays in Debt, Income, and Inequality

Daniel J. Herbst

A Dissertation

Presented to the Faculty

of Princeton University

in Candidacy for the Degree

of Doctor of Philosophy

Recommended for Acceptance

by the Department of

Economics

Adviser: Will Dobbie

June 2018

© Copyright by Daniel J. Herbst, 2018.

All rights reserved.

Abstract

This dissertation is comprised of three essays on various institutions influencing individuals' incomes and the aggregate effects of those institutions on social welfare. Chapter one investigates a large government-sponsored student loan program known as income-driven repayment (IDR). By aligning the repayment burden with the returns to college, IDR may prevent default and improve financial well-being among credit-constrained borrowers but carries a potential cost to social welfare through moral hazard. In this chapter, I estimate the causal impact of IDR on repayment rates, balances, homeownership, and consumption proxies, and find that enrolling in IDR is associated with a dramatic decrease in delinquencies and a significant increase in credit card spending and homeownership rates. Chapter two focuses on the private market (or lack thereof) for income insurance products. Specifically, I empirically test the hypothesis that adverse selection prevents the profitable sale of income-share agreements (ISAs) for college financing. By assessing the predictive power of survey responses concerning expected future income, I find evidence that student borrowers hold private knowledge of their future income which, under reasonable assumptions, can cause a hypothetical ISA market to completely unravel. Chapter three, which is coauthored with Henry Farber, Ilyana Kuziemko, and Suresh Naidu, examines the role of unions in aggregate income trends over the twentieth century. In this chapter, we present a number of results that, across a variety of identifying assumptions, suggest unions have had a significant, equalizing effect on the income distribution over a long sample period.

Acknowledgements

I am extremely grateful to my advisers Will Dobbie and Ilyana Kuziemko for their guidance and mentorship. Their tireless dedication to my development as an economist has helped shape not only this dissertation, but my entire way of thinking. I'd also like to thank Alex Mas, who has fostered my professional growth since my earliest days as a graduate student. Coauthoring with Hank Farber and Suresh Naidu has exposed me to brilliance and wisdom in equal measure, and the constructive comments I received from Leah Boustan, Alan Krueger, David Lee, Chris Neilson, and others were indispensable as I wrote this dissertation. I was also greatly aided by generous financial support from the Princeton Industrial Relations Section and the National Academy of Education Spencer Dissertation Fellowship.

I'd like to thank my colleagues at Princeton for making sure I was never the smartest person in the room. Time spent with Shoumitro Chatterjee, Graham McKee, Emil Verner, and others has made earning my Ph.D. a fulfilling experience, both personally and intellectually. I'd especially like to thank Felipe Goncalves, whose thoughtfulness and good nature has helped sustain my mental health over the past six years.

Finally, my parents, Dale and Kristina Herbst, have been uncompromising in their encouragement and support for all my endeavors. So much of my will to succeed has been driven by their faith in me, which is undeserved but greatly cherished. Thanks, Mom and Dad.

Contents

Abstract iii

Acknowledgements iv

List of Tables ix

List of Figures x

1 Liquidity and Insurance in Student Loan Contracts: The Effects of Income-Driven Repayment on Default and Consumption 1

1.1 Introduction 2

1.2 Background 7

1.2.1 Overview 7

1.2.2 Setting: Student Loan Servicing 10

1.3 Data 11

1.4 Model 15

1.4.1 Setup 15

1.4.2 Liquidity Effects through Reduced Payments 17

1.4.3 Wealth Effects through Debt Forgiveness 19

1.4.4 Social Welfare Implications 20

1.5 Empirical Strategy 22

1.5.1 Difference-in-Differences Specification 23

1.5.2 Instrumental Variables 25

1.6 Results . . . 30
1.6.1 Repayment Outcomes . . . 30
1.6.2 Credit and Consumption Outcomes . . . 32
1.7 Interpretation . . . 33
1.7.1 Loan Forgiveness . . . 36
1.7.2 Alternative Mechanisms . . . 37
1.8 Conclusion . . . 38
1.9 Tables and Figures . . . 40

2 Adverse Selection, Income Insurance, and College Financing 55
2.1 Introduction . . . 56
2.2 The Empirics of Adverse Selection . . . 60
2.2.1 Data . . . 61
2.3 Evidence of Private Information . . . 62
2.3.1 Empirical Approach . . . 64
2.3.2 Results . . . 67
2.4 Model of Adverse Selection . . . 68
2.4.1 Setup . . . 68
2.4.2 No-Trade Condition . . . 69
2.5 Evidence of Market Unraveling . . . 69
2.6 Conclusion . . . 71
2.7 Tables and Figures . . . 72

3 Unions and Inequality Over the Twentieth Century: New Evidence from Survey Data 83
3.1 Introduction . . . 84
3.2 The Economics of Unions and Inequality . . . 89

3.3 Newly Available Data On Union Membership . . . 93
3.3.1 Gallup Methodology before and after 1950 . . . 94
3.3.2 Additional checks on pre-1950 Gallup data . . . 95
3.3.3 Additional data sources . . . 96
3.3.4 Existing time series of aggregate union density . . . 97
3.4 A New Time-series Measure of Household Union Membership . . . 98
3.4.1 The union share of households over time . . . 98
3.4.2 Differences between Gallup and the historical aggregate series . . . 99
3.5 Selection Into Unions . . . 101
3.5.1 Selection into union across time . . . 102
3.5.2 Modeling selection as a parametric function of density . . . 106
3.6 The union family income premium over the twentieth century . . . 108
3.6.1 Estimating the union family income premium over time . . . 109
3.6.2 Robustness and related results . . . 110
3.6.3 Heterogeneous union household income effects . . . 112
3.6.4 Effects on residual income dispersion . . . 113
3.7 The Effect Of Unions On Inequality . . . 114
3.7.1 Distributional regressions . . . 115
3.7.2 Time-series regressions . . . 118
3.7.3 State-year panel regressions . . . 121
3.7.4 Discussion of Estimate Magnitudes . . . 124
3.8 Conclusion . . . 126

A Chapter 1 Appendix **142**

A.1 Additional Tables and Figures . . . 143
A.2 Mathematical Appendix . . . 158
A.2.1 Model Solution . . . 158

A.2.2 Proof of Liquidity Effects 159

B Chapter 2 Appendix **163**

B.1 Technical Appendix 164

B.1.1 Proof of Theorem 1 164

B.1.2 Flexible Contract Model 164

C Chapter 3 Appendix **166**

C.1 Supplementary Figures and Tables Noted in the Text 167

C.2 Sample Selection and Construction of Key Variables 197

C.2.1 Sample Selection 197

C.2.2 Variable Construction 197

C.3 Background on Gallup Data 199

C.4 Existing Measures of Union Density Pre-Dating the Current Population Survey 207

C.4.1 The BLS estimate of early union density 207

C.4.2 The Troy estimates of early union density 209

C.4.3 Other Pre-CPS State-Year Measures of Union Density 210

C.5 Details on RIF Regressions 212

Bibliography **214**

List of Tables

1.1 Summary Statistics 41

1.2 Balance Test 42

1.3 First Stage 43

1.4 Difference-in-Differences and Instrumental Variables Estimates of the Effect of IDR on Loan Repayment Outcomes 44

1.5 Difference-in-Differences Estimates of the Effect of IDR on Financial Outcomes 45

1.6 Difference-in-Differences Estimates of the Effect of IDR on Credit Cards . . 46

2.1 Summary Statistics 72

2.2 OLS Regressions of 2014 Salary on Public Information and Subjective Elicitations 73

2.3 Random Forest Model Statistics 74

2.4 P-values from Pairwise Comparison Tests 75

3.1 Is selection into unions a function of union density? 138

3.2 Aggregate inequality as a function of union density 139

3.3 State-year inequality as a function of union density 140

3.4 How much of the variation in inequality is explained by changes in union density? 141

List of Figures

1.1 Hypothetical Repayment Scenarios: IDR versus Standard Repayment 47

1.2 Pre/Post-Call Trends in Repayment Outcomes 48

1.3 Pre/Post-Call Trends in Credit Scores and Bankruptcies 49

1.4 Pre/Post-Call Trends in Mortgages and Auto Loans 50

1.5 Pre/Post-Call Trends in Credit Cards 51

1.6 First Stage 52

1.7 Difference-in-Differences: Repayment Outcomes 53

1.8 Two-Stage Least Squares: Repayment Outcomes 54

2.1 Correlation Between Subjective Elicitations and Realized Income 76

2.3 CDF of Prediction Differences 77

2.4 Out-of-sample Training/Test Procedure 78

2.5 10-Fold Out-of-sample Training/Test Procedure 79

2.6 10-Fold Out-of-sample Prediction Statistics 80

2.8 Private Future Income Distributions 81

2.9 Predicted Insurance Rejection Rates under Alternative γ 82

3.1 Union density and inequality measures, 1917-2011 128

3.2 Union density over the twentieth century: Comparing our survey-based measures to existing time-series 129

3.3 How does educational attainment predict union household status? 130

3.4 How does race predict union household status? 131
3.5 Estimates of the union family income premium 132
3.6 Differential family union premium by respondent's years of schooling 133
3.7 Differential family union premium for whites relative to minorities 134
3.8 Ratio of residual variance between union and non-union sectors 135
3.9 Effects of union density on the 90/10 ratio of family income 136
3.10 Effects of union density on the Gini coefficient of family income 137

Chapter 1

Liquidity and Insurance in Student Loan Contracts: The Effects of Income-Driven Repayment on Default and Consumption[1]

[1]Versions of this chapter were presented at several research venues, including the CFPB Research Conference, the Federal Reserve Board, the Jain Family Institute, the National Academy of Education Spring Research Conference, the RAND Corporation, and the University of Arizona.

1.1 Introduction

Each year, roughly one million borrowers default on their student loans, and millions more struggle to buy homes ([106]), accumulate wealth ([21]), or choose their preferred career ([116]). Many economists blame the rigid repayment terms on student loans, which require fixed payments averaging $350 per month early in borrowers' careers ([34]).[2] By far, the largest and most far-reaching policy response has been "Income-Driven Repayment" (IDR). First introduced in 2009, IDR allows students to repay a fixed portion of their yearly earnings until either their debt is paid off or some forgiveness period has been reached, offering an alternative to the flat, ten-year repayment schedule under traditional repayment plans. Enrollment in IDR has tripled since 2014, and today over $300 billion in debt is repaid through IDR ([36]).

Even as IDR take-up continues to rise, its effects on borrowers are largely unknown. By aligning debt repayment with the returns to college investment, IDR may provide the liquidity needed to help credit-constrained borrowers smooth consumption over the life-cycle and insure against transitory income shocks. Evidence of high payment-to-income ratios ([101]) and limited credit ([75]) among young borrowers suggests such liquidity may increase borrowers' welfare, allowing them to pay down their student debt while avoiding financial distress or sharp drops in consumption. However, these short-term liquidity benefits may be outweighed by the long-term costs of IDR's forgiveness provisions, which cancel any unpaid debt after twenty-five years of payments. If enrollees' incomes are *permanently* low, the repayment burden ultimately falls on the government, reducing social welfare through redistributive inefficiencies, administrative costs, and distorted incentives (i.e., moral hazard).

Assessing the costs and benefits of IDR requires causal estimates of its impact on repayment and consumption, but two obstacles have prevented researchers from identifying these

[2]For example, [15] write, "The US student loan system is currently in crisis...mainly due to the fact that the US operates mortgage-type student loans: these are repaid over a set period of time, which places high repayment burdens on low earning graduates."

effects. First, an empirical analysis of IDR requires high-frequency repayment data for many borrowers, but until now these data have been unavailable.[3] Second, IDR enrollees are by design a very selected group, as only borrowers with a sufficiently high debt-to-earnings ratio would benefit from enrolling in IDR. Estimates which rely on cross-sectional comparisons of IDR enrollees to non-enrollees could therefore be biased in either direction, as low-income individuals typically have worse financial outcomes, but borrowers with high student debt balances are often highly educated and positively selected ([126]).

In this paper, I use novel administrative data from a large student loan servicing company to estimate the causal effect of IDR enrollment on defaults, credit scores, bankruptcies, and consumption proxies. The data I use link monthly loan records from a large loan servicer ("LLS") to administrative credit bureau information from TransUnion for over one million student borrowers. LLS manages disbursals and payments for over $300 billion in federal student loans. The LLS data include detailed loan information (e.g., balances, payments, delinquency status, repayment plan) at a monthly frequency, as well as contact histories, zip code, age, institution attended, and college enrollment dates. These data are, to the best of my knowledge, the first panel of U.S. federal student loan payments used in public research.

I use two complementary research designs to identify the effect of IDR on student borrowers, both of which exploit plausibly exogenous variation from loan servicing phone calls. First, I estimate the difference-in-differences between IDR-enrollees (the "treatment group") and non-enrollees (the "control group") before and after receiving delinquency calls from LLS informing them of IDR options. The identifying assumption for this design is that, in

[3]Economists have long bemoaned the dearth of available student loan data. In a 2015 meeting, William Dudley, President of the Federal Reserve Bank of New York, remarked, "Do borrowers who use programs like income-based repayment eventually succeed in paying off their debts? How do income-based repayment programs affect important decisions such as labor supply, consumption and household formation? These are important questions for the nation,...but it is very hard to answer these questions with existing data." ([50]). In a column for the New York Times, Susan Dynarski echoes these sentiments, writing, "We are remarkably ignorant about student debt...But at the moment, the federal student loan data remains locked within the walls of the Education Department, with limited metrics trickling out," ([52]) and "Data on student loans are remarkably thin, given the size of this market. They are particularly inadequate for modeling and costing out income-based repayment plans," ([51]).

the absence of IDR, post-call outcomes for treatment and control groups would have exhibited parallel trends. To support this assumption, I show that trends in borrower outcomes are nearly identical across the two groups for several periods before the delinquency call.

My second empirical strategy is an instrumental variables (IV) design exploiting the quasi-random assignment of delinquency calls to debt-servicing agents via an automatic dialing system. I use variability in agents' tendencies to induce IDR take-up as a means of identifying the local average treatment effect (LATE) of IDR on individuals whose repayment plan decisions depend on the servicing agent to whom their delinquency calls are connected. I measure an agent's IDR inducement rate using the leave-one-out, mean IDR status among the borrowers who receive a call from that particular agent. The agent effectiveness measure is predictive of IDR take-up, but uncorrelated with borrower characteristics and pre-call outcomes. This research design resembles [89], which uses random judge assignment to estimate the effect of incarceration length on earnings, as well as subsequent research focusing on consumer bankruptcy ([47]), juvenile incarceration ([6]), pre-trial detention ([44]), family welfare culture ([35]), disability insurance ([91]), and foster care ([11]). While the quasi-experimental nature of the instrumental variables design holds obvious advantages, the increased statistical precision of difference-in-differences estimates makes it my preferred design. Nonetheless, IV estimates provide complementary evidence to the difference-in-differences design, as results are generally consistent between the two strategies.

Results suggest IDR increases loan repayment, credit scores, homeownership, and consumption among student borrowers. Relative to borrowers who remain on standard repayment plans, delinquency rates among IDR-enrollees fall by 21.0 percentage points within seven months of take up. While IDR reduces monthly minimum payments by an average of $140 through a mechanical "first-stage" effect, the effect of reduced minimums on loan balances is dominated by more timely repayment; IDR borrowers pay down $90 more student debt each month, on average, than those on standard repayment plans. IDR enrollees

have credit scores which are 7.5 points higher, hold 0.1 more credit cards, and carry $240 higher credit card balances than non-enrollees one year after the servicing call.[4] While I find no effects on bankruptcies or auto loans, IDR enrollees are 2 percentage points more likely than non-enrollees to hold a mortgage after two years, an increase of 10 percent off of the pre-call mean. Estimated effects on financial outcomes outlast the "first-stage" effect of IDR on payment size: monthly minimum payments return to pre-call levels by fifteen months after the delinquency call. These results suggest that short-term increases in cash-on-hand through IDR have long-term positive impacts on the financial health of credit-constrained borrowers.

In theory, my estimates combine three effects of IDR: (1) a liquidity effect through lower minimum payments, (2) a wealth effect through expected debt forgiveness, and (3) a moral hazard effect, as income-contingent forgiveness can distort labor supply or investment decisions. For two reasons, I argue my estimates capture a pure liquidity effect: First, minimum monthly payments decrease by an average of $140 following IDR take up, but return to standard levels within one year of enrollment, a rate which makes future loan forgiveness unlikely under the terms of the IDR plan I study. Second, because borrowers can opt-in to IDR at any time, any moral hazard from income-contingent loan forgiveness should apply to both treatment and control groups in my sample, as both groups are eligible for IDR and, by construction, aware of its existence. Estimates should therefore net out moral hazard effects, leaving a pure liquidity effect.[5]

My analysis carries four important caveats. First, the effects I estimate hold interest rates and plan-specific repayment terms fixed. In a private market, one would expect the availability of IDR to have general-equilibrium effects on these parameters. For instance, ad-

[4]For reference, [45] finds that the removal of a flag designating Chapter 13 bankruptcy from one's credit report is associated with a 9.8 point increase in credit scores, a $143 increase in credit card balances, and a 2 percentage point increase in the likelihood of holding a mortgage one year after the flag removal.

[5]Note that increased liquidity may itself affect labor supply, even in the absence of forgiveness. I view such effects as conceptually distinct from moral hazard. See [32] and [118] for a discussion of liquidity versus moral hazard in the context of unemployment insurance.

verse selection of individuals with low expected earnings into IDR could have an unraveling effect on a hypothetical "repayment-plan market."[6] In the existing student loans environment, however, these types of effects are unlikely, as interest rates and repayment terms are held fixed by the federal government.[7] My analysis therefore treats loan terms as policy parameters rather than equilibrium objects. Nonetheless, equilibrium effects could potentially operate through political mechanisms like budgetary constraints or political pressure. IDR could also have general equilibrium effects on labor, marriage, or higher education markets. These potential market-level responses are excluded from the partial-equilibrium effects I estimate in this paper.

Second, because my sample includes only IDR-eligible borrowers while they are already in repayment, the treatment effect I identify excludes any ex-ante effects on decisions pertaining to college attendance, occupation choice, or principal borrowing amounts. In many ways, removing these effects is a theoretically attractive feature of my research design, as it allows me to isolate a single mechanism for observed effects – ex-post increases in liquidity. Nonetheless, removing ex-ante responses to expected payment flexibility or forgiveness leaves several unanswered questions important for the social welfare implications of IDR.

Third, my analysis is limited to the repayment and consumption-related outcomes available in my data. While these outcomes highlight the primary channel through which IDR is intended to benefit borrowers, IDR undoubtedly also influences borrower welfare through potential effects on labor supply and household formation. My analysis cannot directly speak to these channels, as my data do not include information on earnings, employment, or marital status.

[6] In fact, the existence of adverse selection may explain why there exists virtually no private market in which to study the phenomenon. In concurrent work, I use private elicitations of expected future income among student borrowers to provide evidence in support of this hypothesis.

[7] A small private student loans market constitutes around ten percent of total student debt, mostly for creditworthy graduate students or borrowers who have exhausted their federal loan limits. In most cases, however, private lenders cannot compete with the subsidized rates offered by the government under the Federal Family Education Loan (FFEL) and Federal Direct Loan programs. Unless stated otherwise, I will use "student loans" to refer to loans originating from these federal programs.

Finally, for several reasons, the outcomes I measure can not perfectly capture welfare improvements through consumption or repayment: (1) The lack of IDR enrollment prior to 2013 makes it impossible to measure effects over long-term time horizons. (2) The credit card, auto loans, and mortgage data I use provide useful proxies for consumption and home-ownership ([72], [39], [87]), but are not direct measures of these variables. (3) While default and delinquency carry negative consequences for borrowers, the welfare implication of these consequences remains an open question, so estimates of improved repayment rates under IDR are difficult to interpret.[8]

The remainder of this paper is organized as follows. Section 2.2 provides a brief overview of federal student loans and student loan servicing in the United States. Section 2.2.1 describes the student loan and credit bureau data and provides summary statistics. Section 2.4 motivates my empirical analysis with a model of consumption under alternative student loan repayment plans and discusses the potential welfare implications of IDR. Section 1.5 describes my empirical strategy. Section 1.6 presents results, and Section 1.7 provides interpretation. Section 1.8 concludes.

1.2 Background

1.2.1 Overview

Over 90% of student loans in the United States are federally subsidized and guaranteed. The government holds the liability on student loans, and interest rates are set by Congress.[9]

[8]Different stages of non-repayment on student loans carry increasingly severe late fees and credit consequences, culminating in an official default classification after 270 days past due. "Default" is a legal distinction, defined as "Failure to repay a loan according to the terms agreed to in the promissory note." Defaulted loans can result in garnished wages, withheld tax returns, and revocation of professional licenses ([36]).

[9]Congress has set rates on student loans since 1965, though automated the process in 2013 with the Bipartisan Student Loan Certainty Act. The legislation mandates a rate-setting rule for student loans, setting interest rates equal to the 10-year Treasury bond rate plus 205 basis points (360 bps for graduate students). Interest rates are fixed throughout the life of a loan and accrue as simple daily interest on principal only.

Student loans are not secured by collateral or subject to any credit check.[10] While the federal government holds the liability on student loans, the Department of Education contracts private debt servicing companies to handle disbursals, billing, and processing. Debt servicers disburse loans to colleges' financial aid offices, which apply disbursed funds directly to students' accounts. While the amount one can borrow from federal sources is capped by semester, virtually anyone attending an accredited institution is eligible to borrow at the same subsidized rate.[11]

The Department of Education sets repayment terms for student loans through several repayment plans, each requiring borrowers make monthly payments to their loan servicer. The default repayment option into which all borrowers are automatically enrolled is known as "standard repayment." Under standard repayment, the minimum amount a borrower is required to pay follows a flat repayment schedule over ten years, so that minimum monthly payments are calculated as:

$$\text{Monthly Standard Payment} = \frac{i * \text{Principal}}{1 - (1+i)^{-(10*12)}} \tag{1.1}$$

where i denotes the monthly interest rate. Until 2010, the vast majority of borrowers in repayment were enrolled in standard repayment plans, with only a small fraction of borrowers choosing alternative financing options. Income-driven repayment (IDR) plans first became available in 2009 as an alternative to standard repayment.[12] Minimum payments under IDR are pegged to fifteen percent of borrowers' discretionary income, defined as the differ-

[10]The exception is PLUS loans, for which parents can serve as cosigners, subject to a credit check.

[11]In the short term, borrowing costs can vary by financial need, as the "Subsidized Stafford Loan" program forgives interest accrued while the borrower is still in school, up to a means-tested limit.

[12]Since 2009, several IDR plans have become available, including Income-Based Repayment (IBR), Pay-As-You-Earn (PAYE), and Revised-Pay-As-You-Earn (REPAYE). The eligibility requirements and repayment terms can vary across these IDR plans, though they share the same general structure. For the purposes of this study, I focus on Income-Based Repayment (IBR), as it is the only IDR plan for which the Federal Family Education Loan (FFEL) borrowers in my sample are eligible, though the discussion and conclusion generalize to the broader concept of IDR.

ence between adjusted gross income (AGI) and 150% of the federal poverty line (FPL).[13] Specifically,

$$\text{Monthly IDR Payment} = 15\% * \left(\frac{\text{AGI} - 1.5 * \text{FPL}}{12} \right) \tag{1.2}$$

Monthly payments are capped at the standard minimum payment amount, and payments continue until the borrower's balance reaches zero. If a borrower successfully makes twenty-five years of payments under IDR, any remaining balance is forgiven.[14]

A graphical comparison of repayment plans can be seen in Figure 1.1. The figure plots IDR and standard repayment paths under alternative income scenarios for a borrower leaving college with $18,000 in student loans. In panel A, the borrower's income is too high to qualify for reduced payments: her payments are identical under standard (dashed blue line) and IDR (dotted red line) plans. In panel B, the borrower's income is low enough to reduce IDR payments in the first five years of repayment, but rises to the standard amount thereafter, extending the repayment period beyond the standard length of ten years but not long enough to qualify for loan forgiveness at twenty-five years. Panels C and D, by contrast, depict scenarios in which the borrower is granted partial loan forgiveness under IDR with panel D highlighting the "dynamic" response of IDR to a temporary earnings shock such as unemployment.

Borrowers can switch to IDR at any point in the repayment process. Opting-in requires completing an online form through the Department of Education, which verifies income and family size using information from a borrower's most recent federal tax return. Borrowers must recertify their income on a yearly basis, although they can adjust their payments more frequently by providing proof of income. If a borrower on IDR goes more than one

[13] Adjusted gross income (AGI) is an individual's total annual income minus specific tax deductions. The federal poverty line (FPL) is a government-specified income threshold determined by household size and state of residence. For 2017, the FPL for a family of four is $24,600 in every state but Alaska and Hawaii.

[14] Forgiveness periods can vary. While the specific IDR plan I study forgives loans only after twenty-five years, loans for more recent IDR borrowers are typically forgiven after twenty years, and non-profit or government employees can often qualify for a ten-year forgiveness period under the Public Service Loan Forgiveness program. Regardless of the period, any forgiven balance is taxed as income in the year of forgiveness.

year without recertifying income and family size, her payments automatically return to the standard payment amount.

Borrowers who fail to meet their monthly payments (i.e., "fall delinquent") under any repayment plan face penalties which increase in severity with the number of days past due. Between one and fifteen days past due, student loan servicers will contact delinquent borrowers through email or post. Between fifteen and ninety days past due, borrowers are charged late fees and contacted by phone at increasing frequency to encourage repayment and discuss repayment options. At 91, 181, and 271 days past due, borrowers are reported to credit bureaus, and loans more the 270 days past due are considered eligible for default. Once in default, all remaining balance on student debt becomes due, and the Department of Education can garnish up to 15 percent of borrowers' wages or withhold their tax returns to collect on defaulted debt. Unlike other forms of consumer debt, student loans cannot be discharged by declaring Chapter 7 or Chapter 13 bankruptcy.[15] Defaulted borrowers are ineligible for any future federal student aid.

1.2.2 Setting: Student Loan Servicing

LLS manages disbursals and payments for over $300 billion in federal student loans. Debt servicing is provided on behalf of the federal government, which hires the servicer through a series of contracts.

As a part of its servicing operations, LLS makes frequent contact with delinquent borrowers to encourage repayment. When borrowers become fifteen or more days past due on their payments, their phone numbers are placed in a dialing queue. An automatic dialer then places calls to each of the numbers in this queue in rapid succession. If a call is unanswered, the dialer places it back at the bottom of the queue. Each answered call is immediately

[15] In rare circumstances, borrowers who demonstrate "undue financial hardship" can discharge their student debt in bankruptcy. Student loans can also be discharged if borrowers are disabled, deceased, or attended an institution which has since closed ([36]).

connected to a debt-servicing agent randomly selected from the pool of available agents. If no agents are available, the dialer places the borrower on hold until one becomes available. Such instances are extremely rare, however, as the dialer places calls at a rate to match agent availability, which is highly predictable over large numbers of agents.

LLS employs over 300 servicing agents across four call centers. Agents are tasked with informing borrowers of their delinquent status, inquiring about their intention to repay, and informing them of repayment options. During a call session, the questions and responses of the agent are guided by a decision tree. The agent first asks if a borrower can make payments under their current plan. If not, the agent "models-out" IDR payments for the borrower, asking about their annual income, family size, and employment status. Borrower responses are entered into the agent's computer, which provides an estimate of monthly IDR payments according to Equation 1.1. The agent then provides the borrower with instructions for online IDR enrollment with the Department of Education.

Agents are incentivized to bring delinquent accounts current, but face penalties if they fail to present borrowers with their best available options. Supervisors periodically monitor agents' calls to ensure they meet federal compliance standards. If an agent does not offer IDR to a borrower deemed suitable for the option, the agent's pay is reduced that month.

1.3 Data

The data I use to estimate IDR link administrative student loan repayment data to credit bureau records for over one million borrowers. Borrowers in my samples are drawn from LLS's FFEL loan portfolio, which includes over $90 billion in loans. The LLS data contain detailed repayment records for each borrower, including principal borrowing amounts, loan balances, minimum payments due, and dates of delinquency at monthly frequency. They also include indicators for type of loan (e.g., Subsidized Stafford, PLUS), current repayment plan (e.g., Standard, IBR), and current loan status (e.g., forbearance, grace period, default). In

addition to loan information, the LLS data contain some borrower characteristics, including year of birth, 9-digit zip code, OPE ID for attended institutions, college attendance dates, and graduation status. Gender is inferred using first names.[16]

LLS data are linked to yearly TransUnion credit bureau records from 2010 through 2017. The TransUnion data provide yearly balances, credit limits, delinquencies, and number of accounts for several categories of consumer debt, including mortgages, credit cards, and auto loans. They also include broader measures of financial health, like credit scores and bankruptcies. Additional details concerning TransUnion data can be found in [46], [13], and [60]. TransUnion data are merged to borrowers in the LLS data by SSN.

The loan data provided for this study consist of two samples of borrowers, each drawn from LLS's FFEL portfolio. The first, which I call the "full sample" ($N = 287,456$), is a random sample of FFEL borrowers with loans disbursed after 1995. While the random sample is not used in my main analysis, descriptive statistics from this sample are representative of the FFEL-borrowing population, and thus provide a useful benchmark to compare against my analysis sample. The second sample, which I call the "call sample" ($N = 164,090$), consists of the universe of LLS's FFEL borrowers who ever received a delinquency call from 2012 onward, excluding those who hold any private or Direct loans.[17]

Borrowers in the call sample are linked to loan servicer contact histories from 2009 onward. Contact history data provide a single observation for each point of contact, and include all incoming and outgoing calls in which the line was connected to a borrower in the sample. For each call in the data, I observe the date, time of day, incoming/outgoing status, and servicing agent identifier associated with the call. Agent identifiers are linked to a small

[16]The online appendix to [120] provides a public-use list of common first names paired with the male-female proportions of New York City Facebook profiles with each name. LLS merged this list to first names in their borrower records at my request.

[17]While borrowers can hold loans from a mixture of FFEL, Direct, and private sources, the database I use only includes repayment information for FFEL borrowers. The Call sample excludes borrowers with mixture of loans, so I can observe their complete repayment profile. Roughly 25% of the servicer's FFEL borrowers also hold Direct loans, and fewer than ten percent hold private student loans.

set of agent characteristics, including work site location and work group ("claims aversion", "skip tracing", etc.).

I construct the instrument and analysis samples on the borrower-by-call level, applying the following selection criteria to all calls placed to borrowers in the call sample: First, I remove borrowers who cannot be matched by zip-code or first name to inferred measures of gender or income, leaving 581,148 calls. To remove calls which may be non-randomly assigned, I remove borrowers marked as non-native English speakers, as well as those more than 140 days delinquent at the time of the call, which leaves 401,088 calls.[18] Then, I remove borrowers who were already enrolled in IDR prior to their delinquency call, as they would not be "eligible" for call-induced IDR take-up. From the remaining group of 362,759 calls, I keep only those which reached the stage at which borrowers were provided information concerning IDR enrollment (i.e., "modeled-out"), leaving 66,248 calls.

While both estimation strategies use balanced panels, I use the larger unbalanced sample to calculate the instrument, excluding those calls made by agents with fewer than 40 total calls. Applying these criteria to the instrument sample reduce measurement error in my instrument because estimates of the mean taken over small cells have high variance. The instrument sample consists of 59,405 calls.

I use monthly and yearly balanced panels centered around delinquency call dates as my analysis samples for repayment and credit outcomes, respectively.[19] For monthly repayment outcomes, I create a balanced panel of 31,113 calls with 15 leads and 10 lags. For yearly credit outcomes, I create a balanced panel of 16,021 calls with 2 leads and 3 lags. The pooled sample of calls from either panel constitutes my analysis sample.

[18]LLS handles "Late Stage" delinquencies of 150 days or more through another division with different agent-assignment procedures.

[19]Note that the calendar-correspondence between monthly and yearly variables can vary dramatically depending on call dates and data collection dates. If a borrower receives a call during the month in which yearly TransUnion data are collected, $t_{year} = 0$ corresponds to $t_{month} \in [-11, 0]$. If she receives the call one month later, $t_{year} = 0$ corresponds to $t_{month} \in [1, 12]$.

Table 2.1 provides summary statistics for the random sample, call sample, and analysis sample. In the full sample, IDR has low take up, with only 4 percent of borrowers enrolled in a plan. That share rises to 13 percent in the analysis sample, as it is constructed to include only borrowers who might benefit from the plan. Unsurprisingly, these borrowers are negatively selected: they have lower credit card limits, higher rates of bankruptcy and live in lower-income zip codes. While negative selection into the analysis sample is partly because borrowers must be delinquent in order to receive a phone call, it also reflects negative selection into IDR; borrowers with high income or low debt balances would not benefit from the plan, and hence would not reach the "modeled out" stage of the delinquency call. Looking within the analysis sample, we see that those who ultimately enroll in IDR are largely comparable to those who do not. IDR enrollees have slightly higher principal borrowing amounts and come from slightly lower-income zip codes, again reflecting characteristics of those who might gain more benefit from IDR.

While the analysis sample is broadly representative of the IDR-eligible population, there are two important caveats concerning external validity. First, while my dataset is nationally representative of borrowers attending college between 1995 and 2010, individuals in my sample are selected along two dimensions: (1) they have at least one instance of late payments, and (2) their loans originated prior to 2010. While the first restriction limits the sample to likely beneficiaries of IDR, the second restriction removes many borrowers for whom we would expect IDR to be most effective, as younger borrowers typically have higher debt-to-income ratios. Estimated effects may therefore be small relative to population-wide effects.

Second, my sample is restricted to individuals eligible for one particular form of IDR. In the U.S., IDR exists through several repayment plans. While all IDR plans follow the same general structure, particulars of the payment calculation formula and forgiveness provi-

sions under each plan can vary.[20] Furthermore, IDR plans include unattractive institutional features like staggered payment adjustments and complicated sign-up procedures. To the extent real-world IDR deviates from an "ideal" system of income-contingent loan repayment, estimates from this paper apply to the former and thus include any effects from these institutional features.[21] Generalizability of my results to alternative IDR plans, existing or hypothetical, are therefore limited.

1.4 Model

To motivate my empirical analysis, in this section I develop a dynamic model of student loan repayment. The model illustrates the various channels through which IDR may affect borrowers' consumption and default decisions.

1.4.1 Setup

Infinitely-lived borrowers are endowed with student debt D_0 and non-student debt assets $A_0 = 0$. Each period, borrowers draw income $y_t \sim N(\mu_t, \sigma)$ and make student debt payments x_t. Individuals can save ($A_{t+1} > 0$) or borrow ($A_{t+1} < 0$) to finance consumption, but face a liquidity constraint L, imposing an upper bound on the amount they can borrow.[22] To simplify exposition, assume no discounting or interest ($\beta = (1 + r) = 1$). The borrower's problem is given by:

[20] Individuals in my analysis sample hold FFEL loans and not Direct loans, which means they are eligible for "Income-Based Repayment" (IBR), and not "Pay-As-You-Earn" (PAYE) or "Revised-Pay-As-You-Earn" (REPAYE) plans. Approximately twenty percent of the five million borrowers in IDR plans are enrolled in IBR ([37]). For a detailed description of each plan, including eligibility criteria and repayment terms, see www.studentaid.gov.

[21] For a more detailed discussion of the limitations imposed by, and potential improvements upon, the current structure of IDR in the U.S., see [53].

[22] While I do not explicitly model default in this model, a simple extension would be to incorporate a floor on consumption whereby borrowers make zero loan payments and incur a penalty to utility in period t if minimum payment x_t exceeds cash-on-hand by some amount, i.e. $x_t > (A_t - A_{t+1} + y_t) + \underline{c}$ In the baseline model, the incidence of binding liquidity constraints can be thought to capture delinquency or default in some loose sense.

$$V\left(y_t, D_t, A_t\right) = \max_{A_{t+1}} \left\{u\left(c_t\right) + E\left[V\left(y_{t+1}, D_{t+1}, A_{t+1}\right)\right]\right\} \tag{1.3}$$

$$\begin{aligned} A_{t+1} &\geq L & (1.4)\\ c_t &= A_t - A_{t+1} + y_t - x_t & (1.5)\\ D_{t+1} &= D_t - x_t & (1.6) \end{aligned}$$

Where minimum payments x_t are determined by a payment scheme modeled as a function of current-period income and remaining student debt:

$$x_t = x(y_t, D_t) \tag{1.7}$$

First order conditions for the borrower's problem yield an Euler inequality:

$$u'(c_t) \geq E\left[\frac{\partial V}{\partial A}(y_{t+1}, D_{t+1}, A_{t+1})\right] \tag{1.8}$$

which binds in period t if and only if $A_{t+1} > L$. The solution to the borrower's problem is characterized by Equations 1.4 through 1.8.[23] Unless prevented by liquidity constraints, individuals will save or borrow to so that the marginal utility of current-period consumption is equal to the expected marginal value of assets next period.

In the following sections, I analyze how a change from Standard ("S") to IDR ("I") repayment schemes affects optimal consumption in the model above, where

$$x^S(y, D) = \min\left\{\frac{D_0}{N}, D\right\} \tag{1.9}$$

$$x^I(y, D) = \min\left\{\theta y, \frac{D_0}{N}, D\right\} \tag{1.10}$$

[23]Proof in Appendix A.2.

For borrowers in repayment, the IDR plan is modeled as the minimum of some share θ of per-period income and the standard payment amount D_0/N. Borrowers under both plans never pay more than their remaining balance D.

1.4.2 Liquidity Effects through Reduced Payments

To isolate the potential channels through which IDR might influence consumption, I first consider alternative specifications for the payment scheme in Equation 1.7 and defer discussion of loan-forgiveness provisions to the following subsection. Let m_t denote income net of loan payments, $m_t \equiv y_t - x_t$. The reduction in minimum payments offered by IDR offers two distinct benefits to borrowers, both of which increase consumption through liquidity. First, IDR weakly increases net income in the short-term relative to the long-term:

$$E\left[m_t^I\right] \geq E\left[m_t^S\right] \quad \forall t \leq N \tag{1.11}$$

$$E\left[m_t^I\right] \leq E\left[m_t^S\right] \quad \forall t > N \tag{1.12}$$

The result is an intertemporal *smoothing effect*: pushing payments farther into the future flattens borrowers' expected net income profiles. As a result, financing consumption in periods $t \leq N$ requires less borrowing, decreasing the likelihood of binding liquidity constraints. With less fear of binding constraints, borrowers place less value on precautionary savings: the marginal utility of assets in Equation 1.8 decreases and consumption increases.[24] Intuitively, individuals face less pressure to borrow because their reduced current-period payments are effectively "borrowed" through higher payments in future periods.

Second, in addition to smoothing income over time, IDR also smooths each period's income over states of the world, providing income insurance for student borrowers.[25] To see

[24] Proof in Appendix A.2

[25] I use the term "insurance" loosely, referring to state-contingent liquidity benefits under IDR. Note that, ignoring loan forgiveness, income-contingent payments provide no insurance against *lifetime*-earnings risk.

how, note that IDR reduces *ex-post* variance in net income:

$$\text{Var}(m_t^I) \leq \text{Var}(m_t^S) = \sigma^2 \quad \forall t \tag{1.13}$$

Reducing income uncertainty increases consumption through an *insurance effect*. This insight comes from the literature on consumption dynamics: individuals facing higher income uncertainty are more likely to face binding liquidity constraints in all periods. A mean-preserving spread in the future income distribution will therefore increase the marginal value of assets and decrease current-period consumption out of choice (precautionary savings) or necessity (the liquidity constraint binds).[26]

The complementary smoothing and insurance benefits of lower minimum payments are summarized in the following expression:

$$\Delta c_t = \underbrace{\sum_{k=0}^{\infty} -\frac{dc_t}{d\mu_k}\Delta E[x_k]}_{\text{smoothing}} + \underbrace{\sum_{k=0}^{\infty} \frac{dc_t}{d\sigma_k}[\sigma_k - \sigma]}_{\text{insurance}} \tag{1.14}$$

Note that assuming quadratic utility ($u''' = 0$) and no debt forgiveness ($\sum_{k=0}^{\infty} \Delta E[x_k] = 0$), removing liquidity constraints would reduce both terms above to zero, as borrowers would simply borrow and save as needed to achieve constant consumption over all periods. In this sense, IDR serves more like a market-correcting measure than a transfer program: Under standard repayment, borrowers who expect future earnings increases would ideally finance current period consumption through borrowing. Market failures may prevent such borrowing, however, as recent college graduates often lack credit or collateral. IDR alleviates this problem by pushing debt obligations further into the future during periods of low income. If borrowers are liquidity constrained under standard repayment, IDR should therefore decrease delinquency rates, increase consumption, or both. Conversely, if borrowers have alter-

[26] See [100]. Proof in Appendix A.2.

native credit options, we would expect student loan repayment and consumption to remain unchanged under IDR. For these unconstrained borrowers, IDR may instead have a deleveraging or savings effect, causing them to spend down higher interest credit cards or increase their savings.

Assessing IDR's potential liquidity effects is important for evaluating its welfare implications. If IDR provides liquidity to borrowers who would otherwise default or sharply reduce consumption, it corrects for an incomplete credit market and may increase total welfare: borrowers benefit from smoother consumption, improved credit, and fewer late fees, and the government saves on the considerable cost of collecting delinquent payments.[27] If, however, borrowers simply use their increased cash-on-hand to pay down existing debt or increase savings, flexible repayment amounts to little more than a redistributive policy transferring interest-savings to student borrowers through taxpayer-subsidized loans.

Observations of the student debt market provide support for potential liquidity benefits. Median monthly payments exceed 14% of median earnings for borrowers beginning standard repayment([101]), and raising limits on credit card balances has twice the expenditure response for those under 44 as those over ([75]). Together, these facts suggest college graduates might benefit from borrowing against future earnings to finance consumption and avoid delinquency in low-income periods.

1.4.3 Wealth Effects through Debt Forgiveness

Note that while the liquidity benefits described above increase borrower welfare, they do not increase lifetime income: reduced payments facilitate wealth transfers *within-borrower* over time, and therefore carry no cost to social welfare. By contrast, the forgiveness provisions

[27]The Congressional Budget Office estimates annual administrative costs of the FFEL student loan program equal 0.67 percent of outstanding balances, or approximately $10 billion per year ([102]).

of IDR, which forgive any outstanding debt after twenty-five years of successful payments, amount to a means-tested government transfer. [28]

To capture the loan forgiveness provisions of IDR, I make a simple modification to the model above. Let T denote the number of periods after which remaining balances are forgiven. In period $t = T + 1$, instead of Equation 1.6, we have:

$$D_{T+1} = 0 \tag{1.15}$$

Note that if borrowers expect loan forgiveness, $\sum_{k=0}^{\infty} \Delta E[x_k] < 0$ and the first term in Equation 1.14 includes a *wealth effect*. In this case, the expression can be decomposed to differentiate between liquidity and wealth effects:

$$\Delta c_t = \underbrace{-\frac{dc_t}{dD_{T+1}} E[D_T^I - x_T^I]}_{\text{wealth effect (+)}} + \underbrace{\frac{dc_t}{dD_{T+1}} E[D_T^I - x_T^I] + \sum_{k=0}^{T} \left(-\frac{dc_t}{d\mu_k} \Delta E[x_k] + \frac{dc_t}{d\sigma_k} [\sigma_k - \sigma] \right)}_{\text{liquidity effects (+)}} \tag{1.16}$$

1.4.4 Social Welfare Implications

While reduced payments and debt forgiveness provisions both increase consumption among IDR enrollees, they hold different implications for social welfare. By providing a ceiling on debt's share of lifetime income, debt forgiveness provides "wealth insurance" against lifetime-earnings risk. As with any social insurance program, however, providing state-contingent transfers carries costs. First, because IDR borrowers pay no "premium" for wealth insurance, the government bears the expected cost of forgiven debts plus some risk premium.[29] Such

[28] Note that IDR may provide income transfers through other channels not included in this model. For instance, the subsidized interest rates on student loans provide a net-present-value benefit to borrowers whose minimum payments drop due to low income.

[29] While insuring *idiosyncratic* risk between borrowers would be diversified away in large numbers, *systemic* risk could be costly to the government if, for instance, uncertainty over the business cycle makes it difficult to predict the amount of forgiven debt.

a transfer would impose a deadweight loss to social welfare through taxation while carrying ambiguous redistributional consequences: forgiveness is progressive conditional on student debt, but college graduates, especially those with high debt balances, have higher lifetime income then the general population. Second, insuring lifetime income for student borrowers can distort labor supply, occupation choice, or college attendance decisions through moral hazard, thereby reducing social welfare. For example, borrowers on IDR may take jobs with higher risk of low pay or unemployment as a result of being partially insured against income losses. By contrast, providing liquidity through reduced IDR payments is revenue neutral and non-distortionary. Income-contingent payments may affect labor supply, but such responses represent a corrective response to incomplete credit markets rather than moral hazard.[30]

A full accounting of IDR's social welfare effects would distinguish between liquidity and wealth effects and measure labor supply, college attendance, and principal borrowing responses through each channel. For several reasons, such an analysis is beyond the scope of this paper. First, IDR has not been available long enough for the twenty-five-year forgiveness period to bind, so assessing the incidence of forgiveness benefits necessarily would necessarily involve some degree of forecasting. Second, labor supply effects, like wages or hours worked, can only be detected insofar as they influence consumption decisions, and the post-graduate timing of IDR enrollment for individuals in my sample means I cannot estimate moral hazard effects on decisions like principal borrowing amounts or institution and college major choice. Third, some degree of wealth insurance is theoretically afforded to all IDR-*eligible* borrowers, not just current enrollees, as individuals on standard repayment plans usually can opt-in to IDR at any time. My research design cannot identify moral hazard through these market-wide effects, as all individuals in my sample are eligible for IDR.

Nevertheless, the repayment path of IDR borrowers in my sample should provide some insight into the relative effects of liquidity versus expected debt forgiveness. The incidence

[30]See [32] for a more complete description of the distinction between moral hazard and liquidity in the context of unemployment insurance.

of debt forgiveness depends on the likelihood of positive balances after twenty-five years of IDR payments. While I cannot directly test for moral hazard or measure realized loan forgiveness, the evolution of payment-to-debt ratios in the two years following IDR enrollment can inform forgiveness likelihood. If IDR payments return to the standard amount within one or two years, it suggests IDR holds little insurance benefit and moral hazard is unlikely to play a large role in borrower behavior. If, instead, payments remain low for a prolonged period following enrollment, the likelihood of loan forgiveness is much higher and IDR could potential carry a high social welfare cost through moral hazard or redistributional inefficiencies.

1.5 Empirical Strategy

Consider the following empirical model of borrower i's outcomes, t periods after receiving delinquency call c:

$$Y_{ict} = \beta_0 + \beta_1 IDR_{ic} + \beta_2 \boldsymbol{X}_{it} + \epsilon_{ict} \tag{1.17}$$

where Y_{ict} denotes the outcome of interest, IDR_{ic} is an indicator for IDR enrollment within three months of the call, $\boldsymbol{X}_{ict}$ is a vector of borrower control variables, and ϵ_{ict} is an error term. Estimating β_1 in Equation 1.17 using OLS would likely yield biased estimates, because preferences over repayment plan choices are correlated with unobserved borrower attributes. For example, borrowers expecting an increase in earnings would likely have a weaker preference for IDR than those with lower earnings expectations. In this case, OLS estimates of IDR's effect on measures of financial well-being would be biased *downwards*, reflecting the selection of individuals low earnings potential into the plan.

The overcome the biases associated with OLS and identify the causal effect of IDR on repayment and credit outcomes, I employ two complementary empirical strategies. First, I estimate the difference-in-differences between IDR-enrollees and non-enrollees before and

after receiving delinquency calls. Second, I use an instrumental variables (IV) design which exploits the quasi-random assignment of servicing agents to these calls.

1.5.1 Difference-in-Differences Specification

My difference-in-differences design compares pre-/post-call differences in outcomes between borrowers who take up IDR (the "treatment" group) and borrowers who remain in standard repayment plans (the "control" group) to identify the causal impact of IDR on repayment and credit outcomes. Formally, the difference-in-differences specification takes the following form:

$$Y_{ict} = \gamma_i + \gamma_t + \left[\sum_{\tau \neq -1} \delta_\tau \cdot IDR_i \cdot \mathbf{1}\{t = \tau\}\right] + \gamma_1 IDR_{ic} + \gamma_2 \boldsymbol{X}_{ict} + \epsilon_{it} \tag{1.18}$$

where Y_{ict} denotes the outcome of interest, γ_i are individual fixed effects, γ_t are event-time fixed effects (months or years relative to call date), IDR_{ic} is an indicator for IDR enrollment within three months of the call, $\boldsymbol{X}_{ict}$ is a vector of borrower control variables (including call date and time fixed effects), ϵ_{ict} is an error term, and δ_t are coefficients on IDR enrollment status which vary non-parametrically by event time. The specification omits α_t and δ_t terms at $t = -1$ to prevent perfect multicollinearity. Estimates can therefore be interpreted relative to the baseline period of one month or year prior to the delinquency call.

Identification in the difference-in-differences specification comes from variation in the propensity to take up IDR following a delinquency call. Such variation can come from a variety of sources, including both between-agent differences captured by the instrumental variables strategy I describe in Section 1.5.2 and *within*-agent variation over calls. While the following section argues that between-agent variation is as-good-as-random, within-agent variation is not. Internal validity of this specification therefore rests upon the assumption that, holding borrower-specific differences fixed, such within-agent variation is exoge-

nous with respect to outcomes. Stated differently, unbiased estimation in my difference-in-differences design assumes that trends in outcomes would be the same in both groups of borrowers had neither taken up IDR.

Figures 1.2 - 1.5 provide graphical evidence in support of the common-trends assumption. The figures plot raw means for monthly repayment and credit bureau outcomes, respectively, for IDR enrollees and non-enrollees, normalized by month- or year-of-call and pre-call mean. Trends in pre-call outcomes appear similar between IDR and standard enrollees for several periods, diverging only after receiving the delinquency call. I also estimate IDR effects in an alternative differences-in-differences specification that controls for group-specific linear trends in months or years prior to call.[31]

Even if IDR and standard borrowers exhibit observably similar pre-trends, other potential violations to the exclusion restriction exist. For example, IDR-enrollees may have experienced a shock at the time of their delinquency call that both induced them into IDR take-up while also influencing outcome variables. I argue that such instances are unlikely. Delinquency calls are *outgoing*, so their incidence is determined by LLS, not the borrower. While the timing of these calls are mechanically non-random *within-borrower*, they do not vary systematically *between* borrowers with observably similar debt characteristics. If IDR borrowers were enrolling as a response to sudden shocks, outcomes should therefore vary from non-IDR borrowers in the months immediately preceding the call.

It is possible that borrowers exhibit the same or similar trends before delinquency calls but make IDR enrollment decisions based on expected *future* shocks to their financial well-being. While my difference-in-differences design cannot rule out this scenario, any bias created by

[31]Estimates from the specification including linear pre-trends can be interpreted as IDR's impact on outcomes relative trend-predicted differences between groups. Formally, the model is given by

$$Y_{ict} = \gamma_i + \delta t \cdot IDR_i \cdot \mathbf{1}\{t < 0\} + \left[\sum_{\tau \geq 0} \delta_\tau \cdot IDR_i \cdot \mathbf{1}\{t = \tau\}\right] + \gamma_1 IDR_i + \gamma_2 \boldsymbol{X}_{ict} + \epsilon_{it} \tag{1.19}$$

forward-looking borrowers should be negative: the benefit of IDR is strictly decreasing in income and available credit, so borrowers who select into IDR based on expected shocks to their financial well-being should, all else equal, exhibit *lower* repayment, consumption, and credit scores relative to standard borrowers, attenuating any positive treatment effects of IDR.

Finally, the quasi-experimental nature of my instrumental variables design provides an informal robustness check to my differences-in-differences design. While differences between the two approaches can be partly attributable to heterogeneous treatment effects between their respective complier populations, similar IV and difference-in-differences estimates would suggest that the variation driving the latter is plausibly exogenous. I make this comparison for monthly repayment outcomes, where IV estimates are sufficiently precise, to lend further credibility to difference-in-differences estimates of IDR's effect on credit outcomes, where my IV approach is too underpowered to conduct meaningful inference.

1.5.2 Instrumental Variables

I complement the difference-in-differences design described above with an instrumental variables (IV) approach which isolates quasi-random variation in IDR enrollment. The instrument used in my IV design is a measure of a quasi-randomly-assigned servicing agent's tendency to induce IDR take-up among the borrowers she calls. I interpret observed differences in outcomes between borrowers assigned to different agents as the causal impact of the change in IDR take-up associated with those agents. The parameter I estimate in this specification is a local average treatment effect (LATE), which identifies the impact of enrolling in IDR among borrowers whose plan choice depends on agent-assignment.

Instrument Calculation I construct my instrument using a residualized, leave-one-out measure of agents' ability to induce IDR take-up that accounts for the timing and ordering of

delinquency calls. This "agent score" is analogous to the measures of judge leniency used by [35] and [44] to instrument for trial outcomes in court case settings. The approach solves two problems one might see in a non-residualized instrument. First, the start and end dates of an agent's employment with LLS might coincide with time trends in IDR take-up unrelated to the agent's ability. For instance, an agent employed before IDR became a popular repayment option might incorrectly be measured as "low-score". Second, the shift choices of agents are non-random and may correlate with borrowers' IDR take-up if, for example, borrowers are more likely to hang up on calls received in the evening or on weekends.

I account for these potential selection effects by removing year, month, weekend, and hour-of-day fixed effects as well as a linear time trend from observed IDR enrollments before calculating the leave-out agent score. Specifically,

$$IDR^*_{ict} = IDR_{ict} - \gamma \boldsymbol{W}_{ict} \tag{1.20}$$

$$= Z_{ijc} + \epsilon_{ict} \tag{1.21}$$

where $\boldsymbol{W}_{ict}$ is a vector of call date and time controls and Z_{ijc} is agent score. I calculate residualized rate of IDR take-up, IDR^*_{ict}, using OLS estimates of γ in Equation 1.20. I then construct agent score Z_{ijc} using the leave-one-out mean of this residualized rate:

$$Z_{icj} = \left(\frac{1}{n_j - 1}\right)\left(\sum_{k=0}^{n_j} IDR^*_{kj} - IDR^*_{ij}\right) \tag{1.22}$$

where n_j denotes the number of calls made by agent j. Removing borrower i's outcomes from the calculated mean is important to avoid reflection bias: including one's own outcome in her assigned agent score would introduce the same estimation errors on both left- and right-hand side of the regression, incorrectly attributing unobserved between-borrower differences in IDR take-up to agent ability. My two-stage least-squares estimation avoids this problem

by using the leave-one-out agent score, calculated as in Equation 1.22, as an instrumental variable for whether a borrower enrolls in IDR.

The residualized agent score distribution can be seen in Figure 1.6. The analysis sample includes calls from 144 different agents in four different call centers. Agents place 216.06 calls on average to borrowers in the analysis sample, with a median of 183 calls.

Identifying Assumptions In order for two-stage least squares estimates to identify a local average treatment effect (LATE) for the causal impact of IDR take-up, the instrument must satisfy three conditions: (1) IDR take-up varies with agent assignment, (2) agent assignment predicts borrower outcomes only through its effect IDR take-up, and (3) agents' tendency to induce IDR take-up is monotonic across borrowers.

To test the first identifying assumption, I empirically evaluate the first-stage relationship between the agent score instrument and observed IDR take-up. Figure 1.6, which plots a local linear regression of IDR take-up against the agent score instrument, provides graphical evidence for this relationship. IDR take-up is monotonically increasing in agent score throughout the distribution of borrowers. The first-stage relationship seen in Figure 1.6 can be represented through following linear probability model:

$$IDR_{ict} = \alpha_0 + \alpha_1 Z_{ict} + \alpha_2 \boldsymbol{X}_{ict} + \epsilon_{ict} \tag{1.23}$$

In Table 1.3, I provide first-stage estimates from an OLS regression of Equation 1.23. Results imply a positive and highly significant relationship between agent score and IDR take-up. Point estimates range from 0.62 to 0.60, depending on the inclusion of borrower controls, and the F-statistic on a test of instrument significance equals 21.89.

The second identifying assumption requires that agent assignment be predictive of borrower outcomes only through its impact on repayment plan choice. One way this assumption could be violated would be if different types of borrowers were systematically assigned par-

ticular agents. Table 1.2 provides empirical evidence that, conditional on call date and time, borrowers do not vary systematically by agent score. Column 1 reports results from an OLS regression of realized IDR take-up against several borrower characteristics and pre-call outcome variables, as well as call date and time fixed effects. Not surprisingly, estimates demonstrate considerable selection into IDR; holding date and time of call fixed, IDR enrollees are significantly more likely to be young, female, low-income, and hold lower balances across several types of debt. Column 2 reports results from an OLS regression of the agent score instrument on the same right-hand side variables. Estimated coefficients on borrower variables in this specification are statistically indistinguishable from zero, and the F-statistic on a test for whether all borrower variables can jointly predict agent score is 1.17.

Even if agents are randomly assigned to borrowers, the exclusion restriction may still be violated if agents can influence borrower outcomes through channels other than repayment plan choice. If, for example, agents who induce high IDR take-up also convince borrowers to make timely payments, two-stage least squares estimates of IDR's effects on repayment would be biased upwards. While it is impossible to rule out agent effects through non-IDR channels, the institutional details in my research setting provide suggestive evidence that such threats to validity are minimal. LLS's delinquency calls are designed solely to provide borrowers with information on their repayment options. Agents provide no advice or counseling to borrowers, and follow a decision tree to present repayment alternatives.

The third identifying assumption requires monotonic agent effects across borrowers. To satisfy the monotonicity assumption, there can be no borrower for whom a higher (lower) score agent decreases (increases) the likelihood of IDR take-up; any borrower who declines the IDR option under a high agent must not be induced into IDR under a comparatively low score agent. This assumption would be violated if certain agents "match" well with certain borrowers. For example, if some borrowers respond more favorably to female agents, their take-up may be higher under low-score female agents compared to high-score male agents.

Agent Variation Variation in agent score can be driven by several potential sources. First, while agents ask questions and provide responses according to a standardized decision tree, several nodes of this tree require agent judgment. For example, before informing borrowers of alternative repayment options, an agent must first ask if a borrower is "capable of making payments under her current plan." This question seldom elicits an unambiguous "yes" or "no" response, so agents must make this determination based on their own judgment. To the extent agents are afforded discretion in their responses to borrowers, their decisions are undoubtedly influenced by the incentive scheme discussed in Section 1.2.2. For instance, more risk averse agents may present the IDR option to a broader swath of borrowers to avoid potential compliance infractions. Second, variation may exist in agents' perceived geniality. Even agents reciting the same words can vary in IDR-conversion through subtle variations in demeanor or tone, and borrowers often hang up or stop listening depending on the interaction. Similarly, while call data are not detailed enough to verify, conversations with agent supervisors suggest that factors like speech patterns and accents play a large role in keeping borrowers on the phone long enough to advise them about IDR. Third, agent score may be influenced by agents' ability to provide clear details regarding plan payments and sign-up instructions. Borrowers must log into the Department of Education website using their social security number, authorize the IRS to transfer their tax return, correctly identify their loan program, and consent to change their payment plan. If an agent fails to properly explain these steps, a borrower may fail to enroll in IDR even if the agent convinces her to do so.

1.6 Results

1.6.1 Repayment Outcomes

Table 1.4 reports difference-in-differences and IV estimates of IDR's effect on minimum payments, changes in loan balances, and indicators for 10 or more and 90 or more days delinquent. For each outcome listed on the left-hand side, Columns 2-4 provide estimated coefficients on the interaction between IDR take-up and consecutive three-month periods from the pooled difference-in-differences specification given by Equation 1.18. Columns 5-7 provide estimated coefficients on IDR take-up from separate two-stage least squares regressions in each three-month period. [32] Figure 1.8 plots coefficients on these outcomes using the same specification, but separated by each month following the date of the delinquency call.

Minimum Payments IDR's effect on minimum payment is mechanical: given income, family size, and debt balance, one could directly calculate a borrower's IDR payments using Equation 1.1. Estimating the change in payments for compliers in my sample, however, can provide insight into the mechanisms behind effects on other outcomes. If we saw no effect on payments, any observed effects on delinquencies or consumption would be entirely attributable to the behavioral effects of expected loan forgiveness.[33]

In the difference-in-differences analysis, required payments for those on IDR fall by $140 relative to those who remain on standard repayment. In the top-left panel of Figure 1.7, payments remain relatively stable in the twelve months following the delinquency call, then sharply increase to their pre-call levels, so that IDR and standard borrowers face similar

[32] Results from the first three months following the call are omitted from Table 1.4 because it typically takes one or two months following contact to process and enroll borrowers in IDR, and the relative timing of successful enrollment, next payment due date, and data collection date at the end of the calendar month adds further lag time before IDR effects can be realized.

[33] The expectation of future liquidity benefits could technically have a behavioral response as well, though upward-sloping wage profiles, and the ability to opt-into IDR at any point in the future make "expected-liquidity" channels unlikely.

minimum payments fourteen to eighteen months following their delinquency calls. The convergence in payments corresponds to the one-year recertification period for IDR, suggesting that IDR-enrollees in my sample do not extend benefits beyond one year. IV estimates of IDR's effect on payments are similar to difference-in-difference estimates. For those induced into IDR by their servicing agents, minimum payments lower by an average of $200 four to seven months following IDR take-up and exhibit the same rapid return to standard levels one year following the delinquency call, as seen in the top left panel of Figure 1.8.

Delinquency I measure IDR's impact on two measures of delinquency: the likelihood of falling ten or more days delinquent, and the likelihood of falling ninety or more days delinquent. The ten and ninety day benchmarks are points of increased penalties: at ten days past due, borrowers begin to accrue late fees for delinquent loans, and at ninety days past due, borrowers are reported to credit bureaus. Figures 1.7 and 1.8 show a large negative effect of IDR enrollment for both delinquency measures. In the difference-in-differences analysis, IDR borrowers are 21 percentage points less likely to fall ten or more days delinquent relative to standard borrowers in the four to seven months following the delinquency call, with a pre-call mean of 22pp. Corresponding estimates for months seven through nine and ten through twelve are -16pp and -4pp, respectively. Estimates for ninety or more days delinquent exhibit a similar pattern: the likelihood of falling ninety or more days delinquent falls by 8pp, 8pp, and 7pp in months four through six, seven through nine, and ten through twelve after the delinquency call, respectively, compared to a pre-call mean of 2pp. IV estimates for the effect of IDR on delinquencies are qualitatively similar to difference-in-differences estimates, though less precise and larger in magnitude.

Change in Balances While minimum monthly payments measure the amount borrowers are required to pay to avoid late penalties, the high incidence of delinquency means that payments *due* provides a poor measure of payments *made*. I estimate effects on monthly

changes in remaining balances to provides a sense of how IDR might impact borrowers' propensity to pay down their debts. Both IV and difference-in-differences estimates of IDR's effect on changes in debt balances suggest that, despite lowering required payments, IDR significantly increases the amount repaid. While not statistically significant, point estimates for the instrumental variables analysis suggest IDR decreases monthly balances by $90 in months four through six and $80 in months seven through nine after the delinquency call, off of a pre-call mean of $130. Difference-in-difference estimates are qualitatively similar and statistically significant: relative to standard borrowers, IDR enrollees pay down $70 to $90 more debt each month than they did before their delinquency calls.

1.6.2 Credit and Consumption Outcomes

For evaluating IDR's impact on credit and consumption outcomes in the TransUnion data, I rely principally on difference-in-difference estimates, as instrumental variables estimates are too low-powered to draw meaningful inference.[34]

Table 1.5 provides estimates of the effect of IDR on credit score, mortgage, and auto loan from the difference-in-differences analysis described in Section 1.5.1. For each outcome listed on the left-hand side, Columns 2-4 report coefficients on IDR in consecutive years from the pooled regression specified in Equation 1.18, beginning with the year of the delinquency call. Columns 5-7 report IDR coefficients from a regression which omits pre-call month dummies and includes a linear time trend.

Relative to those who remained in standard repayment, borrowers who enrolled in IDR experienced a statistically significant 5.01-point increase in credit scores within one year of the delinquency call, off of a pre-call mean of 593.26, followed by increases of 7.53 and 5.93 points one and two years after the call. Estimates of IDR's effect on bankruptcy filings and auto loans are statistically indistinguishable from zero for all three years following the

[34]For completeness, instrumental variables estimates for IDR's effect on credit outcomes are reported in Appendix A.1.

delinquency call. IDR effects on the likelihood of holding a mortgage are also effectively zero in the year of the call, but increase gradually throughout the sample period, rising to 2 percentage points by year two, off of a pre-call mean of 20pp, an increase of 10 percent.

Table 1.6 provides difference-in-difference estimates for the effect of IDR on credit card balances, number of credit cards, and credit card limits. While I find no significant effects in the year of the delinquency call, IDR is associated with statistically significant increases in all three credit card measures one and two years following the call. Compared to standard borrowers, total balances on credit cards held by IDR enrollees increase by $240 (0.33 log points) and $400 (0.38 log points) one and two years after the delinquency call off of a pre-call mean of $1420, corresponding to an increase of 28 percent. Similarly, by the second year of enrollment, IDR borrowers hold 0.14 more credit cards (pre-call mean of 3.00) and have $340 higher credit limits (pre-call mean of $2920) compared to those who remained in standard repayment plans. Estimates of IDR's effect of credit limits are modest and not statistically significant, suggesting liquidity effects are driven principally by the direct increase in cash-on-hand from lower loan payments rather than an increased ability to borrow on credit, at least within the first two years of IDR enrollment.

1.7 Interpretation

In this section, I interpret my results by placing them in the context of the model laid out in Section 2.4. In short, results are consistent with a pure liquidity effect of IDR: reducing short-term payments increases borrowers' cash on hand, allowing them to increase consumption and avoid default in periods of low income.

Minimum Payments

Ignoring the potential effects of debt forgiveness, estimates of IDR's effect on minimum payments are analogous to a "first stage" in an estimation of the effects of increased liquidity:

a necessary condition for IDR to generate observable liquidity effects is that it measurably reduces payments during my sample period. While the alternative scenario seems unlikely, it is possible for high-income borrowers to sign up for IDR solely in anticipation of future income shocks. My results run counter to this hypothetical: the large and significant observed effect on minimum payments suggests increased liquidity drives at least part of results on other outcomes.

Note that while the estimated effects of IDR on minimum payment provide valuable insight into the mechanisms driving IDR's effects on other outcomes, the more specific question of how IDR changes minimum payments can be directly measured, at least for the treatment group in my sample, because the counterfactual standard loan payments for IDR enrollees is a deterministic function of principal debt amounts and payment histories up until point they switch to IDR. More simply, the effect of IDR on payment size is approximately given by observed IDR payments minus payments in the month prior to receiving the delinquency call. Figure A1 provides a graphical illustration of this measured payment effect across the distribution of IDR enrollees in my analysis sample.

Having established that IDR borrowers receive an increase in short-term cash-on-hand, results from delinquency and credit outcomes provide evidence for how borrowers respond to this cash infusion. Recall that borrowers with reduced payments have three options: (1) keep consumption constant and add to their savings, (2) repay their student debt in a more timely manner, or (3) increase short-term consumption. While these options are not mutually exclusive, a repayment and/or consumption response is consistent with liquidity constraints and suggestive of social welfare improvement under IDR. My results provide evidence of both responses: the negative effect on delinquencies and positive effect on credit cards and mortgages suggest borrowers use the liquidity provided by IDR to prevent default, increase non-durable consumption, and purchase homes.

Delinquencies and Balances

Decreased incidence of delinquencies following IDR take up holds two important implications. First, it provides a pathway for increased borrower welfare under IDR, as non-repayment carries negative consequences described in Section 2.2. Note, however, that the magnitude of welfare effects through this channel are difficult to measure, as delinquencies serve only as a noisy indicator for the myriad of consequences associated with various stages of non-repayment. Second, reduced delinquencies under IDR speaks to the long-standing debate over the determinants of default. Increased repayment following a reduction in minimum payments is suggestive of a liquidity motive for default rather than a "stragetic" motive, as lower payments should not influence strategic default decisions.[35]

In theory, IDR could affect balances on student loans in either direction. Lower required payments mean non-delinquent IDR borrowers pay down less of their debt than non-delinquent borrowers on standard repayment plans, increasing balances. However, the reduced likelihood to default while on IDR has the opposite effect on balances. My results suggest the effect of reduced minimums on loan balances is dominated by more timely repayment; IDR borrowers pay down $90 more student debt each month, on average, than those on standard repayment plans.

Credit Outcomes

The positive estimated effects of IDR on credit card balances and number of credit cards are suggestive of a consumption response to increased liquidity, but interpreting these results requires two important caveats. First, I have loosely referred to credit card balances as "proxies for consumption," but balances reported in the credit bureau data capture both flows in credit card spending as well as the stock of unpaid debt. For this reason, the

[35]I interpret "strategic default" as default decisions driven by total outstanding liability, which does not include decisions driven by expected future liquidity constraints. Also note that IDR may affect strategic default through loan forgiveness. I rule out forgiveness effects later in the section.

parameter I identify is not a marginal propensity to consume (MPC) out of liquidity. To the extent lower required payments have a deleveraging effect allowing borrowers to pay down existing debt on credit cards, I likely underestimate the true consumption response to IDR.

Second, the liquidity effects I estimate may be driven by multiple mechanisms. While the immediate increase in cash-on-hand borrowers receive through IDR is the primary channel through which we would expect increases in consumption or homeownership, the increased credit-access associated with higher credit scores provides a potential secondary channel through which individuals might be affected. While this credit channel probably accounts for some of my results, especially over the two and three year time horizons, the estimated impact on credit limits is modest or null, suggesting liquidity effects are driven principally by the direct increase in cash-on-hand from lower loan payments rather than an increased ability to borrow on credit, at least within the first two years of IDR enrollment.

1.7.1 Loan Forgiveness

By increasing expected lifetime income, debt forgiveness provisions under IDR can also affect borrower behavior. If borrowers expect their loans to be forgiven, they may increase short-term consumption regardless of liquidity constraints. While expected loan forgiveness could, in theory, drive effects of IDR on repayment or consumption, my results are more consistent with a pure liquidity effect. In order for the forgiveness period to bind, a borrower must make 300 complete monthly loan payments on IDR and still hold an outstanding balance on her student loans. Minimum payments under standard plans would pay off balances after just 120 payments. The rapid return of minimum payments to standard levels therefore suggests forgiveness is very unlikely for borrowers in my sample. Importantly, this rapid rise in payments does not necessarily imply an increase in income; borrowers may qualify for continued reduced payments but not recertify their income after twelve months of IDR. Even so, such borrowers should not expect loan forgiveness. It is possible that payments

may drop again in the future, making debt forgiveness more likely. Forgiveness effects on consumption under these circumstances would require one of two scenarios, both of which I argue are unlikely. First, if rising payments reflect increased incomes, payments might fall again in the future if income declines more permanently. While I cannot rule out this possibility, it seems unlikely, as income profiles do not generally exhibit sharp, transitory *upward* shocks followed by permanent decreases. Second, if rising payments reflect a failure to recertify income, borrowers may have permanently low incomes through the end of the sample period and choose to reenroll in IDR at a later date. These borrowers may qualify for forgiveness if they successfully make enough reduced payments following their reenrollment. However, if consumption were effected through the expectation of future forgiveness via potential future reenrollment, one would expect those effects to exist for both treatment and control borrowers in my sample, as those who remain in standard repayment plans following the delinquency call are aware of the forgiveness benefit they would receive from IDR should they decide to enroll at a later date.

1.7.2 Alternative Mechanisms

While my results suggest that IDR improves welfare by providing liquidity to credit-constrained borrowers, I cannot rule out all alternative mechanisms driving the positive effect on repayment and consumption. Behavioral mechanisms, in particular, may be consistent with my results while holding different welfare implications. For example, myopia can drive effects on consumption upwards while reducing borrower welfare if increased cash-on-hand induces overspending in the short-run. Likewise, enrolling in a new repayment plan may make a borrower's debt more psychologically salient, prompting her to keep current on payments even if the plan provides no real benefit.

Sixty-nine percent of IDR-enrollees in my sample face IDR payments of zero dollars and thus cannot fall delinquent on their loans. While this mechanical result would still

be characterize as a "liquidity effect" under a neo-classical model, to the extent borrowers facing payments of $\epsilon > 0$ dollars would face higher delinquency rates than zero-payment borrowers, the result may be driven in part by pyschological frictions or "hassle-costs." To attenuate this mechanical effect, I conduct my difference-in-differences analysis for repayment among a subsample of individuals with predicted nonzero IDR payments using the following procedure: First, I regress an indicator for positive IDR payments among IDR enrollees on a full set of demographic controls and pre-call student loan and credit variables. Second, I use these estimates to predict the likelihood of having positive IDR payments among *all* borrowers in my analysis sample. Finally, I conduct my difference-in-differences analysis on the subgroup of individuals with greater than fifty percent predicted likelihood of positive IDR payments. Realized IDR payments are nonzero for more than seventy percent of treated individuals in this subsample, yet the repayment effects of IDR persist. Appendix Table A1 reports delinquency results for this subsample, and continues to find a large and significant effect on repayment rates.

1.8 Conclusion

In this paper, I use new administrative data from a large student loan servicing company and exploit two sources of plausibly exogenous variation from servicing phone calls to isolate the causal effect of IDR enrollment on default, bankruptcy, consumption proxies, and household balance sheets. I find evidence of increased liquidity on both default and consumption margins. IDR decreases delinquency rates by 16 percentage points within seven months of take up. Relative to borrowers who remain on standard repayment plans, IDR enrollees have credit scores which are 7.53 points higher, hold 0.1 more credit cards, and carry $240 higher credit card balances one year after the servicing call, implying increased short-term consumption out of liquidity. Minimum monthly payments decrease by an average of $140 following IDR take up, but return to standard levels within one year of enrollment, sug-

gesting limited forgiveness incidence and minimal moral hazard effects. My results suggest IDR improves borrower welfare principally through a liquidity channel, providing short-term increases to cash-on-hand in periods of financial distress.

The results of my analysis raise several questions for future research. First, while this paper establishes the benefits of IDR on consumption, the same benefits may carry over into other outcomes. Income, employment, and occupation choice might all be affected by reduced loan payments, especially since liquidity benefits are concentrated in the years after college when labor market decisions are most consequential. Likewise, IDR may have ex-ante effects on borrowers' college attendance decisions, major choices, or principal borrowing amounts. In both cases, the relative effects of liquidity and forgiveness hold important implications for social welfare: a borrower may make "riskier" career decisions because she can borrow against future income, or because her future income is insured.

Second, while the apparent lack of forgiveness-eligible borrowers in my sample provides clear evidence of liquidity benefits, it also leaves unanswered questions regarding moral hazard and student debt forgiveness more generally. A number of existing and proposed student loan programs offer more generous forgiveness provisions than those under the IDR plan studied in this paper. For instance, teachers and public service employees can apply for debt forgiveness after only ten years of payments ([36]), and President Trump has proposed a version of IDR with a fifteen-year forgiveness period ([48]). Further empirical work on loan forgiveness is needed to assess the welfare benefits and potential distortionary effects of these policies.

Third, evidence for IDR's benefits raises the natural question of why so few students enroll in the plan. According to classic economic theory, IDR should be weakly preferred to standard repayment for any borrower. There are no pecuniary costs to enrolling, yet my results suggest many individuals who would benefit from IDR do not take it up. There are a number of potential reasons for this low take-up, such as "hassle-costs" of enrollment,

inattention, or incomplete information.[36] Further research into these psychological frictions may reveal policy interventions to help borrowers take advantage of the benefits offered by IDR.

1.9 Tables and Figures

[36]Prior work has demonstrated the importance of information in similar contexts, including principal borrowing decisions for student loans ([104]) and college major choice ([76]).

Table 1.1: Summary Statistics

	Full Sample	Analysis Sample		
	(1) Pooled	(2) Pooled	(3) non-IDR	(4) IDR
Panel A: LLS Data				
IDR	0.0405	0.141	0	1
Female	0.570	0.713	0.708	0.742
Zip Median Income	59.23	52.25	52.42	51.24
Age	32.92	43.26	43.28	43.14
Amount Borrowed	19.11	27.48	27.09	29.84
10+ Days Delinquent	0.500	0.998	0.999	0.995
90+ Days Delinquent	0.287	0.737	0.754	0.635
Days Delinquent	93.99	167.3	171.6	141.0
Number of Calls		5.183	5.328	4.298
Panel B: Credit Data				
Credit Score	676.3	588.9	587.6	596.4
Bankruptcy	0.127	0.196	0.194	0.207
Derogatory Rating	0.331	0.553	0.551	0.568
Number of Credit Cards	5.020	3.081	3.066	3.169
Credit Card Balances	4.104	1.581	1.582	1.575
Number of Mortgages	1.156	0.682	0.691	0.631
Mortgage Balances	67.25	25.87	26.68	20.94
Credit Card Limits	20.03	4.835	4.839	4.806
Number of Auto Trades	1.923	1.613	1.618	1.584
N	148666	24985	21452	3533

Note: This table reports summary statistics at the borrower level. The Full sample is a random sample of the population of borrowers in LLS's FFEL portfolio who made any loan payments from 2010 onward. The analysis sample is a subsample from the same population, selected according to the following criteria: borrowers received an IDR-modeled delinquency call from 2009 onward, have observable repayment histories three years years prior and two years following the phone call, hold no private or Direct loans, are not recorded as non-english speakers, and were assigned to agents with at least 40 observed phone calls. IDR is and indicator for whether the borrower ever enrolled in IDR. Female is a measure of likelihood-female inferred from first name following [120]. Initial zip income is the median 2010 income for the borrower's 5-digit zip code as of her first payment. Days delinquent is the maximum number of days the borrower was ever past due on payments as of the end of the month, and ever delinquent is an indicator for whether days delinquent is greater than 10. Number of calls is the total number of outgoing calls recorded for the borrower. Credit scores, bankruptcies, derogatory ratings, credit card, mortgage, and auto loan information are taken from TransUnion credit bureau data.

Table 1.2: Balance Test

	(1) Agent Score∗100	(2) IDR∗100
Female	0.0192267	1.9057462***
	(0.0364808)	(0.4107607)
Amount Borrowed	−0.0002625	−0.0753578***
	(0.0019822)	(0.0215861)
Age	0.0000204	−0.0349901
	(0.0014395)	(0.0235802)
Lag Zip Median Income	−0.0020918***	−0.0265073***
	(0.0007458)	(0.0094746)
Lag Minimum Payment	0.2533364	6.3175518***
	(0.1757279)	(1.9366660)
Lag Remaining Balance	0.0006257	0.0869712***
	(0.0011584)	(0.0160278)
Lag Credit Score	−0.0005425**	0.0356230***
	(0.0002723)	(0.0040647)
Lag Credit Card Balances	−0.0069284	−0.0396615
	(0.0054446)	(0.0724199)
Lag Number of Auto Trades	−0.0025185	−0.2650774**
	(0.0078358)	(0.1166223)
Lag Any Mortgage	0.0469042	−0.9608241
	(0.0554542)	(0.8276821)
Lag Any Credit Card	−0.0058517	0.2461155
	(0.0362830)	(0.5559533)
Lag Mortgage Balances	−0.0005516	−0.0108382**
	(0.0004285)	(0.0047757)
Lag Number of Credit Cards	−0.0079420	0.0485748
	(0.0050360)	(0.0868937)
Lag Credit Card Limits	0.0035912*	−0.0407309
	(0.0021653)	(0.0392379)
Mean Dep.	-0.096	10.864
F-stat	1.17	15.86
P-value	0.3029	0.0000
R-squared	0.014	0.016
N	31113	31113

Note: This table reports balance test results. The regressions are estimated on the call sample described in the notes to Table 2.1. Agent score is estimated using data from other phone calls placed by the same agent following the procedure described in Section 1.5.2. Column 1 reports the estimated coefficients from an OLS regression of agent score multiplied by 100 against the variables listed and call year, month, and hour fixed effects. Column 2 reports estimates from an identical regression, except with the dependent variable equal to realized IDR take-up as of six months after the call, multiplied by 100. Robust standard errors two-way clustered at the borrower and agent level are reported in parentheses. The p-value reported at the bottom of columns 1-2 is for an F-test of the joint significance of the variables listed in the rows. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 1.3: First Stage

	(1) IDR	(2) IDR
Agent Score	0.6179307***	0.6042973***
	(0.1320810)	(0.1309608)
Female		0.0189413***
		(0.0040865)
Amount Borrowed		−0.0007520***
		(0.0002164)
Age		−0.0003500
		(0.0002361)
Lag Zip Median Income		−0.0002524***
		(0.0000948)
Lag Minimum Payment		0.0616446***
		(0.0194076)
Lag Remaining Balance		0.0008659***
		(0.0001603)
Lag Credit Score		0.0003595***
		(0.0000403)
Lag Credit Card Balances		−0.0003547
		(0.0007229)
Lag Number of Auto Trades		−0.0026356**
		(0.0011642)
Lag Any Mortgage		−0.0098917
		(0.0082987)
Lag Any Credit Card		0.0024965
		(0.0055521)
Lag Mortgage Balances		−0.0001050**
		(0.0000480)
Lag Number of Credit Cards		0.0005337
		(0.0008721)
Lag Credit Card Limits		−0.0004290
		(0.0003948)
Mean Dep.	0.109	0.109
F-stat	21.89	21.29
P-value	0.0000	0.0000
R-squared	0.010	0.018
N	31113	31113

Note: This table reports first stage results. The regressions are estimated on the call sample described in the notes to Table 2.1. Agent score is estimated using data from other phone calls placed by the same agent following the procedure described in Section 1.5.2. IDR is an indicator for IDR take-up as of six months after the call. Robust standard errors two-way clustered at the borrower and agent level are reported in parentheses. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 1.4: Difference-in-Differences and Instrumental Variables Estimates of the Effect of IDR on Loan Repayment Outcomes

		Difference-in-Differences			Instrumental Variables		
Dependent Variable	(1) Mean $t = -1$	(2) Mos. 4 to 7	(3) Mos. 8 to 11	(4) Mos. 12 to 15	(5) Mos. 4 to 7	(6) Mos. 8 to 11	(7) Mos. 12 to 15
Minimum Payment	0.22	-0.14^{***}	-0.14^{***}	-0.05^{***}	-0.20^{***}	-0.22^{***}	-0.14^{**}
		(0.00)	(0.00)	(0.00)	(0.02)	(0.03)	(0.06)
Remaining Balance	27.83	-0.04	-0.38^{***}	-0.29^{***}	-0.27	-1.35	-1.32
		(0.03)	(0.04)	(0.05)	(0.79)	(1.21)	(1.37)
Δ Remaining Balance	0.13	-0.09^{***}	-0.08^{***}	0.07^{***}	-0.15	-0.11	0.03
		(0.01)	(0.01)	(0.01)	(0.16)	(0.15)	(0.13)
10+ Days Delinquent	0.22	-0.21^{***}	-0.16^{***}	-0.04^{***}	-0.37^{**}	-0.29^{***}	-0.09
		(0.00)	(0.00)	(0.00)	(0.15)	(0.11)	(0.13)
90+ Days Delinquent	0.02	-0.08^{***}	-0.08^{***}	-0.07^{***}	-0.13	-0.07	0.06
		(0.00)	(0.00)	(0.00)	(0.10)	(0.09)	(0.09)
Call Time FE		Yes	Yes	Yes	Yes	Yes	Yes
Controls		Yes	Yes	Yes	Yes	Yes	Yes
N	31113	124447	124447	124447	124447	124447	124447

Note: This table reports difference-in-differences and two-stage least squares estimates of the effect of IDR enrollment on monthly loan repayment outcomes. Column 1 reports the dependent variable mean in the month prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive three-month periods, beginning with four months following the delinquency call from the pooled OLS regression specified in Equation 1.18. Each of Columns 5-7 report estimates from separate two-stage least squares regressions on outcomes in the same months. Regressions are estimated on the analysis sample as described in the notes to Table 2.1. Two-stage least squares models instrument for IDR with the agent score calculated using data from other calls made by the agent following the procedure described in Section 1.5.2. All specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 1.5: Difference-in-Differences Estimates of the Effect of IDR on Financial Outcomes

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Score	593.26	5.01***	7.53***	5.93***	4.95***	7.48***	5.87***
		(1.06)	(1.40)	(1.63)	(1.53)	(1.82)	(1.98)
Bankruptcy	0.19	−0.00	−0.00	−0.00	−0.00	−0.01	−0.01
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Any Mortgage	0.20	0.00	0.02**	0.02***	0.00	0.02**	0.02***
		(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Any Auto Trade	0.71	−0.00	−0.00	0.01	−0.00	−0.00	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

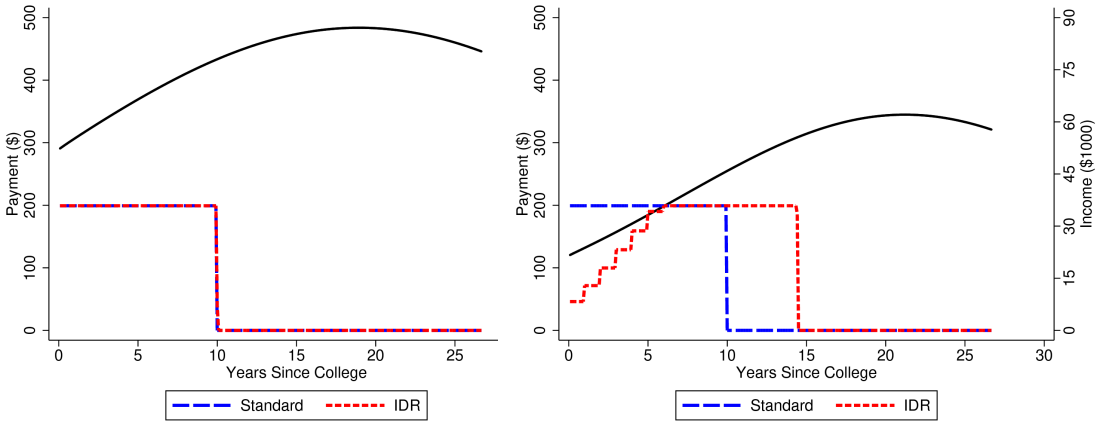
Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on yearly financial outcomes. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same yearly effect for a regression which omits pre-call year dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 124,447 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table 1.6: Difference-in-Differences Estimates of the Effect of IDR on Credit Cards

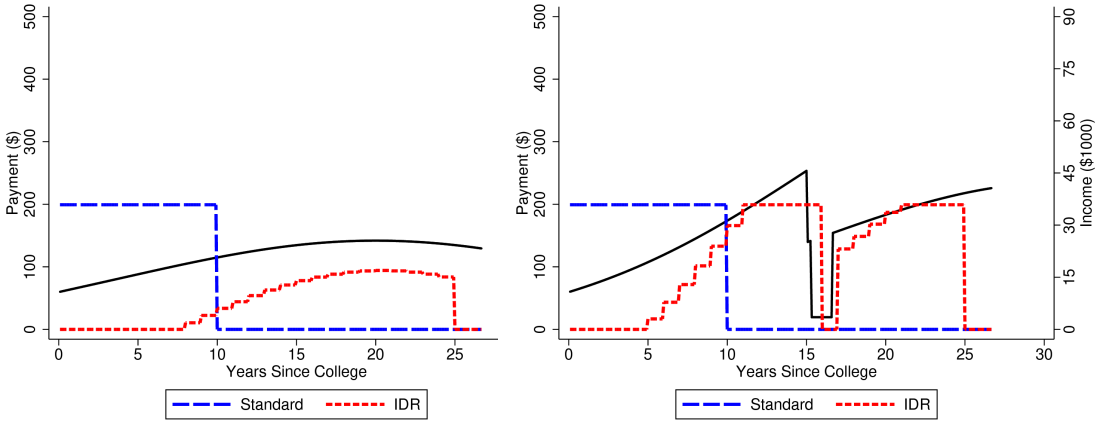
		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Card Balances	1.42	0.05	0.24***	0.40***	0.02	0.21**	0.37***
		(0.05)	(0.08)	(0.11)	(0.07)	(0.09)	(0.12)
Log Credit Card Balances	-2.51	0.07	0.33***	0.38***	0.06	0.32***	0.37***
		(0.05)	(0.07)	(0.07)	(0.06)	(0.08)	(0.09)
Number of Credit Cards	3.00	−0.01	0.07	0.14**	−0.03	0.06	0.12**
		(0.03)	(0.04)	(0.06)	(0.03)	(0.05)	(0.06)
Credit Card Limits	2.92	−0.13	0.14	0.34	−0.23	0.04	0.24
		(0.12)	(0.16)	(0.21)	(0.14)	(0.18)	(0.23)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on yearly credit cards. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same yearly effect for a regression which omits pre-call year dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 124,447 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure 1.1: Hypothetical Repayment Scenarios: IDR versus Standard Repayment



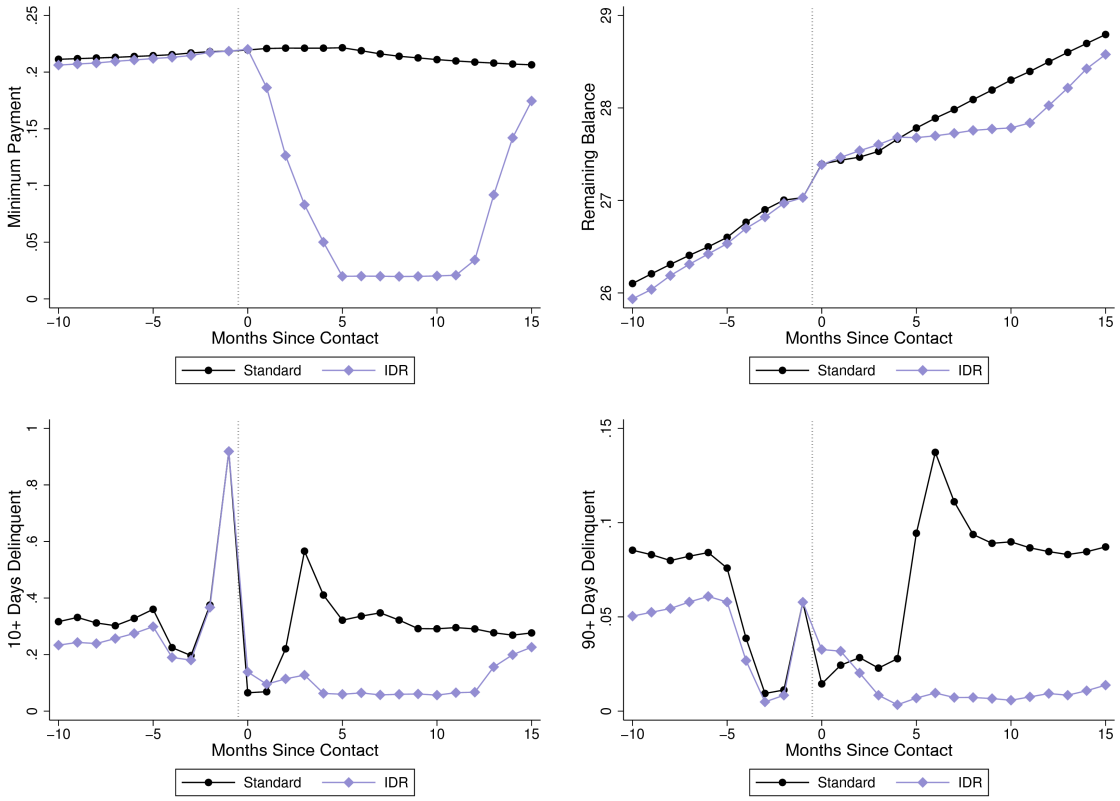
(A) (B)



(C) (D)

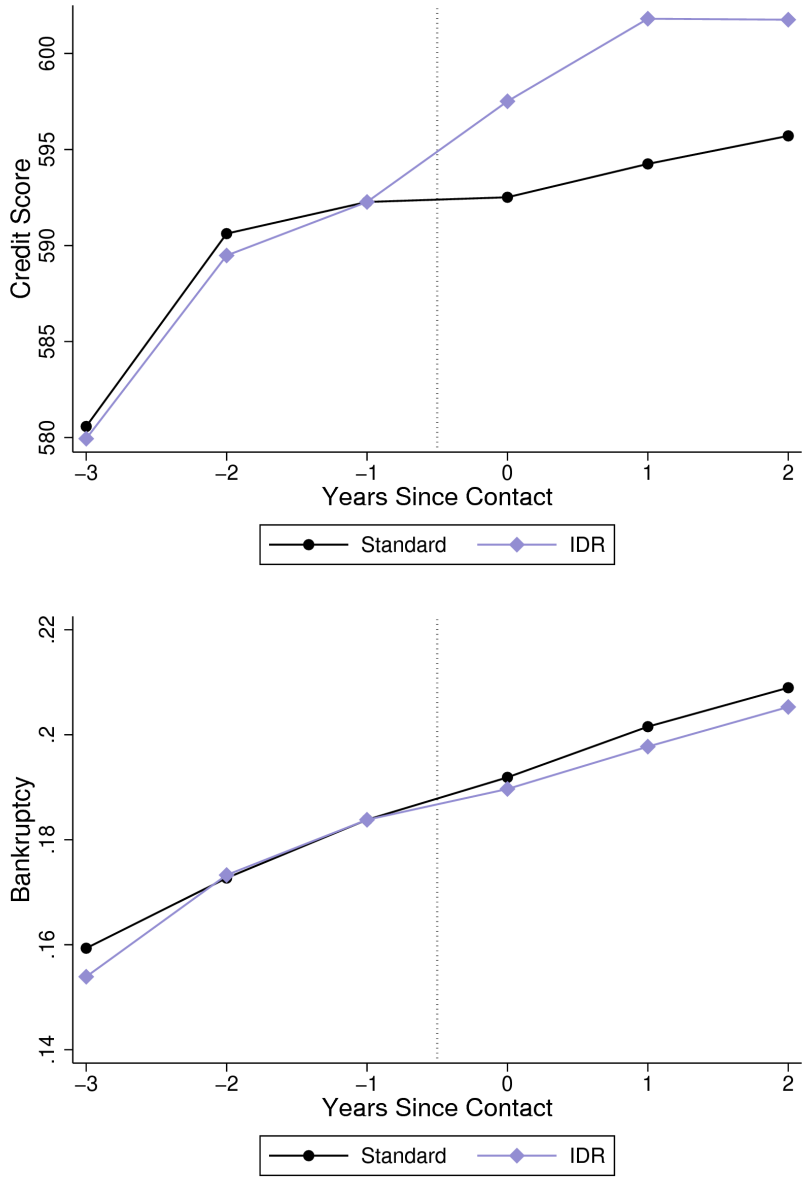
Note: This figure plots hypothetical repayment paths for standard and IDR plans under various income scenarios. Each panel represents an alternative income/repayment scenario for a borrower holding $18,000 of student debt at the time they leave college. The solid black line, plotted against the right axis, represents annual post-college income. The dashed blue and dotted red lines, plotted against the left axis, represent monthly payments under standard and IDR plans, respectively. The x-axis denotes years since leaving college. Repayment paths assume a 6.0 percent interest rate and no late payments.

Figure 1.2: Pre/Post-Call Trends in Repayment Outcomes



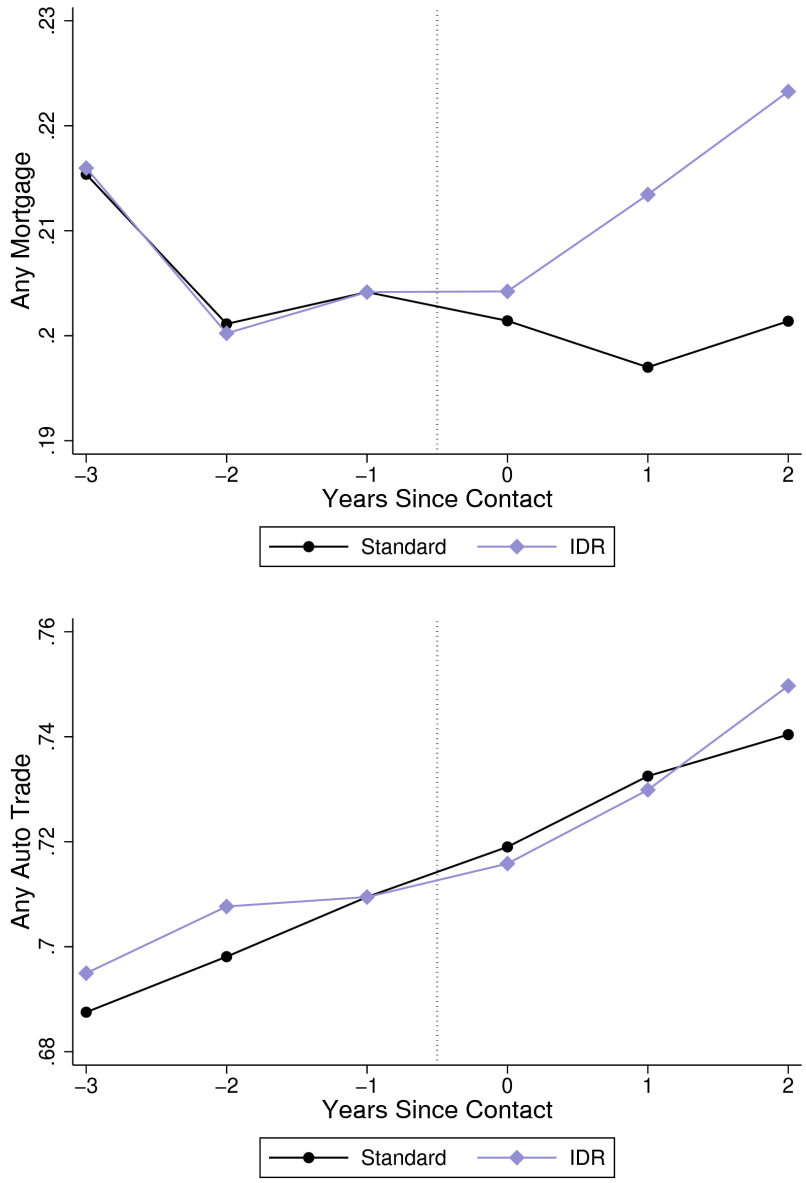
Note: This figure plots the average loan repayment outcomes for IDR enrollees and non-enrollees in the analysis sample. The horizontal axis denotes time, in months, relative to the month of the loan servicing call. Outcomes are normalized to the average value of the outcome for non-enrollees in the month prior to the call. See Table 2.1 notes for additional details on the outcome measures and sample.

Figure 1.3: Pre/Post-Call Trends in Credit Scores and Bankruptcies



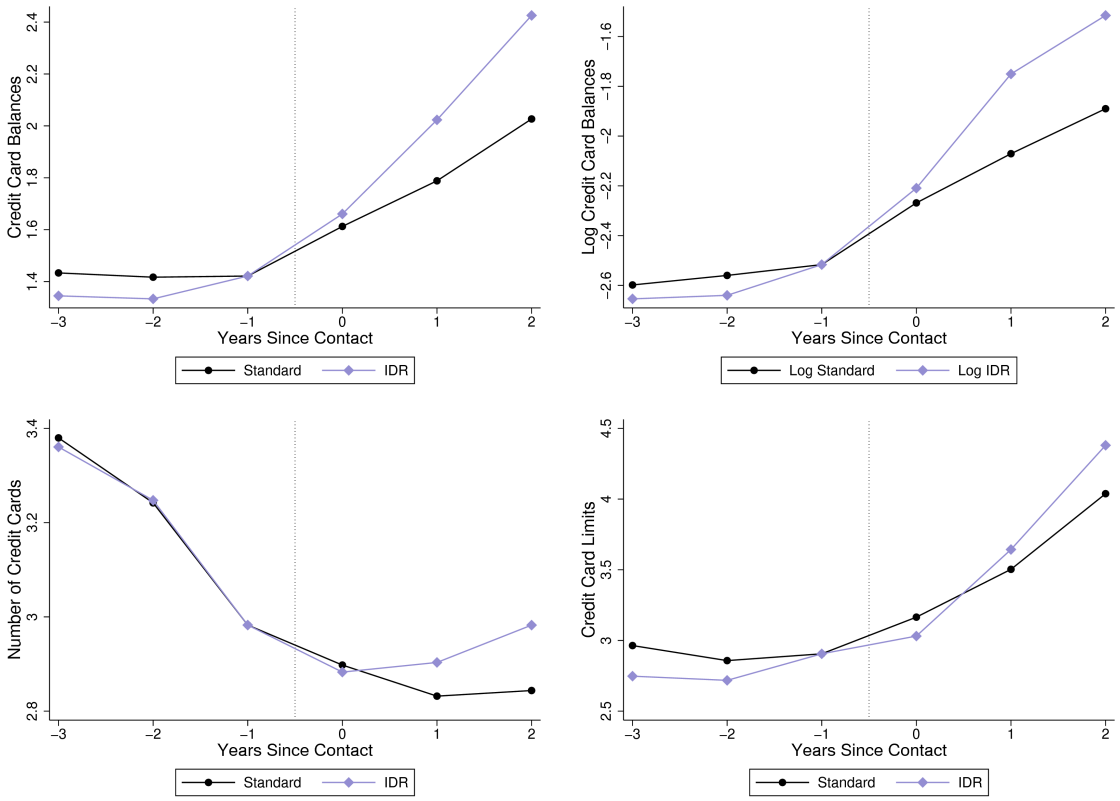
Note: This figure plots the average credit scores and bankruptcies for IDR enrollees and non-enrollees in the analysis sample. The horizontal axis denotes time, in years, relative to the year of the loan servicing call. Outcomes are normalized to the average value of the outcome for non-enrollees in the year prior to the call. See Table 2.1 notes for additional details on the outcome measures and sample.

Figure 1.4: Pre/Post-Call Trends in Mortgages and Auto Loans



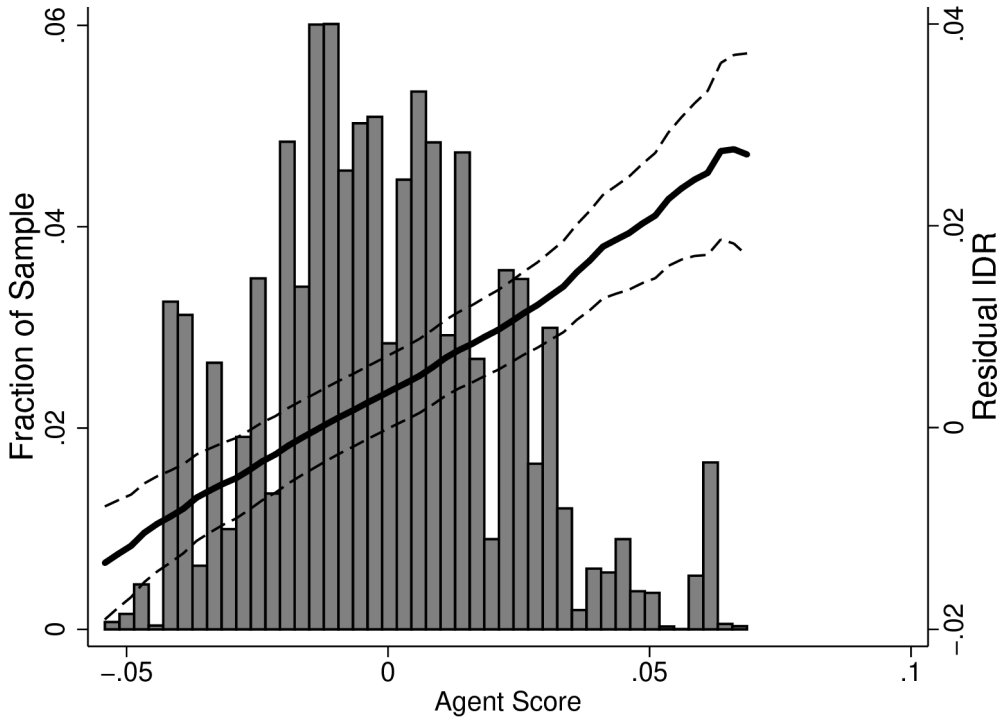
Note: This figure plots the share of IDR enrollees and non-enrollees holding mortgages or auto loans in the analysis sample. The horizontal axis denotes time, in years, relative to the year of the loan servicing call. Outcomes are normalized to the average value of the outcome for non-enrollees in the year prior to the call. See Table 2.1 notes for additional details on the outcome measures and sample.

Figure 1.5: Pre/Post-Call Trends in Credit Cards



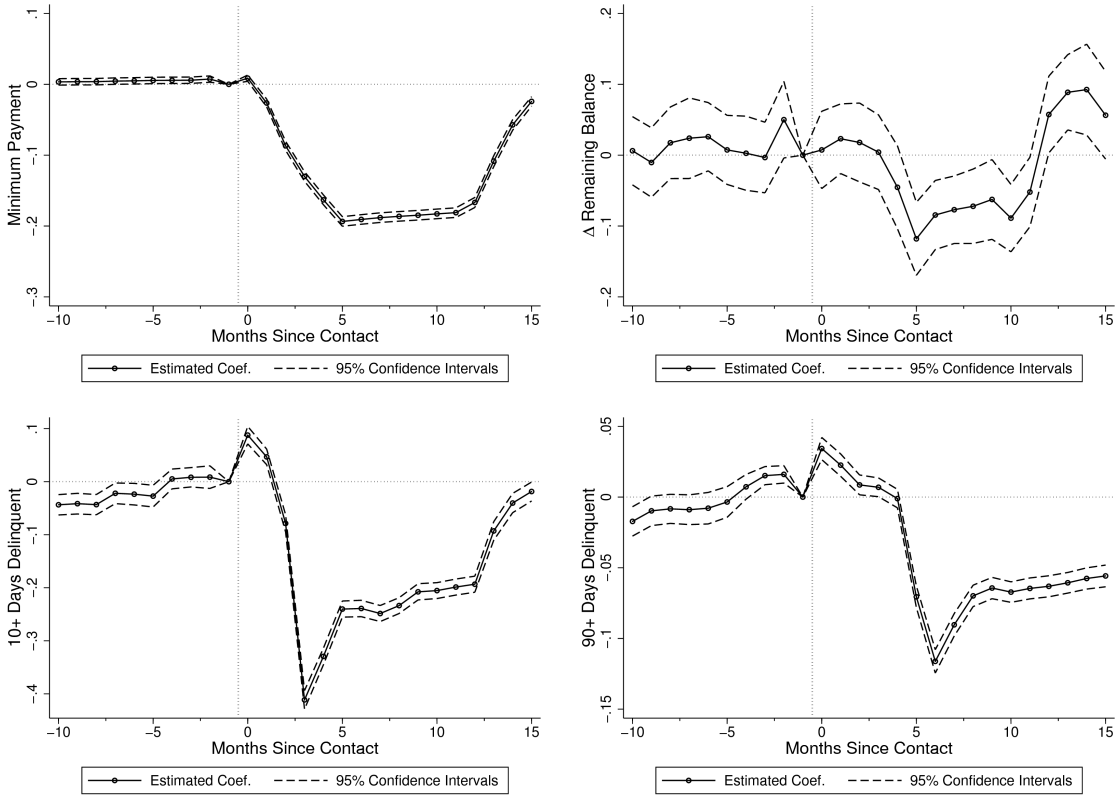
Note: This figure plots the average credit card outcomes for IDR enrollees and non-enrollees in the analysis sample. The horizontal axis denotes time, in years, relative to the year of the loan servicing call. Outcomes are normalized to the average value of the outcome for non-enrollees in the year prior to the call. See Table 2.1 notes for additional details on the outcome measures and sample.

Figure 1.6: First Stage



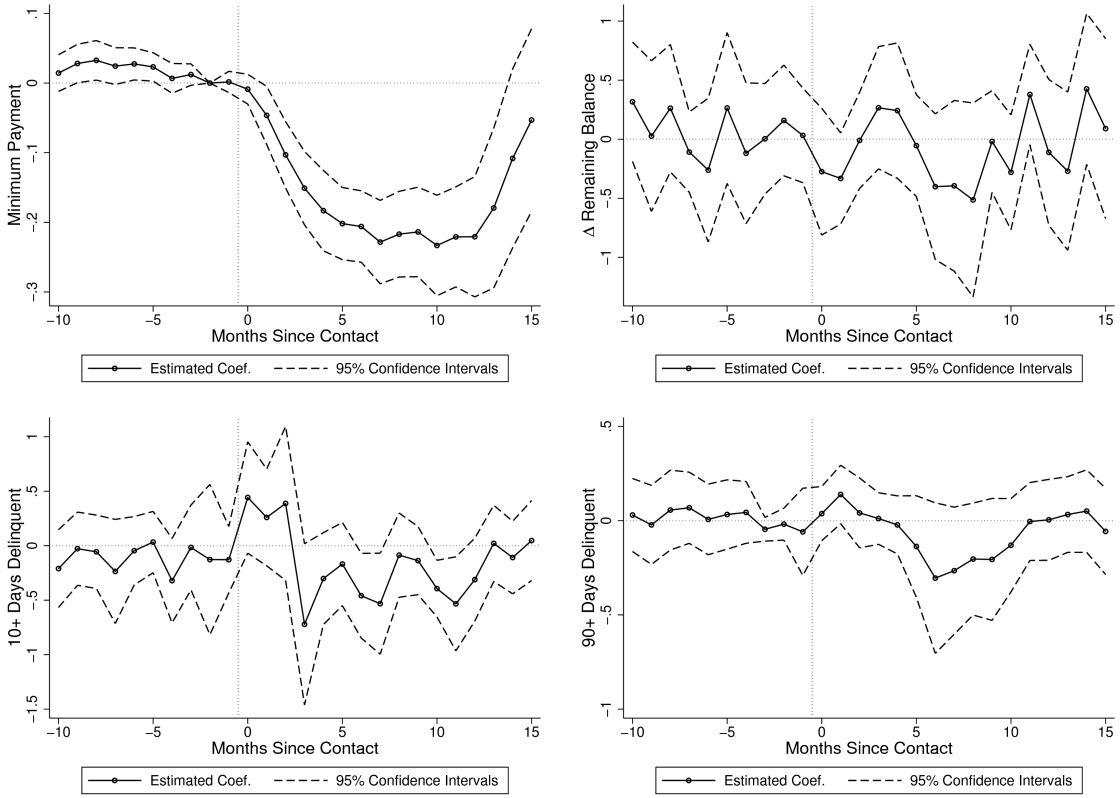
Note: This figure reports first-stage effects and distribution of agent scores across delinquency calls, where agent score is the leave-out mean IDR take-up calculated using data from other calls made by the agent following the procedure described in Section 1.5.2. The solid and dashed lines, plotted against the right axis, represent predicted means with 95% confidence intervals from a local linear regression of residualized IDR take-up on agent score. The histogram, plotted against the left axis, provides the distribution of agent scores across all delinquency calls in my analysis sample. All regressions include the full set of call date and time fixed effects.

Figure 1.7: Difference-in-Differences: Repayment Outcomes



Note: This figure reports estimated coefficients from a series of two-stage least squares regressions. Each point represents the estimated effect of post-call IDR status on the outcome variable at a given time period relative to the date of delinquency call. Relative months are plotted along the x-axis. Dashed lines represent 95% confidence intervals. All regressions include the full set of call date and time fixed effects. Robust standard errors are two-way clustered at the borrower and agent level.

Figure 1.8: Two-Stage Least Squares: Repayment Outcomes



Note: This figure reports estimated coefficients from a series of two-stage least squares regressions. Each point represents the estimated effect of post-call IDR status on the outcome variable at a given time period relative to the date of delinquency call. Relative months are plotted along the x-axis. Dashed lines represent 95% confidence intervals. All regressions include the full set of call date and time fixed effects. Robust standard errors are two-way clustered at the borrower and agent level.

Chapter 2

Adverse Selection, Income Insurance, and College Financing

2.1 Introduction

Investing in college carries a huge risk for millions of potential enrollees. While average returns to postsecondary education are generally positive, the realized incomes of observably equivalent college attendees vary dramatically, in part reflecting ex-ante earnings uncertainty. One way to abate this "college investment risk" would be to offer some form of income insurance to college enrollees. In his 1962 book *Capitalism and Freedom*, Milton Friedman discusses the benefits of such insurance:

> "[Human capital] investment necessarily involves much risk. The average expected return may be high, but there is wide variation about the average...The device adopted to meet the corresponding problem for other risky investments is equity investment plus limited liability on the part of the shareholders. The counterpart for education would be to "buy" a share in an individual's earning prospects; to advance him the funds needed to finance his training on condition that he agree to pay the lender a specified fraction of his future earnings." [69]

This line of reasoning provides a rationale for recent governmental efforts to implement income-driven repayment (IDR) for student loans. IDR programs peg minimum monthly student debt payments to post-graduate income, typically requiring borrowers to repay a fixed portion of their yearly earnings until their student debt is paid off or some forgiveness period has been reached. By dynamically adjusting one component of college costs–student debt repayment–IDR smooths wealth not only over time but across states of the world, acting as a sort of implicit income insurance. Indeed, chapter 1 suggests that IDR does, in fact, improve borrower welfare.

But if borrowers value income-contingent debt repayment, why does there not exist a private market providing such products? It is not hard to imagine a firm which repays borrowers' student loans at the point of graduation and collects future payments based on the their realized incomes. Indeed, so-called "income share agreements" (ISAs) do exist in

a few small, selected markets.[1] If borrowers would pay a premium to reduce their post-graduate earnings risk, why are ISA contracts so uncommon?

In this paper, I formally develop and test the following hypothesis: adverse selection prevents income share agreements from begin sold profitably. To see how this "no trade" condition might be met, suppose student borrowers hold private information concerning their future earnings potential that is unknown to any would-be insurers. This information asymmetry can prevent efficient contracting if the cost of insuring those with low earnings potential exceeds the willingness to pay for those with higher earnings potential, driving "high-type" individuals out of the market and further increasing the cost of coverage–a process known as market unraveling.

First, I provide empirical evidence that individuals hold private information concerning their future income. Using survey data which include responses to questions concerning the personal income expectations of over 3,000 college freshman, I test how well these subjective elicitations predict *realized* incomes among similar borrowers. Under reasonable assumptions concerning borrowers' elicitations and true beliefs, predictive power would imply borrowers hold private knowledge of their future incomes, suggesting ISA markets would be adversely selected.

Applying both traditional econometric strategies and machine learning techniques, I reject the null hypothesis that private beliefs are uninformative in predicting future income. A simple Wald test in an OLS framework rejects mutually zero coefficients with a p-value of 0.001, and a 10-fold cross-validated random forest model finds a statistically significant improvement in both R-squared (p-value = .037) and RMSE (p-value = .037) when subjective elicitation variables are added to a rich set of "public" income predictors. Under assumptions of "unbiasedness" and "self-awareness" of private information, these results can be seen

[1]For example, Purdue's "Back-A-Boiler" program offers discounted tuition in exchange for up to 15% of monthly income, depending on major of study [20].

as evidence of ex-ante private information, and hence, adverse selection in a hypothetical insurance market.

Next, having established the existence of private information, I next seek to determine whether it can explain market unraveling. To inform a test of market unraveling, I develop a model of student debt repayment in which borrowers hold privately-known beliefs concerning their future income distributions and an insurer offers contracts to maximize expected profits in light of this asymmetric information. Retaining the assumption of unbiased beliefs, the model characterizes the set of feasible allocations as a function of the distribution of private information, generating a "no trade" condition under which the only potential contract is each borrower's endowment.

Finally, I use a novel machine learning technique known as Regression Quantile Forests (QRF) to empirically test the "no-trade" condition informed by my model. The QRF approach allows me to non-parametrically estimate the conditional CDF of each borrower's future income, with and without the inclusion of private elicitations. At a crude level, the unraveling condition relies on the relative variances of these distributions: the higher *publicly* expected income uncertainty is relative to *privately* expected income uncertainty, the higher are adverse selection costs of providing insurance. Using these estimates, I predict the incidence of market unraveling under multiple specifications for borrowers' risk preferences in a CRRA utility framework.

My results carry several important caveats. First, it should be emphasized that this paper explores *one* potential reason for the absence of ISAs. In reality, there are a myriad of other factors which may prevent or attenuate exchange of ISA contracts, including moral hazard, legal restrictions, and administrative barriers. My results do not deny the relevance of such factors, but rather suggest that removing these barriers without also addressing the adverse selection phenomenon would fall short of creating a market for ISAs. In other words, adverse selection and other types of market failure are not mutually exclusive.

Second, while my framework allows me to test for market unraveling under any number of demand specifications, in practice I assume borrowers hold homogeneous preferences exhibiting constant relative risk aversion (CRRA). Admittedly, an ideal analysis would directly estimate borrowers' willingness to pay, but estimating demand for a range of non-existent products imposes obvious practical constraints. Similarly, I restrict attention to linear income share contracts, as a richer set of income-contingent contracts greatly complicates the model's solution.[2]

Third, my analysis rests upon the assumption that borrowers hold unbiased beliefs regarding their future income. While I remain agnostic about the relationship between elicitations and beliefs, the exercise nonetheless assumes borrowers would make the best possible guess about their future income distribution and optimize expected utility accordingly. While this is a strong assumption, it is one that is made in innumerable economic studies, especially those concerning asymmetric information (e.g. [55]).

Finally, while private elicitations arguably capture *some* private information, they certainly do not capture *all* of it. As such, any predictive power they hold can be interpreted as a lower bound on the true amount of private knowledge in a hypothetical ISA market. It follows that the no trade condition provides a sufficient, though not necessary, condition for market unraveling under the model's assumptions.

The remainder of this paper is organized as follows. Section 2.2 provides some brief background on adverse selection, my empirical setting, and the NCES dataset used in my analysis. Section 2.3 provides empirical evidence of private knowledge over future income and demonstrates the predictive power of private elicitations. Section 2.4 presents a model of adverse selection in a market for ISAs and generates a "no trade" condition under which

[2]By "linear" ISA, I mean a contract in which the firm offers a lump sum payment towards college before or during enrollment in exchange for a constant share of income after graduation. In theory, a hypothetical insurer might offer a more complicated income-contingent contract with non-linear mapping of realized income into required payment size, though the few ISAs that have been implemented tend to follow a linear structure [20].

the market unravels. Section 2.5 designs an empirical test for this condition and presents results from that test. 2.6 concludes.

2.2 The Empirics of Adverse Selection

The adverse selection phenomenon, whereby private information influences one's preferences over contract terms and distorts market efficiency, has featured prominently in a variety of economic fields since the seminal work by [7]. In recent decades, its influence has extended into applied work on auto insurance, health insurance, and other markets with asymmetric information. Traditionally, empirical strategies to detect adverse selection in these markets have involved observed correlations between demand and realized risk ([54], [33]). But using data on existing markets to detect adverse selection poses limitations, as the phenomenon can prevent otherwise efficient transactions from ever taking place. If an "unraveled" market never exists, there is little inference that can be drawn from observed prices and demand. To overcome this limitation, a more recent literature considers adverse selection in *hypothetical* insurance markets ([59], [78], [79]). These papers use subjective elicitations of individuals' own risks to assess whether, and by how much, private information would drive demand for an imaginary insurance product.

The nature of income-contingent contracts makes them susceptible to adverse selection because salaries depend upon individuals' own productivities, which cannot be directly observed. Moreover, because they offer a large potential benefit to student borrowers, yet seldom exist in the real world, ISAs are a prime candidate for an unraveled market. For this reason, I employ an approach similar to [78] and others, using subjective elicitations to test for adverse selection in a *hypothetical* ISA market. The approach carries very specific data requirements–namely, individual realized income data, income-predictive survey responses, and an exhaustive set of "screenable" characteristics.

2.2.1 Data

The information required for my empirical strategy can be classified into three categories:

1. *Subjective elicitations*: survey responses reflecting information on expected future income which would be ex-ante known to the individual but publicly unobserved.

2. *Realized income*: ex-post realizations of each individual's income after college.

3. *Public information*: any publicly observed individual characteristics a hypothetical insurer might use to screen or risk-rate individuals.

The data I use to meet these criteria come from the 2012 Beginning Postsecondary Study (BPS), a dataset from the National Center for Education Statistics (NCES). The data consist of administrative loan and financial aid data linked to survey responses for a representative sample of first-year college students in the spring of 2012, followed-up in 2014 and 2017. My analysis sample is selected to be enrolled students who have a nonzero loan balance, have left school as of 2014, and report full or part-time employment. [3] The analysis sample consist of 3,610 students.

Subjective Elicitations As a proxy for privately held beliefs regarding future income, I use survey questions of the following form:

> *We have some questions about the range of salary you expect to make once you finish your education. What is...*
>
> 1. *Your expected yearly salary?*
> 2. *Your highest expected yearly salary?*

[3] Data from the 2017 interview are not yet available, so I am limited to students for whom I can observe earnings within 2.5 years of their first year of college. While I hope to incorporate data from the 2017 wave in later versions of this paper, 2-year college students (and 4-year college dropouts) are a population of interest and currently comprise roughly half of the population of first-year college students.

3. *Your lowest expected yearly salary?*

Binned scatter plots for elicitations versus realized income are given by Figure 2.1. The predictive power held by these elicitations can be seen in the positive correlation evident in each panel. Loosely speaking, the empirical tests described in the following section amount to testing the significance of this correlation after flexibly controlling for all available public information on borrower risk.

Realized Income To measure realized income I use the annualized salary of current or most recent job the respondent held in 2014. I restrict to salary income for two reasons. First, annual

Second, variation in salary income is more informative of future earnings potential, so results are more generalizeable to longer time horizons than they would be using gross income, which is more prone to idiosyncratic shocks.

Public Information To control for public information, I use a rich set of covariates covered by the various data sources contained in BPS. Among them are demographic characteristics (age, gender, parental income, marital status, etc.), institutional details (college type, ranking, size), academic performance (GPA, SAT, ACT), loan balance, work-study status, employment history, and more. These variables come close to full set of publicly available upon which a hypothetical insurer could conceivably contract. Note that federal protected classes like race and religion are excluded from the analysis. Summary statistics for selected variables are available in Table 2.1.

2.3 Evidence of Private Information

I begin my empirical analysis by first providing evidence that borrowers do, in fact, hold private knowledge of their future income, as such knowledge is a necessary condition for

adverse selection to take place. I develop a testable condition for the existence private information, first providing the assumptions under which the test is valid, then applying both traditional econometric and machine learning techniques to show the condition is satisfied.

Let borrowers be characterized by parameter θ, which captures all private information held by the borrower. A borrower with information θ holds ex-ante beliefs concerning the likelihood of any given future income realization $Y = y$, represented by the CDF $\mathcal{F}_\theta(y)$ with mean μ_θ. Let Z denote survey responses to elicitations, and X be a vector of observable public information. The true distribution of future income Y is given by CDF $F(y)$. To facilitate an empirical test for the existence of private information, I make two key assumptions.

Assumption 1. *Z contains no additional information than θ about Y,*

$$F\left(y|X,\theta,Z\right) = F\left(y|X,\theta\right)$$

Assumption 1 states that the information contained in elicitations Z can contain *no more information* than true beliefs θ. Put differently, if I could observe one's entire belief distribution, knowing her elicitation responses would not improve my prediction of future income. In some sense, this can be thought of as a "self-awareness" assumption; if respondent's answers reveal information about their future income, I assume they are aware of this information and would act on it in a hypothetical insurance market. This seems reasonable given the questions explicitly ask for future income expectations.[4]

Assumption 2. *Beliefs θ are unbiased,*

$$F\left(y|X,\theta\right) = \mathcal{F}_\theta(y)$$

[4]If instead the question asked "What's the population of Germany?", answers may well predict future income, but such correlation would most likely reflect unobserved ability of which the respondent may not be aware, in which case Assumption 1 would be violated.

Assumption 2 states that true beliefs are unbiased in the sense that, were they observable, respondents' expected likelihood of earning any given income would be correct. This assumption requires there be no systematic over-confidence or pessimism in one's beliefs of future income. While this is admittedly a restrictive assumption, it is a weaker form of rational expectations and is made in many empirical studies of asymmetric information. Note also that it does not require *elicitations* to be unbiased; I remain agnostic about the relationship between Z and θ.

2.3.1 Empirical Approach

To establish the existence of private information, I must reject the following null hypothesis:

$$H_0 : \mathcal{F}_\theta(y) = F\left(y|X\right) \quad \forall y \tag{2.1}$$

Theorem 1 provides a sufficient condition for rejecting H_0 in 2.1.[5]

Theorem 1. *Suppose Assumptions 1 and 2 hold. Then* $\mathcal{F}_\theta(\cdot) \neq F\left(\cdot|X\right)$ *if*

$$E\left[Y|X,Z\right] \neq E\left[Y|X\right] \tag{2.2}$$

Testing the condition given by 2.2 requires estimates of the conditional expectation function (CEF) of Y, under public information (conditioning on X) and private information (conditioning on X and Z). To estimate these CEFs, I use two approaches.

[5]Proof in Appendix B.1.

OLS Test

I first test (2.2) using ordinary least-squares estimation, which provides the best linear unbiased estimate of CEF $E[Y|X,Z]$. I run an OLS regression using the following specification:

$$Y_i = \beta X_i + \Gamma(\boldsymbol{Z_i}) + \epsilon_i \tag{2.3}$$

where $\Gamma(\boldsymbol{Z_i})$ represents vector of second-order polynomials for each elicitation Z_i. I run OLS on (2.3) and conduct a Wald test on the joint hypothesis that $\Gamma(\boldsymbol{Z_i}) = 0$. The drawback of this approach is the inflexibility inherent to a linear model. Indeed, the Wald test for significance of $\Gamma(Z)$ compares the predictive power of 2.3 to the analogous model given by

$$Y_i = \beta X_i + \epsilon_i \tag{2.4}$$

But neither (2.4) nor (2.3) represent the *best possible predictor* of Y given $\{X\}$ or $\{X,Z\}$, merely the best *linear* predictor. Without a more flexible estimator for $E[Y|X]$, the Wald test may reject the null if Z_i is correlated with non-linear predictors in X_i, regardless of whether private information exists. In theory, one could allow for greater flexibility parametrically by including polynomials and the like, but limited sample size presents a dimensionality problem which makes this infeasible.

Machine Learning Test

The burgeoning literature applying machine learning techniques to prediction problems in economics offers a solution to dimensionality problem posed above.[6] The basic principle behind this approach is to flexibly fit the best predictive model to the data while repeatedly cross-validating to prevent over-fitting. Unlike OLS methods, machine learning algorithms place virtually no restrictions on the functional form of $E[Y|X]$.

[6]For a review of machine learning's applicability to prediction problems in economics, see [108].

For two reasons, I argue that this non-parametric, machine learning approach is well-suited for my context. First, I am unconcerned with point estimates relating elicitations to future income. In my framework, the relevant question is not "What is the effect of Z on Y?", but rather "How well does Z predict Y?" Machine learning methods, while poorly equipped to answer the former, can often improve upon traditional econometric methods in answering the latter. Second, machine learning methods more closely resemble the decision-making processes I am modeling. A rational, utility-optimizing agent would make decisions using the her best possible prediction of future income, not one restricted to a linear functional form. Likewise, hypothetical insurance companies would make the best prediction of future income possible when assessing a borrower's risk. In fact, existing insurance companies, credit rating agencies, and financial firms already make use of machine learning techniques in their underwriting operations.

To provide machine learning-based estimate of $E[Y|X, Z]$, I conduct the following procedure:

1. Randomly split analysis sample into "training" and "test" sets.
2. Estimate $\hat{E}[Y|X]$ on training set using the random forest algorithm.[7]
3. Predict $\hat{y}_i = \hat{E}[Y_i|X = \boldsymbol{x}_i]$ using $\boldsymbol{x}_i$ from the test set.
4. Repeat steps 2 and 3 including adding private elicitations $\boldsymbol{z}_i$.

A graphical representation of the procedure above is given by Figure 2.4. Comparing estimated CEFs $\hat{E}[Y|X]$ and $\hat{E}[Y|X, Z]$ provides one test of 2.2. I also assess the relative predictive powers of the two models by computing R-squared and RMSE statistics for the out-of-sample test set. For robustness, I rerun the procedure for ten different training-test pairs, iteratively computing R-squared and RMSE statistics on a 10% holdout sample ten

[7]The random forest algorithm was chosen for its broad applicability and improved performance relative to comparable algorithms. My specification allowed for 2001 trees with 22 variables (of 68 total) randomly sampled at each split. See [22] for details on random forest.

separate times, without overlapping holdout samples (see Figure 2.5). The result are sets of 10 different R-squared/RMSE computations for each model, where the two models' statistics were computed on the same respective training/test samples.

2.3.2 Results

Both OLS and machine learning approaches find evidence of private information. OLS results are shown in Table 2.2. Across all specification, a Wald test rejects the null hypothesis with at the five-percent level. While the "Lowest expected salary" elicitation appears to hold the greatest predictive power, each of the three elicitations are individually significant in separate regressions.

Random forest model statistics are provided in Table 2.3. Including private elicitations Z in the model increases out-of-sample R-squared from .1615 to .1776, and decreases the out-of-sample RMSE from 9416 to 9333. Table 2.4 shows results from paired t-tests and paired Wilcoxon tests comparing the two models. The first row tests for equality in values $\hat{y}_i$, rejecting the null with p-values 0.002 and 0.001 for t-test and Wilcoxon test, respectively. The second row tests for improvement in prediction errors $|y_i - \hat{y}_i|$, and rejects the null that private information does not improve prediction with p-values of 0.001 and 0.003, respectively.

Scatter plots of each model's iteratively-computed out-of-sample R-squared and RMSE statistics for each training-test pair are shown in Figure 2.6. In each panel, the x-axis measures the statistic's value using the public information model, while the y-axis measures the statistic's value using the public + private information model. The model including private information has a higher R-squared in all ten training-test pairs and a lower RMSE for nine of ten training-test pairs.

2.4 Model of Adverse Selection

The model, similar to [115] and [78], aims to characterize the *set of feasible allocations* as a function of the distribution of private information.

2.4.1 Setup

There exists a unit mass of borrowers.[8] Each borrower's post-graduate income y is drawn from distribution $F(y)$. As in Section 2.3, borrowers hold private knowledge θ from which they generate unbiased beliefs over their future income distributions $\mathcal{F}_\theta(\cdot) = F(\cdot|\theta)$. The parameter θ is itself distributed across the population of borrowers according to joint c.d.f. $\boldsymbol{G}(\theta)$, which is common knowledge.

Before graduating, borrowers can contract with an "insurer" to receive lump-sump college payment b in exchange for pledging λ-share of their future income once it is realized.[9] Borrowers of each type θ choose from menu of contracts $(\lambda, b) \in \mathcal{C}$ to maximize expected post-graduate utility:

$$\max_{(\lambda,b)\in\mathcal{C}} E\left[u\left(y\left[1-\lambda\right]+b\right)|\,\theta\right] \tag{2.5}$$

where $u(\cdot)$ holds standard properties for von Neumann-Morgenstern preferences.

The insurer offers a menu of contracts $\mathcal{C} \equiv \{(\lambda_\theta, b_\theta)\}$, where each contract is indexed by to the type θ of its intended borrower. The insurer chooses menu $\mathcal{C}$ to solve the following problem:

$$\max_{\{(\lambda_\theta,b_\theta)\}} \int_{\theta\in\Theta} E\left[\lambda_\theta y - b_\theta|\,\theta\right] d\boldsymbol{G}(\theta) \tag{2.6}$$

$$\text{s.t.} \quad \begin{array}{lcll} E\left[u\left(y\left[1-\lambda_\theta\right]+b_\theta\right)|\,\theta\right] & \geq & E\left[u\left(y\left[1-\lambda_{\hat{\theta}}\right]+b_{\hat{\theta}}\right)|\,\theta\right]\ \forall\theta,\hat{\theta} & \text{(IC)} \\ E\left[u\left(y\left[1-\lambda_\theta\right]+b_\theta\right)|\theta\right] & \geq & E\left[u(y)|\theta\right]\ \forall\theta & \text{(IR)} \end{array}$$

[8] For expositional purposes, the model excludes publicly observable risk characteristics X_i, though the logic extends if you condition on observables. In particular, let $G(\theta) \equiv G(\theta|X = x)$.

[9] In Appendix B.1.2, I modify the model to allow for a fully flexible mapping of income to payments, $x(y)$.

In words, the risk-neutral insurer chooses the set of available contracts to maximize expected profits subject to the constraints imposed by borrower behavior. Specifically, the incentive-compatibility constraint (IC) ensures that $(\lambda_\theta, b_\theta)$ is the optimal choice for borrower type θ given the complete menu of insurance options, and the individual rationality constraint (IR) requires that no borrower's contract be no worse than her outside option of remaining uninsured.

2.4.2 No-Trade Condition

Any set of contracts $C = \{(\lambda, b)\}$ which can be feasibly bought and sold must solve the above maximization problem. The "no trade" scenario in which no insurance is provided corresponds to the case where the IR constraint binds for all types Θ. The conditions for such a scenario are summarized in the following theorem:

Theorem 2. *The endowment* $\{(0,0)\}$ *is the only feasible contract if:*

$$\frac{1}{E\left[\mu|\mu < \mu_\theta\right]} > \frac{E\left[u'(y)|\,\theta\right]}{E\left[yu'(y)|\,\theta\right]} \forall\ \theta \tag{2.7}$$

Theorem 2 states that, for market to exist, type θ's willingness-to-pay for insurance against low-income states must exceed average cost of insurance provision *to all riskier types.*

2.5 Evidence of Market Unraveling

An empirical test the unraveling condition provided in Theorem 2 requires the following three objects:

1. $F(y|\theta, X)$, the private distribution of future income *within* each type.

2. $H(\mu_\theta|X)$, the distribution of private future income means *across* types.

3. A specification for $u(\cdot)$

Ideally, I would have some measure of each borrower's subjective probabilities of earning salary y for every possible y. While this is obviously impossible, I am able to leverage the unique nature of the data to find suitable alternatives. Specifically, I use elicitations Z, which contain private information held by the borrower, as a proxy for θ. I then use a generalization of the random forest machine learning algorithm known as Quantile Regression Forests (QRFs) to generate non-parametric estimates of the conditional distribution functions $\hat{F}(y|X,Z)$ and $\hat{H}(\mu_Z|X)$.[10] The technique provides the best out-of-sample predictor of income distribution $F(y|X,Z)$, allowing me to test a sufficient condition for Theorem 2, in effect answering the question "Would the market unravel if borrowers *only* knew X and Z?" Finally, I assume constant relative risk aversion utility, $u(y) = \frac{y^{1-\gamma}}{1-\gamma}$, and test for unraveling under alternative values of γ.

My procedure is outlined as follows:

1. Use QRF to predict private future income distributions out-of-sample:

$$\begin{aligned} \hat{F}_i(y) &= \hat{F}(y|X_i, Z_i) \\ \hat{\mu}_i &= \hat{E}(y|X_i, Z_i) \end{aligned}$$

2. Flexibly estimate distribution of private future income means:

$$\hat{H}_i(\mu) = \hat{H}(\mu|X_i)$$

3. Calculate no trade condition in Equation 2.7 for each i under alternative γ using CRRA utility.

[10] For details on the quantile regression forest algorithm, see [105].

Results

Figure 2.8 presents quantile regression forest estimates of individual private probability density functions, $f(y|X, Z)$, for the future incomes of three randomly chosen individuals in my sample. Figure 2.9 provides rejection rates under different values of risk-aversion parameter γ. Values less than one imply a completely unraveled market where no contracts are profitable, while values greater than 1.4 imply all individuals will buy insurance contracts. Values in the interior suggest a screening equilibrium–contracting upon observables X would allow insurers to profitable sell to *some* segments of the population.

2.6 Conclusion

Taken together, my results suggest that a hypothetical ISA market would be adversely selected, as borrowers hold private information they would likely use in contracting decisions. Tests of the explanatory power of subjective information provides empirical evidence that borrowers do, in fact, hold private knowledge concerning their future income. I combine this empirical insight with a theoretical model to find that the magnitude of private information is enough to completely unravel many ISA markets under reasonable calibrations for risk-aversion. These results provide a rationale for public provision of IDR or other means of income insurance for student borrowers. Sufficiently large subsidies can prevent the unraveling mechanism from taking place, allowing millions of individuals to benefit from lower earnings risk and potentially encourage greater college enrollment.

2.7 Tables and Figures

Table 2.1: Summary Statistics

	Full Sample		Employed Only	
	Mean	SD	Mean	SD
Female	0.58	0.49	0.54	0.50
Black	0.22	0.41	0.19	0.39
Age	26.5	7.91	26.0	7.42
Married	0.18	0.38	0.17	0.38
Earned Degree	0.44	0.50	0.47	0.50
Year of Degree	2013.0	0.66	2013.0	0.66
Four Year College	0.49	0.50	0.49	0.50
EFC	3323.5	7815.1	3886.8	8247.8
Total Student Loans	7408.2	4220.8	7498.1	4267.2
Expected Salary	48459.3	25012.4	49037.4	25056.3
Lowest Exp. Salary	36347.1	20578.0	36577.2	20374.8
Highest Exp. Salary	65680.6	31184.1	66553.0	31181.5
Exp. Job Likelihood	8.49	1.83	8.51	1.77
2014 Salary	21686.5	10877.3	21686.5	10877.3
2014 Employment	0.67	0.47	1	0
N	5590		3610	

Note: This table reports summary statistics at the borrower level. The full sample includes all 2012/14 BPS respondents who are no longer attending school as of the 2014 interview. The employed only sample includes only those individuals in the full sample who report holding a job as of the 2014 interview. Married is a dummy variable which equals one if the respondent is married as of the 2014 interview. Earned degree is a dummy variable which equals one if the respondent reports having earned their degree as of the 2014 interview. EFC reports the "Expected Family Contribution" from each respondent's completed FAFSA. Total student loans reports the sum of all principal borrowing amounts from both federal and private loans for higher education. Expected salary, lowest exp. salary, highest exp. salary, and exp. job likelihood report survey responses to the private elicitations described in the main text. 2014 salary and 2014 employment report salary and employment status as of the 2014 interview.

Table 2.2: OLS Regressions of 2014 Salary on Public Information and Subjective Elicitations

	(1) 2014 Salary	(2) 2014 Salary	(3) 2014 Salary	(4) 2014 Salary	(5) 2014 Salary	(6) 2014 Salary
Expected Salary		0.0246***			0.0160	0.0182
		[0.00798]			[0.0136]	[0.0387]
Lowest Exp. Salary			0.0333***		0.0258*	0.0695*
			[0.00958]		[0.0139]	[0.0387]
Highest Exp. Salary				0.00903	-0.00763	0.0348
				[0.00597]	[0.00822]	[0.0283]
Controls?	Yes	Yes	Yes	Yes	Yes	Yes
$P^2(\boldsymbol{Z})$	No	No	No	No	No	Yes
F-stat $\Gamma(Z) = 0$		4.228	4.880	2.428	3.507	3.766
P-value $\Gamma(Z) = 0$		0.00204	0.000634	0.0458	0.00184	0.000967
R-squared	0.153	0.157	0.157	0.155	0.158	0.160
N	3610	3610	3610	3610	3610	3610

Note: This table reports results from an ordinary least-squares regression of realized 2014 salary on subjective elicitations of expected future salary Z and public information variables X. The first six rows report estimated coefficients and standard errors on elicitation variables across various specifications, while the bottom six rows report specification details and regression statistics, including p-values from a Wald test of joint significance for all subjective elicitation variables and their interactions, represented by $\Gamma(Z)$.

Table 2.3: Random Forest Model Statistics

	PUBLIC	PUBLIC + PRIVATE
Mean $\hat{y}$	20850	20880
Mean $\lvert\hat{y} - y\rvert$	7167	7098
R^2	.1615	.1776
RMSE	9416	9333

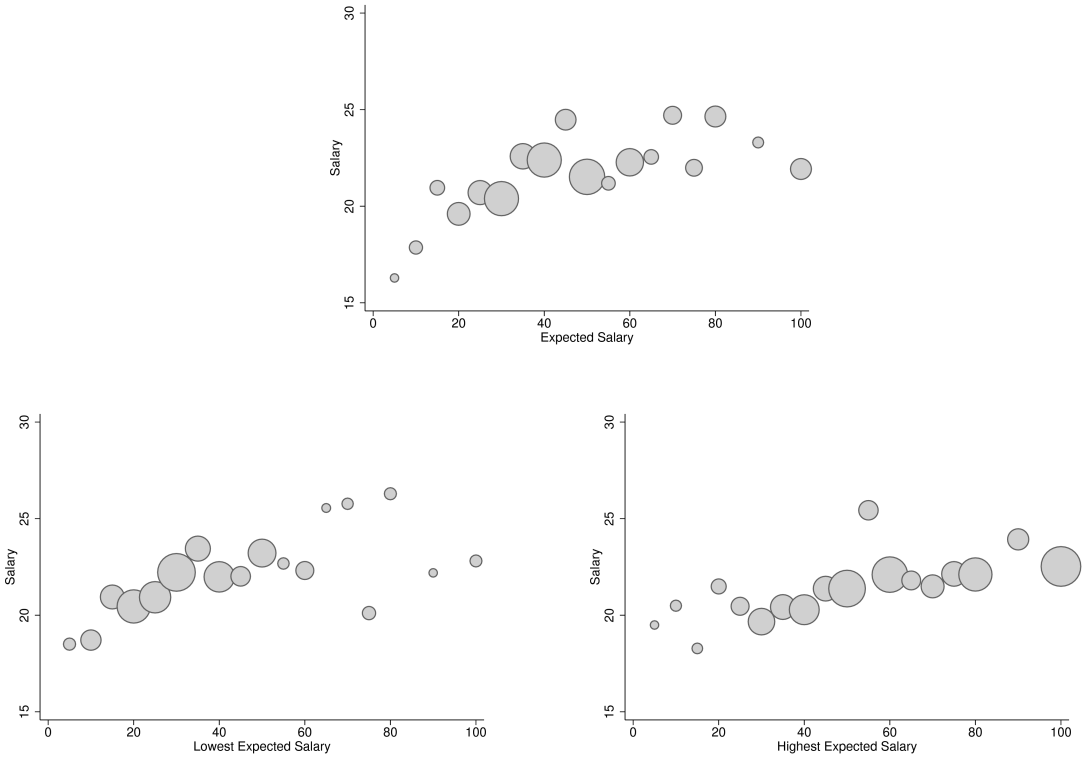
Note: This table reports summary statistics for out-of-sample predictions of two random forest models–one trained on a random fifty-percent sample using only public information variables, and one trained on the same sample using both public and private information variables. For classification of public and private information variables, see Section 2.2.1.

Table 2.4: P-values from Pairwise Comparison Tests

	t-test	Wilcoxon test
$H_0 : \hat{y}_{XZ} = \hat{y}_X$	0.002	0.001
$H_0 : \lvert e_{XZ} \rvert \geq \lvert e_X \rvert$	0.001	0.003

Note: This table reports results from paired t-tests and paired Wilcoxon tests comparing the public information and public + private information models. The first row tests for equality in values $\hat{y}_i$. The second row tests for improvement in prediction errors $|y_i - \hat{y}_i|$.

Figure 2.1: Correlation Between Subjective Elicitations and Realized Income



Note: This figure presents binned scatter plots of unconditional correlations between subjective elicitations and realized income. Circles represent equally spaced bins along the x-axis, which measures response values for each of three elicitations–"Expected", "Lowest Expected", and "Highest Expected" salaries, respectively. The size of each circle corresponds to the relative number of respondents falling into each bin. The y-axis measures realized 2014 salaries. All units are in thousands of dollars.

Figure 2.3: CDF of Prediction Differences

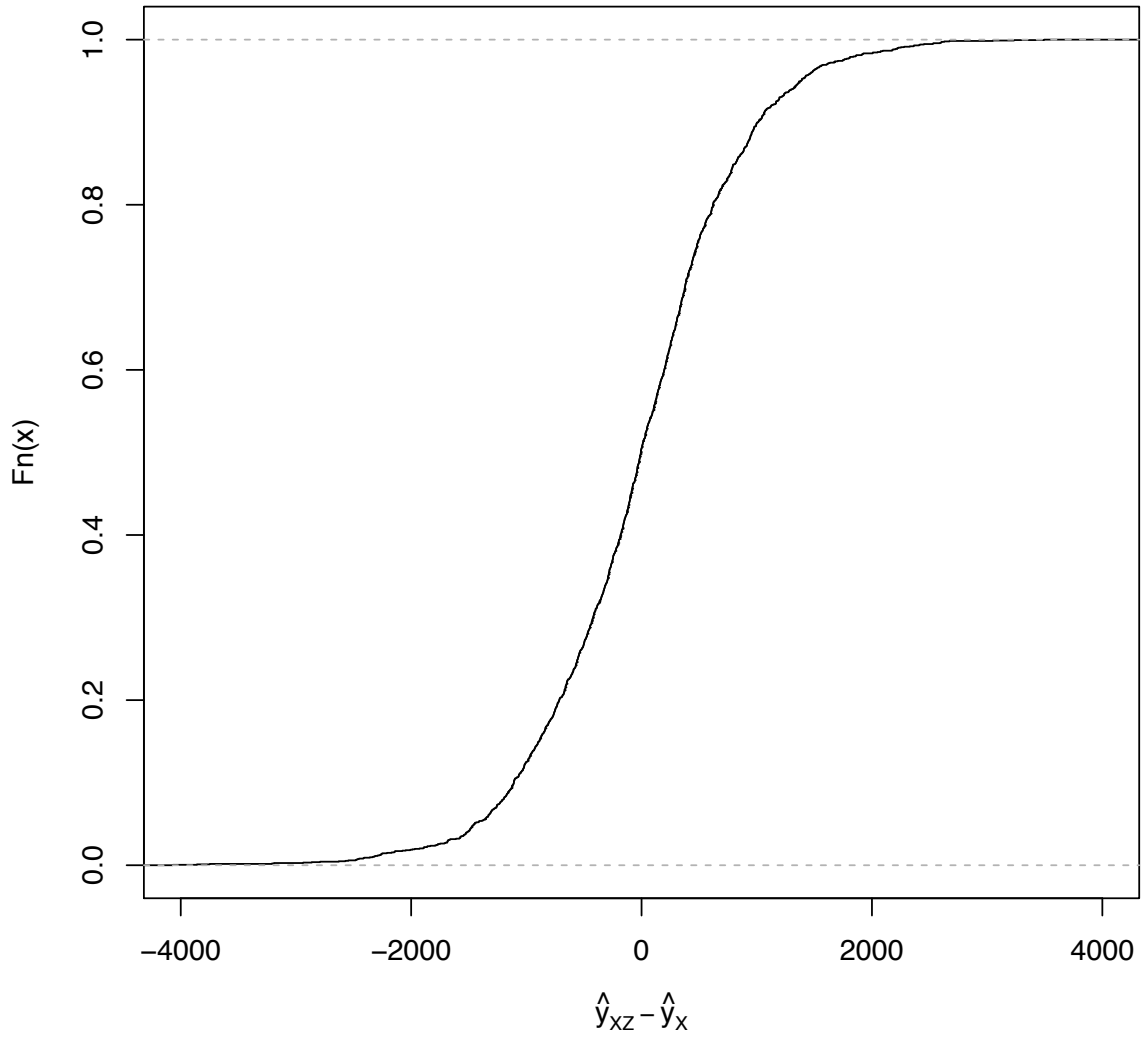


Figure 2.4: Out-of-sample Training/Test Procedure

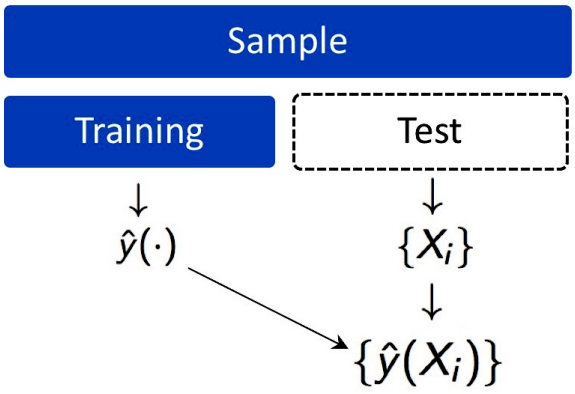


Figure 2.5: 10-Fold Out-of-sample Training/Test Procedure

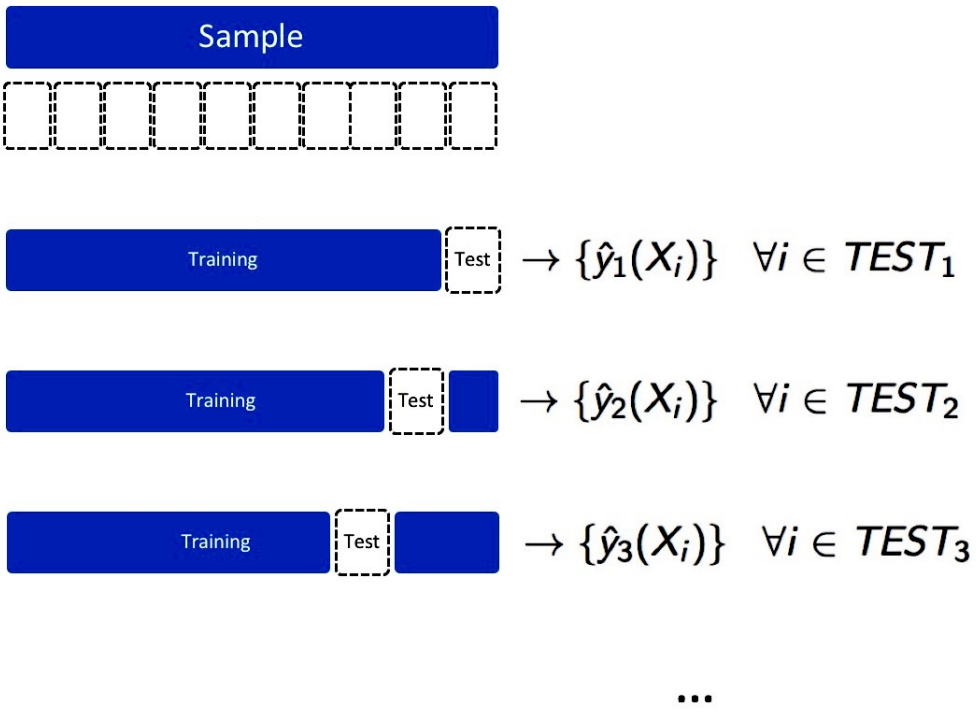
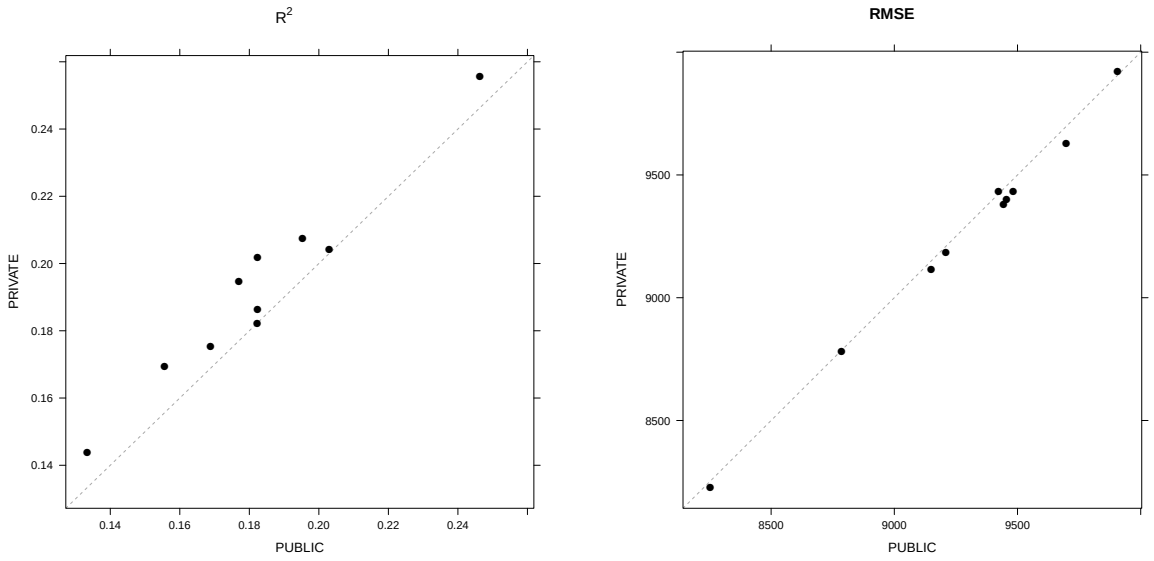
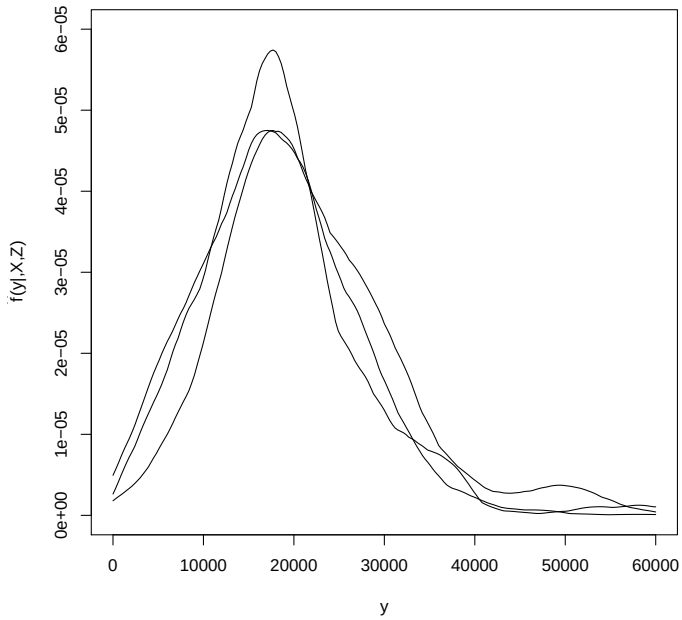


Figure 2.6: 10-Fold Out-of-sample Prediction Statistics



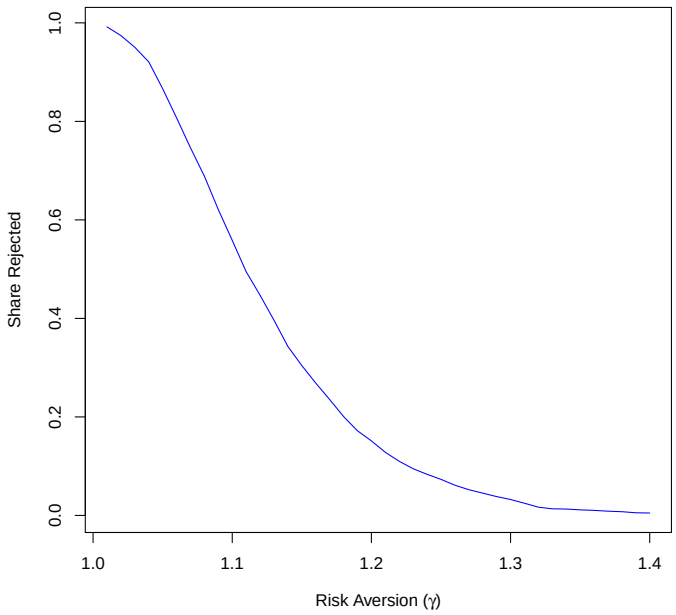
Note: This figure presents scatter plots of each model's iteratively-computed out-of-sample R-squared and RMSE statistics for each training-test pair. In each panel, the x-axis measures the statistic's value using the public information model, while the y-axis measures the statistic's value using the public + private information model.

Figure 2.8: Private Future Income Distributions



Note: This figure presents quantile regression forest estimates of individual private probability density functions, $f(y|X, Z)$, for the future incomes of three randomly chosen individuals in my sample.

Figure 2.9: Predicted Insurance Rejection Rates under Alternative γ



Note: This figure presents rejection rates under different values of risk-aversion parameter γ.

Chapter 3

Unions and Inequality Over the Twentieth Century: New Evidence from Survey Data[1]

[1]This chapter is coauthored with Henry Farber, Ilyana Kuziemko, and Suresh Naidu. We thank our research assistants Obaid Haque, Elena Marchetti-Bowick, Chitra Marti, Brendan Moore, Tamsin Kantor, Amy Wickett, and Jon Zytnick and especially Fabiola Alba, Divyansh Devnani, Elisa Jacome, Amitis Oskoui, and Maryam Rostoum. We have benefited from comments by seminar participants at Columbia, Georgetown, Harvard, INSEAD, the NBER Development of the American Economy meetings, McGill University, Princeton, Rutgers, Sciences Po, Universitat Pompeu Fabra, Society of Labor Economists Annual Meeting, Stanford, and Vanderbilt. We are indebted to Devin Caughey and Eric Schickler for answering questions on the early Gallup data.

3.1 Introduction

Understanding the determinants of the U-shaped pattern of U.S. income inequality over the 20th century has become a central goal among economists over the past few decades. While there is a substantial literature in labor economics and sociology that argues for a causal relationship from redistributive institutions such as labor unions to lowered inequality (27; 40; 124), the consensus among economists is that more fundamental drivers, namely technological developments that increase the demand for skilled labor, better explain the time-series variation in inequality [3].[2]

In the aggregate, there is a well-documented inverse relationship between income inequality and union membership in the US (see Figure 3.1). But moving beyond this aggregate relationship has proven difficult. While aggregate measures of union density date back to the early twentieth century, it is not until the Current Population Survey (CPS) introduces a question about union membership in 1973 that labor economists have had a consistent source of microdata that includes union status. Put differently, it is not until unions are in steady decline that they can be studied with U.S. micro data. It is perhaps not surprising that, instead, economists focus on models of inequality that emphasize the supply and demand for skill, as the U.S. Census has tracked education and wages consistently since 1940 (see, e.g., the seminal work by 74). Indeed, to the extent recent papers have tried to explicitly model the relationship between unions and inequality, they have argued that declining union

[2] As [3] write of the post-1970 period: "Most economists, however, discount the role of unions in the increase in inequality." Indeed, literature searches and polls of actual economists support the idea that economists favor alternative explanations of the variation in inequality over time. For example, a search for articles from 1990 onward on "inequality and technology" in Google Scholar yields 1.75 million hits, compared to 183,000 for "inequality and unions." In 1997, the Federal Reserve conducted a poll of economists, asking them their view on the key driver on increased inequality. Forty-four percent said "technological change," fifteen percent said "other," twelve percent said "international trade," nine percent said "decline in real minimum wage," nine percent said "decline in unionization," and the rest said "rising immigration." See 1997 Economic Report of the President, Figure 5-3.

density is largely an artifact of rising skill-biased technological change (SBTC), and thus unions themselves have little if any direct effect on inequality.[3]

In this paper we bring a new source of household-level data to the study of unions and inequality. While the Census Bureau did not ask about union membership until the 1973 CPS, public opinion polls regularly asked household union membership, together with extensive questions on demographics, socio-economic status and political views. We harmonize these surveys, primarily Gallup public opinion polls, going back to 1936. Our new dataset draws from over 500 surveys over the period from 1936-1986 and has over 980,000 observations, each providing union status at the household level. We combine these data with more familiar micro-data sources (e.g., the CPS) to extend the analysis into the present day.

These new data sources allow us to the revisit the role of unions in shaping the income distribution and contribute to the long-running "institutions versus market forces" debate on the causes of inequality. Bringing new micro data to the study of unions allows us to perform a number of novel tests on the plausibility of unions playing a significant role in reducing income inequality. For example, suppose we find that during the "Great Compression" period (roughly 1940–1970, when unions were at their peak and inequality at its lowest), union members were relatively educated or disproportionately white.[4] Then this pattern of selection, with generally low skill and low wage workers being underrepresented in the union sector during the twentieth century's most egalitarian years, would argue against unions playing an important role in reducing inequality. By contrast, the results we do find push against the idea that unions are simply an artifact of larger market forces and instead favor the idea that they may have their own causal effect in reducing inequality. We support this hypothesis with a number of new empirical results, which fall into two broad sets.

[3]We will discuss these papers in greater detail in the next section.

[4]While a widely used term, there does not appear to be a consensus on the exact years that the "Great Compression" begins and ends. The term appears to originate with [73] and in their paper refers more specifically to the 1940s. The Wikipedia entry, however, suggests the "Great Compression" ends around 1970. In this paper, we will use it to refer to, roughly, 1940–1970, but nowhere does this exact definition matter for any result.

Our first set of results makes heavy use of the fact that our data sources are all *microdata* at the household or individual level, allowing us to examine over an eighty-year period how household union status covaries with characteristics like education, race, and household income. We consider many of the stylized facts about unions established with CPS data, and extend them to earlier decades. We begin by showing that patterns of selection into unions has varied substantially over time: the education of union members relative to non-union members has followed a marked U-shape pattern, mirroring the pattern of inequality itself and the sharp inverse of union density. That is, during the Great Compression period, when density was the highest, unions were drawing in the least skilled workers. Today, as in the 1930s, unions are smaller and their members at least (if not slightly more) educated than the rest of the workforce. A similar pattern emerges for minorities: unions were relatively less white during the Great Compression period than either before or after.

A key stylized fact about CPS-era unions is that members enjoy a wage premium, but did this advantage exist as union density was growing in the 1930s and 1940s or during unions' Great Compression heyday? We show that the income advantage accruing to union households relative to non-union households with the same demographics and skill proxies is roughly constant (between fifteen and twenty log points) over our eighty-year period, despite the huge swings in union density and composition. While extrapolating selection patterns from the CPS era back to the more distant past would have been problematic, remarkably, the union premium from CPS data is a decent estimate for the entire post-1935 period. The union premium is larger for the less-educated (for every additional year of education, the union premium falls by roughly four log points), an effect that is also remarkably stable over our entire sample period. Finally, we observe similarly constant patterns for the ratio of *residual* income variance in the union versus non-union sectors (roughly 0.60 over our entire sample period), suggesting that not only have unions consistently reduced the differentials

paid to observed characteristics like education, they have a similar (and stable) effect on the return to unobservable traits.

Together, the U-shape in selection by skill and the constant patterns in union premia suggest that during the Great Compression period, unions were conferring a substantial advantage to what would otherwise have been low-income households, thus compressing the income distribution. Importantly, these stylized facts also relate to the potentially endogenous relationship between unions and SBTC posited in recent papers. These papers model union density as an explicit function of SBTC and predict certain co-movements in density, the union premium and selection into unions by skill. For example, [3] predict that skilled workers will differentially exit unions as SBTC increases, as their non-union market wages are bid up, whereas our data show the opposite pattern. Similarly, this model also implies that the union premium falls as SBTC rises, as the scope for rent-sharing between high- and low-skill workers narrows, a prediction that is also rejected by our data. While union membership undoubtedly responds to market forces, we argue that, in light of these new stylized facts, such forces alone cannot explain patterns in union membership over the 20th century.

In our second set of results, we move beyond documenting stylized facts consistent with the role of unions depressing inequality, and instead more explicitly model inequality as a function of union density. We begin by estimating distributional regressions following [61], where we show how measures of inequality vary as union density is perturbed. This exercise is most directly related to the stylized facts we document on selection and the union premium. Roughly speaking, for each year of our data, we estimate the effect of unions on the unconditional household income distribution, accounting for the changing position of union households in the counterfactual non-union income distribution as well as the union income premium. Across our eighty-year sample period, we find a consistent negative effect of reweighting the full income distribution toward the union income distribution on both the

Gini coefficient and the 90/10 ratio. As would be expected given the changes in selection documented earlier, the negative effect of unions on inequality is especially large in the Great Compression period, when unions were organizing the most negatively-selected workers.

Next, we turn to regression analysis where instead of micro data we employ annual aggregated data and include union density as an explanatory variable. We begin by simply adding union density to the canonical regressions estimated by [85], [12] and [74], who use aggregate time-series regressions to show that the supply of skilled workers is a strong, negative predictor of the college-wage premium. We then refine our time series analysis by introducing variation at the state level. This parallel analysis, aggregated at the state-year level, is carried out controlling for state and year fixed effects. We are able to implement this analysis only because the Gallup data always include a state-of-residence variable and thus allow us to calculate union density at the state-year level (existing measures of union density before the CPS are only consistently available annually at the national level). In both the annual and the state-year regression analyses, union density has a negative effect on standard measures of inequality such as the skill premium, the 90/10 ratio, the Gini coefficient, and the top-ten-percent income share. While each of these exercises depends on a different set of (admittedly strong) identifying assumptions, they each yield a negative and significant effect of union density on measures of income inequality, in many cases comparable in magnitude to and independent of the effect of skill shares.

The rest of the paper is organized as follows. In the next section, we discuss the existing literature on inequality and unions and our contribution. Section 3.3 describes our data sources, in particular the Gallup data. Section 3.4 presents our new time-series on household union membership. Section 3.5 analyzes selection into unions, focusing on education and race. Section 3.6 estimates household union income premiums over much of the 20th century, and Section 3.7 presents our evidence on the effect of unions on the shape of the overall income distribution. Section 3.8 offers concluding thoughts and directions for future work.

3.2 The Economics of Unions and Inequality

Economists' interest in the effects of unions on workers goes back at least to Adam Smith, who noted asymmetries between combinations of employers versus workers, and John Stuart Mill, who claimed that in the presence of imperfections "trade unions...are the necessary instrumentality of [the] free market."[5] The literature on unions is vast, and below we provide a selective review of those papers most directly related to the relationship between unions and inequality.

The modern empirical neoclassical approach towards measuring the effect of labor unions was pioneered by [97]. This literature has largely focused on estimating the union premium: the differential in wages between union and (otherwise similar) non-union workers. The early analysis by Lewis generally focused on industry-level differences, as consistent sources of microdata were not yet available. [66] were among the first to use CPS microdata to estimate determinants of union membership and the union premium with individual-level data. They find a union premium of roughly sixteen percent, averaging across studies in the 1970s. In general, a 0.10 to 0.20 log-point union premium—controlling for Mincer-type covariates and estimated on cross-sectional wage data such as the CPS—has been found consistently in the literature.[6]

The underlying market structure is important for interpreting the welfare impacts of a significant union premium. A classical view is that labor unions are monopoly sellers of labor to the firm, lowering employment in exchange for higher wages at firms that face otherwise competitive conditions [114]. Unions thus create deadweight losses, and their decline should be associated with greater efficiency. Another view is that unions induce firms to share with workers the rents from product market power. Globalization, technological change, and

[5] See [107] [p. 319 (original publication year 1848)].

[6] As noted, there is very limited work examining the pre-CPS era. [25] is a related recent paper that uses detailed microdata from a survey of six cities in 1951 to estimate a union premium comparable in magnitude to what we find in Section 3.6. [98] is an earlier survey of the literature on union premia, with the earliest micro estimates from the late 1960s.

deregulation all may have lowered rents available to be redistributed, but there could still remain a role for unions to change the split of the remaining rents between workers and firms (as in [1] and [2]), potentially mitigating increases in inequality induced by market forces, and with ambiguous welfare implications. Consistent with this older rent-sharing literature, a recent literature has documented both pervasive firm- and establishment-specific wage premia [119, 29, 16], with unionized firms having larger premia [81]. Further, if, as a recent and rapidly growing literature has suggested (see [103] and [109] for recent surveys), labor markets are monopsonistic, then union bargaining could raise worker wages to their efficient level (as the quote above from Mill suggests).[7] While we lack the data to distinguish these interpretations of the union premium, they should be kept in mind throughout our paper.

A key challenge in this literature is separating any causal effect of union membership on wages from non-random selection into unions based on counterfactual non-union wages. On the one hand, if higher union wages create excess demand for union jobs, then union-sector employers have their pick of queueing workers and unobserved skill could be higher in the union sector, overstating the union premium. On the other hand, a higher union wage premium for less-skilled workers and union protections against firing might differentially attract workers with unobservably less skill and motivation. [26] uses panel data (allowing him to examine workers as they switch between the union and non-union sectors) to show that the union premium remains significant even after accounting for negative selection at the top and positive selection at the bottom.[8] Recent identification of the effect of unions

[7] A recent paper by [17] shows that local employer concentration has increased considerably across labor markets over the last forty years. Moreover, this increased concentration is associated with significantly lower wages *except* in highly unionized labor markets.

[8] [95] performs a similar exercise using Canadian data, with the added advantage that he can focus on *involuntary* switchers. He finds that estimates that account for non-random selection are in fact quite close to OLS estimates of the union premium.

on firms and workers has used discontinuities in union-recognition election results in U.S. firms [42, 93, 63].[9]

A related empirical literature focuses more directly on the connection between unions and the shape of the income distribution. It is helpful to separate this literature into two conceptual categories. First, assume that unions affect the wages of *only* their members and that estimates of the union premium can recover this causal effect (so, put aside the selection issues raised in the previous paragraph and potential spillovers to the wages of non-union workers). Then, simple variance decompositions can estimate the counterfactual no-union income distribution and thus the effect of unions on inequality. For example, so long as unions draw from the bottom part of the counterfactual no-union wage distribution, then their conferring a union premium to this otherwise low-earning group reduces inequality. Moreover, residual wage inequality also appears to be lower among union workers, suggesting that unions reduce inequality with respect to unobservable traits as well [27]. [40] and [61] take this approach and find that unions substantially reduce wage inequality, especially for men.

A second category of papers argues that unions affect the wages of *non-union* workers as well. Unions can raise non-union wages via union "threat" effects (57; 121) or by the setting of wage fairness norms throughout an industry [124]. Conversely, unions can lower non-union wages by creating surplus labor supply for uncovered firms [97]. Unions might also affect the compensation of *management* (113; 70) and the returns to capital (2; 93; 41), thus reducing inequality by lowering compensation in the right tail of the income distribution. Finally, as an organized lobby for redistributive taxes and regulation, unions might affect the income distribution via political-economy mechanisms (94; 4). Given these diverse mechanisms, the effect of union membership on the income distribution might be larger or smaller than that

[9] Although see [64], who cautions against validity of this IV by showing that unions disproportionately lose (win) "close elections" in years when Republicans (Democrats) have a majority on the National Labor Relations Board.

implied by micro-data analysis of the union wage premium and selection into unions, a point we return to in Section 3.7.

A final group of papers of particular interest to us is a small macro-theoretical literature on the relationship between unionization, technology, and inequality. These papers argue that any empirical relationship between unions and inequality is spurious, driven by variation of an omitted variable—skill-biased-technological change—that simultaneously reduces union density and increases inequality. For example, in [3], skilled workers earn more in the non-union sector relative to the compressed union sector, and this wedge grows with SBTC. In [5] both this mechanism as well as selection of firms into the union sector are at play, as in [56], and increases in SBTC contribute to both high-skill workers not wanting to be in unions as well as firms being increasingly unwilling to hire low-skill workers at the union wage. These papers both predict that skilled workers should opt out in increasing numbers as the skill-premium increases, lowering union density. While the opting out of skilled workers is an equilibrium result in Acemoglu *et al.* and Açıkgöz and Kaymak, [43] rule out the unionization of skilled workers by assumption (they argue that skilled workers are too diverse for unions to organize). Thus, union density in their model declines as SBTC rises because unskilled, unionized workers become relatively more expensive to employ and unions have, by construction, no option to organize the more skilled.

We see three key contributions in extending micro-data analyses of unions back to the 1930s. First, economists' understanding of the basic economics of U.S. labor unions—the size or stability of the union premium, selection into unions by education or other proxies for non-union wages, differences in residual wage variance between the covered- and non-covered sectors—relies almost entirely on CPS data and is thus limited to 1973 and later. We use our new micro-data to examine these stylized facts going back to 1936 (Sections 3.5 and 3.6). Importantly, tracing out how selection and the union premium varies during the decline, at the nadir, and then during the rise of U.S. income inequality sheds light on whether

unions are a plausible factor in explaining the time-series pattern of inequality. Second and related, the recent macro-theoretical models described above have distinct predictions for the co-movement of union density, skill composition, and premia that could only be tested in a limited way with CPS-era data. In fact, our data reject many of these predictions, suggesting that union density is not merely an artifact of skill-biased technological change. This finding helps motivate our third contribution, in which we model inequality as an explicit function of union density (Section 3.7) in regression analyses.

3.3 Newly Available Data On Union Membership

As noted in the introduction, one of the main contributions of the paper is the introduction of newly available household-level data that include information on union membership. We draw much of these data from public opinion polls, which have recently been posted online by the Roper Center at Cornell.[10]

Polling has a long history in American life. The earliest systematic polls were conducted by magazines, in particular *Literary Digest*, which would include a returnable postcard with opinion questions to conduct "straw polls" on the issues of the day [82].[11] Beginning in the late 1930s, George Gallup, Elmo Roper, and Archibald Crossley began importing techniques from market research into the domain of public opinion polling.

Gallup established the American Institute of Public Opinion (AIPO) and set out to conduct nationwide surveys of American opinions on a range of social and political issues.[12] Gallup was scrupulously non-partisan, never running polls on behalf of a particular party. AIPO also devoted considerable efforts to develop neutral, easy to understand question

[10] See `https://ropercenter.cornell.edu/ipoll-database/`.

[11] The Earliest *Literary Digest* poll we could find referenced was a poll to predict the outcome of the 1916 presidential election.

[12] Similar organizations were formed at roughly the same time: Roper's company was steadily employed by Fortune magazine starting in 1935, Henry Cantril started the Organization of Public Opinion Research (OPOR) in 1940, and the University of Chicago's National Opinion Research Center (NORC) was founded in 1941.

wording. By 1940, about eight million people had encountered Gallup's tri-weekly polling report, *America Speaks!* Gallup and other pollsters made money by selling their results to businesses for consumer research and newspapers for public opinion.

3.3.1 Gallup Methodology before and after 1950

Before 1950, Gallup used so-called "quota-based" sampling. Survey-takers had to fill quotas for each pre-determined strata thought to capture distinct political views. Enumerators were given both hard (e.g., gender, must have one-third female) and soft (e.g., age, "get a good spread") quotas, but within each quota, interviewers had a lot of discretion. As [18] notes, "interviewers preferred to work in safer areas and tended to question approachable respondents," which likely led to Gallup over-sampling, within each quota strata, more prosperous and well-off respondents.[13]

Gallup once noted that the "the voting public...is the universe of the opinion researcher," suggesting his aim was to be representative of *voters*, which implies substantial underrepresentation of certain segments of the population. Presumably because the South had low turnout (given many of its elections during this time did not even manage a Republican challenger), it was under-sampled. Southern blacks were differentially underrepresented among Southerners, consistent with their near total disenfranchisement during this period. Gallup purposely over-sampled men because of a belief that women merely adopted their husbands' opinions on Election Day.[14]

Consistent with discretion within the quota-based sampling leading to oversampling of the well-to-do, Gallup over-predicts the Republican vote share in 1940 and 1944, though

[13] [18] provides great detail on Gallup's quota-based sampling procedures, from which we draw much of the information in this subsection.

[14] It is worth noting that any oversampling of men is not a substantial problem for our purposes since we are interested in measures of union status and income at the *household* level. Since most men and women are combined in households, particularly in the earlier years, reports of "any union members in the household" and "household income" should not be affected by whether the surveyed individual in the household was male or female.

in both cases he still correctly predicts Roosevelt victories. In 1948, this over-sampling of Republican voters leads him to incorrectly call the election.

From 1950 onward, Gallup uses modern-day probabilistic sampling procedures. Weights are often provided, but their documentation is not consistent. As a result, in our analyses of the Gallup data we use weights that we generate from the Census, as detailed in Appendix C.3. Before 1942, our weights adjust Gallup demographics at the $region \times race$ level. From 1942 onward, we adjust at the $region \times race \times education$ level (because, before 1942, education variables are not consistently available in Gallup).

Appendix Table C.1.1 and C.1.2 compare *unweighted* Gallup data with their closest Census year. We provide greater detail for the 1940 comparison (when Gallup is thought to be at its worst in terms of representation of the full adult population). As expected, Gallup oversamples the educated and substantially undersamples women, though these biases are far smaller by 1950. As we only have two surveys in 1940 that include education, we also show detailed occupational breakdowns (doing our best to match Gallup and Census categories). Reassuringly, major categories like farming appear to be well represented in Gallup in 1940. Again as expected, the South is undersampled in 1940, though outside the South, the remaining regions are all represented roughly equal to their population and importantly the black share *outside* the South is very similar in Gallup and the Census.

In the coming sections, we focus on results using *weighted* Gallup data, but we show some unweighted results for the sake of comparison (and all other unweighted results are available from the authors). In general, weighting makes little difference to our regression results.

3.3.2 Additional checks on pre-1950 Gallup data

Given readers may have natural concerns about the earlier Gallup data, we present a few more plausibility checks that we can assess in *non*-Census years. In Appendix Figure C.1.1, we show that our Gallup unemployment measure matches in changes (and often in levels) that of

the official Historical Statistics of the United States (HSUS). The concordance is reassuring, given that a concern voiced by pollsters at the time was that the surveys were prone to sample the relatively well-off (both due to enumerator discretion and quotas tilted towards voters). We in fact see large shares of Gallup respondents during the so-called "Roosevelt recession" of 1937-38 report that the household head is unemployed (note that levels are not directly comparable as Gallup and HSUS use different definitions of unemployment in the 1930s).[15] Note that, as with most of our results, weighting the Gallup data has little effect.

As another test of whether Gallup can pick up high-frequency changes in population demographics, Appendix Figure C.1.2 shows the "missing men" during World War II deployment: the average age of men increases nearly three years, as millions of young men were sent overseas and no longer available for Gallup to interview.

3.3.3 Additional data sources

While we rely heavily on the Gallup data, we supplement Gallup with a number of additional survey data sources from the 1930s onward. Gallup does not ask family income for much of the 1950s, but the American National Election Survey (ANES) asks both family income and union household status throughout that period, so often we will we augment our Gallup data with the ANES.[16] We also found a handful of additional survey datasets that ask union status as well as the other variables we need to estimate a family income premium (i.e., education, family income, state of residence and basic demographics). The first is an expenditure survey that the BLS conducted in 1936–1937, which we will henceforth refer to as the 1936 Expenditure Survey.[17] The second is a 1946 survey performed by the U.S.

[15]Gallup unemployment is measured as the share of households who report being unemployed (including, sometimes, those who are in fact out of the labor force), while the HSUS reports unemployed workers as those who had no employment, were available for work, and had made "specific efforts" to find employment. The HSUS unemployment rate is the share unemployed out of the employed plus unemployed population.

[16]The ANES has a relatively small sample size in any given year so that our ability to use the ANES to provide detailed breakdowns of union status and income by geography or demographics is limited.

[17]Note that interviews conducted in 1937 refer to consumption and income in 1936. The survey asks about *union dues* as an expenditure category, and thus can be used to measure household union membership.

Psychological Corporation. In 1947 and 1950 we use data from National Opinion Research Corporation (NORC) as a check on our union density estimates from Gallup, but, as these data do not have state identifiers, we do not use them in micro-data regression analysis. From 1977 onward, we can use the CPS to examine household measures of union membership.[18]

Summary statistics for the CPS, ANES as well as for these additional data sources appear in Appendix Table C.1.3. In general, at least along the dimensions on which Gallup appears most suspect in its early years (share residing in the South, share white, education level), these data sources appear more representative. The table shows all data sources unweighted, though we will use ANES and CPS weights in years they are provided, to follow past literature. We do not weight the other additional surveys.

3.3.4 Existing time series of aggregate union density

As noted in the introduction, aggregate counts of union members exist going back to the early 20th century. We describe the two main time series—the BLS and Troy series—in greater detail in Appendix C.4 and give only a very brief summary here.

The BLS series is based on unions' reports of their aggregate membership counts. A key issue, which we describe in detail in the Appendix, is that individual unions had an incentive to overstate membership. Especially from 1937–1955, when the main U.S. labor umbrella organization had split into two warring factions—the American Federation of Labor and the Congress of Industrial Organizations—the two federations over-stated their membership in attempt to gain advantages over the other.

The Troy series was compiled by Leo Troy under the auspices of the NBER. He attempts to correct some of these over-reporting issues in the BLS, but in many years he is forced to rely on the same membership statistics, while in other years he derives membership by

[18]We could in principle go back to 1973 using the CPS, but only beginning in 1977 does the CPS include both the union membership question and state (as opposed to merely state groups or regions) identifiers, and as we are generally controlling for state in all our results, we want to remain consistent with analysis from our other data sets.

dividing total revenue by average dues rates. However, as we discuss in Appendix C.4, revenue data were also believed to be vastly overstated.

In fact, unions had no obligation to report membership to the government until 1959, when every union involved in interstate commerce had to file reports. From this point onward, these aggregate counts are believed to be more accurate [122].

3.4 A New Time-series Measure of Household Union Membership

The typical Gallup union question is "Are you (or is your husband) a member of a labor union?", with the choices most often being: "neither," "yes, I am," "yes, he is," "yes, both are." In 1959, "husband" changes to "husband/wife." In some years, however, the question does not ask which member or members of the household is or are in a union, so we cannot, for example, always measure individual union status. We harmonize these questions to form a measure of *household* union status, where we code a household as union if either household head or spouse is a union member. While technically the implied unit of observation is *couple* (or individual if the respondent is not part of a couple), we will generally refer to this measure as *household union status*. Importantly, Gallup asks this question of *all* respondents, not skipping those in, say, agricultural occupations or who are unemployed.

3.4.1 The union share of households over time

Figure 3.2 plots our (weighted) Gallup-based measure of the union share of households, by year, alongside several other series.[19] The Gallup series bounces around between eleven and thirteen percent from 1937 to 1940. Between 1941 and 1945, the years the U.S. is involved in World War II, union membership in our Gallup data roughly doubles. Union membership

[19] Appendix Figure C.1.3 shows this series alongside its unweighted analogue. In most years, the two series look very similar.

continues to grow at a slower pace in the years immediately after the war, before enjoying a second spurt to reach its peak in the early 1950s. After that point, union membership in the Gallup data slowly but steadily declines.

Also presented in Figure 3.2 are our supplemental survey-based series. Note that each of these series generally has fewer observations per year than Gallup. The ANES sits very close to Gallup, though as expected is noisier. The 1936 expenditure survey is very close to our earliest Gallup observation, in 1937. The U.S. Psychological Corporation appears substantially lower than our Gallup measures in 1946, whereas the two NORC surveys (from 1947 and 1950) sit somewhat higher than Gallup estimates for those years.

Finally, Figure 3.2 plots the widely-used historical data series described in Section 3.3.4, the BLS and the Troy series.[20] To avoid clutter, we do not plot our CPS series in this figure. Appendix Figure C.1.4 shows the Gallup and CPS series, zoomed in on the period of their overlap (1977-1986), so that readers can more easily assess their degree of concordance. Reassuringly, in the years when Gallup and the CPS overlap, they are quite close.[21]

3.4.2 Differences between Gallup and the historical aggregate series

While the Gallup measures do not always agree with the BLS and Troy series in levels, they are, for the most part, consistent in changes.

The largest differences exist in the early years of the series, with much more agreement in the 1950s, when both union reporting and Gallup sampling improve. As noted in Appendix C.4, we have reason to believe that both the Troy and the BLS series over-stated member-

[20]Recall that these series give aggregate union *counts* of membership, so we divide by estimates of total U.S. households (geometrically interpolated between Census years) to make the numbers as comparable as possible to Gallup. This transformation will obviously overstate the union share of households if many households had multiple union members.

[21]Give the labor-intensity of reading in the Gallup data, we do not continue past 1986 and beyond this point rely on the CPS. We cut off at 1986 in order to have a ten-year period where Gallup and CPS overlap, so we can check consistency of Gallup over a substantial period of time.

ship. While unions had incentives to overstate membership, respondents themselves had no incentive to tell Gallup survey-takers they were union members when they were not, so this bias is unlikely to affect the Gallup numbers. However, there are reasons that Gallup and other opinion surveys may *under-state* union membership relative to these historical series based on union reports.

First, there is a legitimate possibility that individuals are union members without knowing it, especially during certain historical moments, meaning union reports would accurately count them as members but they would (truthfully, but potentially inaccurately) tell Gallup that they were not currently members. During World War II, some unions default-enrolled all new workers and automatically collected dues from workers' paychecks (workers would have to actively take steps to dis-enroll). Workers could thus be members of unions without knowing it. During the war, a period of rapid union growth, the National War Labor Board (NWLB) gave unions this privilege—default enrollment of new workers—in war-related plants, in exchange for a no-strike pledge [99]. It is thus perhaps not surprising that the increase in membership during the war, while large in the Gallup data, is even larger in the union-membership-reports-based series.

Second, and related, moments of high unemployment complicated calculations of union density. Until Congress mandated annual reporting in 1959, unions had great discretion in how to count a union member who became unemployed, whereas an unemployed respondent in Gallup (who is no longer paying dues) might well consider himself no longer a member.[22] Indeed, Figure 3.2 shows that the Gallup estimate diverges from the BLS and Troy estimates the most in 1937-38 recession period. Gallup shows essentially no growth during this period, whereas the BLS and Troy show robust growth, despite soaring unemployment.[23]

[22]As noted, Gallup and ANES did *not* skip over the unemployed or those otherwise out of the labor force when fielding their union question, and many unemployed and retired respondents in these surveys nonetheless identify as union members.

[23]Indeed, it is well documented that at least among the largest locals where data are available, *dues payments* plummeted for CIO unions during this period, as millions of workers were laid off [99].

Third, as noted, Gallup over-samples the well-to-do, especially before 1950, which likely biased union membership toward zero. But offsetting sampling biases—most notably the under-representation of the South, a region historically hostile to unions—likely overstate density estimates. While we can never fully discount the possibility that non-representative sampling is causing Gallup to understate density, that we only find a marginal changes in density after applying weights suggests it is not a major factor (see Appendix Figure C.1.3).

In summary, we are not surprised that the union density measures based on opinion surveys differ slightly in levels from the more widely used measures in the literature, given non-trivial differences in methodology. However, in almost all cases they firmly agree in changes, and, like the Troy and BLS series, Gallup exhibits the same inverted U-shape over the twentieth century. Moreover, as we will show in Section 3.7, the relationship between aggregate union density and inequality is very similar whether we use our new, micro-data-based measures of household unionization rates or the traditional, aggregate measures.

An important advantage of our series, however, is that it is based on micro data, which allow us far greater scope to sort among theories relating inequality to unionization. It is to this task we now turn.

3.5 Selection Into Unions

The analysis in this section exploits a key feature of our microdata sources: they allow us to examine the joint distribution of unions and other covariates over time. Of particular interest is the relative skill of union members relative to non-union workers. First, how selection into unions varies across time is informative about the plausibility of the claim that unions have played an important role in compressing the income distribution. For example, suppose we find that the expansion of unions from, say, 1940 to the 1960s was all among relatively *high*-skill workers. Then it would cast doubt on the claim that unions contributed to the mid-century Great Compression in wages, as they would be conferring an advantage

to already highly-paid workers. Second, as discussed in Section 3.2, some existing models of unionization, technology and inequality predict that the relative skill of union members will decline (rise) as union density declines (rises). Plotting selection by skill across time allows us to test these predictions.

We take two approaches to examining the relationship between selection into unions and union density. We begin by estimating a basic union selection model (regressing a union indicator on demographics and skill proxies) by year, and reporting the coefficient on skill proxies across time. This approach allows the data to "speak for themselves" in a transparent manner. If there is indeed a positive (negative) relationship between union density and the relative skill of union members, then these coefficient estimates across time should take an (inverted) U-shape, as union density across time is itself inverted-U-shaped. We then take a more parametric approach and include density as an explanatory variable in a union selection regression model.

3.5.1 Selection into union across time

We begin our education selection analysis by estimating the following equation by survey-source d and year y:

$$Union_{hst} = \beta_{dy} Educ_h^R + \gamma_1 Female_h^R + f(age_h^R) + \mu_s + \nu_t + e_{hst}. \tag{3.1}$$

In this equation, subscripts h, s, and t denote household, state and survey-date, respectively. The superscript R serves to remind readers that in many cases, a variable refers specifically to the *respondent* (not necessarily the household head, though via assortative mating by race, age and education, the respondent, if she is not the head herself, will often have similar demographic characteristics). $Union_{hst}$ is an indicator for whether anyone in the household is a union member (and is the underlying household-level variable we use to construct the

aggregate time-series in the previous section). $Educ_h^R$ is the respondent's education in years.[24] $Female_h^R$ is a female dummy, $f(age_h^R)$ is a function of age of the respondent (age and its square when respondent's age is recorded in years, fixed effects for each category when it is recorded in categories), and μ_s and ν_t are vectors of state and survey-date fixed effects, respectively.[25]

The vector of estimated β_{dy} values tells us, for a given year y and using data from a given survey source d, how own education predicts whether you live in a union household, conditional on basic demographics and state of residence.[26] Note here that we are not yet controlling for race.

Figure 3.3 shows these results across our key datasets. A marked U-shape emerges, with the year-specific point-estimates remarkably consistent across all data sources.[27] In the earliest years (1936 through the mid-1940s, say) the coefficients suggest that an additional year of education reduces the likelihood of living in a union household by about 2 to 2.5 percentage points. At the trough of the U (around 1960), we estimate that an additional year of education reduces the likelihood of living in a union household by roughly four percentage points. Since the 1960s, the negative marginal effect of education on the probability of living in a union household declines steadily: it reaches zero around 2000 and is now positive and statistically significant, though small.

This change may reflect, in part, the substantial growth of public sector labor unions since the 1960s. Since at least the early 1970s, the public sector workforce has had higher

[24] Where a specific survey does not collect information directly on years of education but reports specific ranges or credentials, we use simple rules to convert these measures to years of education. The note to Figure 3.3 describes how we impute years of education in these cases. We also carry out some analyses with alternative measures of education (e.g., a high school dummy) to probe robustness.

[25] Year fixed effects are effectively included in the model in the form of the *survey-date* fixed effects (the ν_t).

[26] For the ANES, given the small sample sizes, we constrain the coefficients on education (β_{dy}) to be equal across six-year bins in order to reduce sampling error. For the Gallup and other surveys, we estimate the coefficients on education (β_{dy}) by estimating separate regressions for each *survey source* $\times$ *year* combination.

[27] This pattern holds when other education measures are used instead of years of schooling. Appendix Figures C.1.5, C.1.6, and C.1.7 show similar patterns when log years schooling, a high-school dummy or college dummy serve as the education measure.

education than the private sector workforce. And, importantly, unions have been much more successful organizing workers in the public sector than in the private sector.[28] While we do not know sector for the Gallup, Psych. Corp., and 1936 expenditure surveys, we can compare our baseline selection patterns from the ANES and CPS to those when we drop any household with a public sector worker. As Appendix Figure C.1.8 shows, while the *levels* of the selection effect change slightly for this sample, the *increase* in the skill of union households from 1970 onward is unchanged.

Another possibility is that the steep decline in manufacturing employment since the 1960s drives the most recent half of the U-shape in our selection results. Appendix Figure C.1.8 also shows selection patterns after dropping households with *either* a public-sector or a manufacturing worker. A large majority of the up-skilling effect remains.

As a final robustness check on our education selection results, we experiment with different weighting schemes. As noted in Section 3.3, sampling before 1950 in the Gallup data may be suspect (and as we showed in Appendix Table C.1.2, especially suspect in terms of over-sampling the educated, at least when we compare them to Census estimates in 1940). However, Appendix Table C.1.4 shows that selection by education in Gallup before 1950 barely changes no matter how we weight (including not weighting at all).

We next examine selection by race, which is important for at least two reasons. First, given that school *quality* is an often unobserved dimension of skill [30] and blacks have always attended lower-quality schools than whites, race may serve as another proxy for skill. Second, selection of union members by race over time is an important (and unresolved) historical question. Historians disagree on the degree to which unions discriminated against black workers during the pre-World War II and Great Compression period, with some arguing that unions exacerbated white-black wage gaps and others arguing that, at the very least,

[28] Over the period from 1973-2016, tabulation of CPS data indicates that 5.3 percent of college graduates employed in the private sector were members of labor unions. In contrast, fully 39.7 percent of college graduates employed in the public sector are union members.

unions were less discriminatory than the non-union labor market more generally during this period (10, 110; 62; 88; 86).

We analyze selection by race in the same manner as selection by years of schooling, and simply replace $Educ_h^R$ with $White_h^R$ in equation (3.1).[29] We start by examining selection by race *unconditional* on education, because Gallup always asks about race but only beginning in 1942 does it consistently ask about education and we want to maximize the time period over which we can estimate selection by race. We return to selection by race conditional on education later in the section.

The estimated coefficients on $White$ across time and data sources are presented in Figure 3.4. Again, a U-shape emerges. In the beginning of our sample periods, whites are (conditional on our covariates) more likely to be in union households than non-whites. This advantage disappears and in fact reverses during the war years and continues to grow more negative until about the 1960s. Since then, whites gain on non-white households until today there is essentially no differential.

While not quite as consistent as for education, selection by race again agrees for the most part across data sources. There is some disagreement between the ANES and Gallup in the 1950s and 1960s, though the large standard errors around the ANES estimates certainly do not allow us to reject equality. There is some disagreement between Gallup and CPS, whereby Gallup shows minimal negative selection with respect to education by the early 1980s, whereas CPS shows that whites are still somewhat less likely to live in union households. However, by the end of the sample period, there is no remaining selection by race in the CPS either.

Appendix Figure C.1.9 replicates this analysis, but excludes the South, given that Gallup significantly under-sampled blacks in the region at least through 1950 and its correction of

[29]Results are essentially exactly the inverse when instead of $White$ we use a black dummy. We use $White$ instead because sometimes Gallup uses "negro" and sometimes "non-white" and thus $White$ would appear, in principle, a more stable marker.

this bias over time might make comparisons of union selection by race across time difficult. In fact, the results are very similar.

3.5.2 Modeling selection as a parametric function of density

So far, we have shown that, graphically, selection into unions by proxies for skill takes an inverted U shape from 1936 to today, the inverse of the shape of the union density time-series. In this section, we pool our data and include union density itself as a regressor. Relative to the graphical evidence, this analysis allows an easier quantification of the relationship between density and selection and also facilitates robustness checks. Robustness checks are especially important because other key trends over the twentieth century are U-shaped, so any relationship between density (inverted-U-shaped) and selection (U-shaped) could easily be spurious and driven by some third factor.[30]

We begin by examining selection by education, modifying equation (3.1). Here, we pool all years of CPS and Gallup data (results are unchanged if we also add ANES and our one-off datasets) and estimate variants of the following equation:

$$\begin{aligned} Union_{hst} = &\beta_1 Educ^R_{hst} \times Density_t + \beta_2 Educ^R_h + \\ &\beta_3 Educ^R_h \times Year_t + \beta_4 Educ^R_h \times Year^2_t + \\ &\gamma_1 Female^R_h + \gamma_2 White^R_h + f(age^R_h) + \nu_t + \mu_s + e_{hst}, \end{aligned} \tag{3.2}$$

where $Density_t$ is aggregate union density in year t as estimated in our Gallup/CPS data and all other notation follows naturally. Including the $Educ^R_h \times Year_t$ and $Educ^R_h \times Year^2_t$ terms allows selection by education to take an arbitrary U-shape of its own. Note that, given the evidence we just documented that selection into unions by race is also U-shaped,

[30] Two important examples of U-shaped trends over twentieth-century America are the share foreign born and political polarization in Congress.

we include a control for race. As noted, a surprising number of trends appear to take a U or inverted-U shape over the period we study. This specification allows us to examine whether there actually is a direct relationship between union density and the skill and demographic composition of union members, or is composition following, perhaps coincidentally, its own inverted U pattern or being driven by some other U-shaped trend.

To provide a baseline, col. (1) of Table 3.1 estimates equation (3.2) with the restriction that $\beta_3 = \beta_4 = 0$ (so, not allowing selection by education to follow its own U-shape). The coefficient of interest on $Educ^R_{hst} \times Density_t$ is large and negative, as we would expect from the graphical evidence in the previous subsection. Col. (2) allows selection by education to follow its own U-shape, and indeed both of these terms are significant and of the expected sign. As we would expect given the collinearity, the coefficient on $Educ^R_{hst} \times Density_t$ falls (by just over half) but remains large, negative and highly significant. As a final check, in col. (3) we drop any household in the CPS that has a public-sector worker, and the coefficient of interested is largely unaffected.[31] Col. (4) shows that higher union density is associated not only with more unskilled workers joining unions, but also not non-whites joining as well (even allowing selection by race to follow its own, arbitrary U-shape), though the final column of this table shows that this relationship is sensitive to dropping public-sector households.

Thus, much of the prominent U-shape in Figures 3.3 and 3.4 *appears correlated with variation in union density* (which itself has an inverted U-shape), and is not merely following some arbitrary U-shape of its own. As a further robustness check, we confirm in Appendix Table C.1.5 that *state-year* union density also predicts greater unionization rates among the less-educated and minorities.

[31] We cannot drop public-sector workers in Gallup, as Gallup does not designate sector. However, given the rise of public sector unions is not until the 1970s and in these regressions we only use Gallup observations until 1976, the vast majority of unionized households in our Gallup data are working in the private sector.

3.6 The union family income premium over the twentieth century

The analysis in Section 3.5 shows that union households were much lower skilled during the period of peak union density than either before World War II or in recent periods. We now use the survey microdata to estimate union premia, both on average and by skill level over over nine-decade period. Given that almost all evidence on the size of the union premium is from the CPS era, estimating union premia from earlier periods is interesting it its own right.

Measuring the premium over longer periods of time is important for at least two other reasons. First, it helps us assess the plausibility of the argument that unions reduce inequality: for example, if the union premium was much smaller in the immediate post-war period, then it would be harder to argue that unions were conferring advantages to otherwise low-wage workers during the Great Compression period and beyond, as we claim. Second, some models of SBTC-driven union density imply that the union premium should increase (shrink) as density grows (declines), a prediction we can test with our data.

As in the previous two sections, we draw from a variety of data sources. From Gallup, we can estimate a premium in 1942 and then each year beginning in 1961 (though we drop any year in which the top-coded category contains more than thirty percent of observations). Gallup occasionally asked household income before 1961, but only in 1942 do we have a survey that also includes the union question and education. Our other major data source is the ANES. We can estimate a household union income premium in the ANES in 1952, and then every other year after 1956.[32]

[32]As seen in Figure 3.2, we can estimate union density in the ANES in 1948, but that survey does not include household income.

As noted in Section 3.3, we supplement these surveys with any other survey we are able to find that asks union membership, household income and education.[33] Recall that these supplemental surveys are the 1936 Expenditure Survey and the 1946 U.S. Psychological Corporation survey.[34]

3.6.1 Estimating the union family income premium over time

Across all these surveys, we are able to estimate the following regression equation (as with selection, we estimate it separately by data source d and year y):

$$\begin{aligned} ln(y_{hst}) =& \beta_{dy} Union_h + \gamma_1 Female_h^R + \gamma_2 Race_h^R + f(age_h^R) + \\ & g(Employed_h) + \lambda_h^{eduR} + \nu_t + \mu_s + e_{hst}. \end{aligned} \tag{3.3}$$

While we are estimating a *household income* function, we do our best to mimic classic Mincerian controls. In the above equation, y_{hst} is household income of household h from survey date t in state s; $Union_h$ is an indicator for whether anyone in the household is a union member; $Female_h^R$ and $Race_h^R$ are, respectively, indicators for gender and fixed effects for racial categories *of the respondent*; $f(age_h^R)$ is a function of age of the respondent (age and its square when respondent's age is recorded in years, fixed effects for each category when it is recorded in categories); $g(Employed_h)$ is a flexible function controlling for the number of workers in the household; λ_h^{educR} is a vector of fixed effects for the educational attainment of the respondent; and μ_s and ν_t are vectors of state and survey-date fixed effects, respectively. Note that for the 1946 U.S. Psychological Corporation and for the Gallup surveys from 1961

[33] In fact, the real constraint are surveys that ask both household income and union status. Almost all surveys we have found that ask for household income also ask respondents' education.

[34] The 1936 Expenditure Survey is the only one in which income is not binned. To make it comparable to our other data sources, we winsorize at 95% and 5% percent and then create 15 equally sized income bins.

onward, we cannot control for the number of workers per household, but below we show that this bias should be small.

As with our selection results in the previous section, we show premium results separately by survey source and year in Figure 3.5. Remarkably, the household income premium has not appreciably changed from 1936 to today, holding stable at roughly fifteen to twenty log points. While not a perfectly flat line, given the standard errors around each estimate, the family union premium does not appear to follow any discernible pattern over time. Of the sixty-some point estimates, only a handful are greater than 0.20 or less than 0.10. For none of these estimates do the confidence interval intersect zero.

[27], using CPS data, noted as a puzzle that the union wage premium was surprisingly stable between 1973 and 1993, even as private-sector union density declined by half. Our results, if anything, deepen this puzzle, as we show that the premium is constant over a *nine-decade* period that saw density (as well as the degree of negative selection) both increase and then decrease. This stability of the premium is not predicted by the existing models in which SBTC determines union density rates, as in those models the premium should dwindle as density declines. This result is also hard to rationalize with models that assume a union objective function that is a positive function of both union wages and membership, such as [43]. We have no clear resolution of this puzzle and indeed find it hard to write down a model of collective bargaining outcomes with standard union and firm objective functions that yields a constant premium in the face of declining density. We flag this question as a potentially fruitful area for future research.

3.6.2 Robustness and related results

While we do not focus on them, Appendix Table C.1.6 shows the coefficients on the Mincer equation covariates in equation (3.3). For ANES (since it covers a long time period), we split the sample in two, so that coefficients can be compared across time. In all cases, the

coefficients on the covariates have the same signs as we typically see from an individual earnings regression.

In Appendix Figure C.1.10, we show results after controlling for occupation of the household head. As noted, occupation categories vary considerably across survey sources, which is why we relegate this figure to the Appendix. The appendix figure reports coefficients that are somewhat larger than in the main Figure 3.5, consistent with unions differentially drawing from households where the head has a lower-paid occupation.

As noted earlier, we cannot control for the employment status of household members in the Gallup and the Psychological Corporation data. Appendix Figure C.1.11 shows that any bias is likely very small: in the ANES, *not* controlling for employment status increases the estimated union premium only slightly, relative to the baseline results where these controls are included.[35]

The family income premium may not fully capture changes in the household's economic well-being. Union families may benefit from other forms of compensation such as health benefits or vacation, as has been documented in the CPS-era (see 67 and 23 among others). Unfortunately, Gallup and our other sources do not consistently ask about benefits. One exception is from a 1949 Gallup survey that asked about paid vacation. As we show in Appendix Table C.1.7, Gallup respondents in union households are over twenty percentage points (about forty percent) more likely to report receiving paid vacation as a benefit.

On the other hand, the union premium may also reflect compensating differentials for workplace dis-amenities, which would suggest that our estimated premia are *overstating* the differential well-being of union households. Some evidence against this claim comes from another Gallup survey in 1939 that asks respondents how easily they could find a job "as

[35]Union households are more likely to have at least one person employed (likely the union member himself), which explains why controlling for household employment has a (slight) negative effect on the estimated union household premia. However, living with a union member is a *negative* predictor of own employment (results available upon request), which likely accounts for the fact that controlling for total number of workers in the household has only a small effect on the estimated premium.

good" as their current one. As we show in Appendix Table C.1.8, union households are far more likely to say it would be hard for them to find a job just as good. To the extent respondents considered non-wage job characteristics (safety, working conditions, benefits, etc.) this result is an additional piece of evidence that union members felt their jobs were better—in a broad sense—than non-union members.

Our estimates of a sizeable and relatively stable union premium contrast with recent papers using NLRB regression discontinuities in close representation elections to estimate the causal effect of unionization on firm-level outcomes [42, 63]. These papers have found little evidence of positive union premia. What explains the discrepancy? A possibility is that the LATE identified by the RD papers is not informative about the average treatment effect of unions. Importantly, most existing union workplaces were organized earlier and most elections are not very close. It is reasonable that a clear (sizeable) union victory in an election reflects workers' expectations of substantial advantage while a very close election reflects workers' expectations of more limited advantage. As such, the LATE identified by the RD papers is likely not informative (and likely understates) the average advantage of unionization. We do not mean to imply that we have identified the true average causal effect of unions on wages, but neither is it the case that the small effects found in the close-election RD analyses identify are appropriate when applied broadly.

3.6.3 Heterogeneous union household income effects

We have so far assumed that unions confer the same family income premium regardless of the characteristics of the respondent. We now explore heterogeneity by years of education and race (whites versus non-whites).

We begin by augmenting our family income equation (3.3) by adding an interaction term between years of schooling and the household union dummy. Figure 3.6 presents the coefficient on this interaction term, as usual, separately by survey-source and year. The results are

consistent throughout the period and show that less-educated households enjoyed a larger union family income premium. Over the nine decades of our sample period, this differential effect appears relatively stable. For each additional year of education, the household union premium declines by roughly four log points.

The analogous results when instead of adding $Years\,of\,educ_h^R \times Union_h$ to equation (3.3) we add $White_h^R \times Union_h$ are shown in Figure 3.7. The interactions are not statistically significant in the earliest surveys (the 1936 BLS Expenditure Survey and the 1942 Gallup Survey), though their signs suggest that white workers enjoyed larger premiums. However, in the 1946 Psychological Corporation survey and in succeeding Gallup, ANES and CPS surveys, there is consistent evidence of a larger union family income premium for nonwhites over the next five decades. This racial differential in the union effect on household income has declined somewhat since the 1990s and in the most recent CPS data it cannot be distinguished from zero.

Our conclusion from the heterogeneity analysis is that, at least for most of our sample period, disadvantaged households (i.e., those with respondents who are non-white or less educated) are those most benefited (in terms of family income) by having a household member in a union. Ignoring this differential effect would tend to underestimate the effect of unions on inequality, especially during the Great Compression period and beyond (1940-1990 or so) when the differential premium for black households appears largest.

3.6.4 Effects on residual income dispersion

An influential view of unions is that they lower the return paid not only to observed skill, as we document above, but also unobserved skill. Supporting this view is the fact that, at least in the CPS era, the union wage distribution is compressed even after conditioning on observable measures of human capital (e.g., 66 and 27).

We implement an analogous analysis at the household level to determine if unions performed a similar function in earlier decades. Separately for union and non-union households, we regress log family income on all the covariates (except union) in equation (3.3). As before, we perform this analysis separately by survey-source and year. We then calculate residuals for each sector and compute the ratio of variances between the union and non-union residuals (which has an F-distribution with degrees of freedom given by the two sample sizes, allowing us to construct confidence intervals). If unions compress the distribution of unobserved skill, then this ratio should be less than one.

Figure 3.8 shows, over our sample period, the ratio of variance of residual log family income between the union and non-union sector, together with 95% confidence intervals. The ratio is uniformly below one, and often below 0.5, with confidence intervals that always exclude equality of the variances. Like the union premium estimates, there does not seem to be a strong pattern over time in the union-nonunion difference in residual income inequality. Instead, it appears that the CPS-era pattern of unions compressing residual inequality in fact holds in a very similar manner throughout the post-1936 period.[36]

3.7 The Effect Of Unions On Inequality

We have so far documented that, in their effect on household income, unions have exhibited remarkable stability over the past eighty years. During our long sample period, the union premium has remained between ten and twenty log points, with the less-educated receiving an especially large premium. Moreover, the effects of unions on residual income variance are large and also relatively stable over time. By contrast, selection into unions is not constant across time. In the Great Compression period, when unions were at their peak and inequality at its nadir, *disadvantaged* households were much more likely to be union

[36]For example, [27] estimates a union-non-union variance ratio of around 0.61 in 1973 using individual male earnings, very similar to what we find in the 1970s for household income.

members than either before or since. These pieces of evidence suggest, at least indirectly, that unions were a powerful force pushing to lower income inequality during the heyday of the labor movement. In this section, we explore in a more direct manner the relationship between unions and income inequality.

First, and most directly related to the results in the previous two sections, we estimate distributional regressions following [61], where we show how measures of inequality change with small increases in union density. This regression jointly accounts for where union households are in the income distribution as well as the effect of union membership on a household's position in the income distribution. The identifying assumptions are, first, that, conditional on our controls, union membership is not otherwise correlated with determinants of income and, second, that union membership affects *only* the income of union households (i.e., no "spillovers" to other workers or households).

Second, we turn to more aggregate analysis. We follow some of the canonical work on the effect of skills shares on the college premium, and add union density to these standard, aggregate, time-series estimations. Note here that aggregate analysis does not rule out spillovers, but instead rests on the (strong) identifying assumption that conditional on our time-series controls, union density is exogenous. Finally, as the Gallup data always has state identifiers (thus allowing us to create, for the first time, state-year level density measures going back to the 1930s), we also conduct a parallel analysis at the state-year level.

The next three subsections describes the implementation of and results from each of these exercises. Note that each asks a slightly different question, so we would expect magnitudes to differ. We defer a more detailed comparison of magnitudes to Section 3.7.4.

3.7.1 Distributional regressions

In this section we present the effects of union membership on inequality using recentered influence functions (RIF) as in [61]. We provide a brief overview of the methodology below,

with greater detail provided in Appendix C.5. In our case, the RIF is a transformation of the dependent variable that allows the coefficient on the household union dummy to recover the effect of union density on some aggregate statistic of the family income distribution. We present effects of union density on the 90-10 ratio of log wages as well as on the Gini coefficient.

The virtue of RIF regressions is that they provide a tractable method to estimate the marginal effect of a variable on a statistic of the whole distribution. Let the joint distribution of family income y and covariates X be $F(y, X)$, and let the probability of a household being a union household be represented by $p \in [0, 1]$, so that $\mathbb{E}(union) = p$, where $union$ is our usual household union dummy. Then, under the assumption of no spillovers, we can decompose the family income distribution into the $union = 1$ and $union = 0$ distributions so that

$$F(y, X) = p \cdot F(y, X \mid union = 1) + (1 - p) \cdot F(y, X \mid union = 0). \tag{3.4}$$

Note that the assumption of no spillovers is incorporated into this expression by making the union and non-union joint distribution of income and characteristics independent of the share of the population that is unionized.

Consider some distributional statistic $\nu(F)$, such as the difference between the 90th and 10th percentile, $Q_{90} - Q_{10} = F^{-1}(\frac{90}{100}) - F^{-1}(\frac{10}{100})$ or the Gini coefficient $Gini(F)$. We can write the effect of a small change in union density on this statistic as $\frac{d\nu(F)}{dp}$. [61] show that the value of this derivative is obtained from a regression of the RIF of the statistic ν on the independent variable of interest. As with the selection and union premium results in the previous two sections, we run this estimation separately by survey source d and year y, using a specification parallel to equation (3.3) but with the RIF of a given inequality statistic ν instead of family income as the dependent variable:

$$RIF(y_{hst}, \nu(F)) = \beta Union_h + \gamma_1 Female_h^R + \gamma_2 Race_h^R + f(age_h^R) + g(Employed_h) + \lambda_h^{eduR} + \nu_t + \mu_s + e_{hst}. \quad (3.5)$$

In our setting, the RIF of a given statistic is, roughly speaking, the derivative of ν as the distribution in expression (3.4) is slightly perturbed toward the $union = 1$ distribution (i.e., as p increases slightly), an expression that must be derived for each statistic of interest.[37]

Figure 3.9 shows the effect of moving the family income distribution toward that of union families on the difference between the 90th and 10th percentile of log family income. Interestingly, while noisy, these estimates also display a U-shape over the 20th century. Given the constant income premium and U-shaped selection documented in the previous two sections, this pattern is unsurprising; when unions drew from the lower skilled, a constant wage premium would compress the wage distribution more. The implied effect on the 90-10 ratio varies from roughly -0.5 in the 1960s to -0.2 in earlier years as well as the CPS period.

In Appendix Figure C.1.12 we show the effects separately by the tenth, fiftieth and ninetieth percentile. Effects are uniformly positive for the 10th percentile, and negative or close to zero for the 90th percentile. The effect on the median is always positive, but depending on the data source and time period is sometimes above and sometimes below the effect on the tenth percentile. As we discuss in Appendix C.5, some of these differences may arise because of variation across data sources and across time in how coarsely the income data are binned, and we show robustness to changing bin sizes.

We conduct a parallel analysis for the Gini coefficient, which summarizes changes in inequality coming from all parts of the distribution, though, relative to the 90/10 ratio, it is especially sensitive to changes in the middle part of the distribution. Figure 3.10 shows the effect of union density on the Gini coefficient in each year. For the bulk of the sample

[37] For the sake of illustration, we derive the RIF of the 90/10 ratio in Appendix C.5.

period, the effect is quite stable, roughly -0.075, with magnitude falling as we approach the present, again, consistent with the up-skilling of union households since the 1960s.

3.7.2 Time-series regressions

While the distributional regressions capture the effect of union density on inequality, they require a strong assumption that there are no spillovers, threat effects, or political economy mechanisms that alter wages for non-union workers. The plausibility of these more macro mechanisms warrants an aggregate analysis, complementing the individual household regressions estimated above.

We begin our aggregate analysis of the effect of unions on inequality by adapting regressions from the time-series literature on the college wage premium, such as the analysis in [74], which spans the whole 20th century. Following [85] and [73] and using a mix of data from the Decennial Census, the CPS and a 1915 survey from Iowa, [74] show that the evolution of the college premium between 1915 and 2005 is well-explained by the relative supply of college workers, controlling for flexible functions of time. [12] confirm this analysis using data from the CPS in the 1963-2005 period and adding more covariates, and argue that the non-market factors stressed by [28], [92], and [96] have limited explanatory power in explaining the rise of inequality, measured as the 90-50 or 50-10 ratios. However, these papers do not consider unions as a potential factor in their analysis.

Motivated by these earlier papers, we begin by estimating the following equation:

$$\log\left(\frac{wage_t^{Col}}{wage_t^{HS}}\right) = \beta UnionDensity_t + \gamma log\left(\frac{N_t^{Col}}{N_t^{HS}}\right) + f(t) + \lambda X_t + \epsilon_t. \tag{3.6}$$

The dependent variable is the log college wage premium, which we specify as a function of the supply of skilled workers, $log(N_t^{Col}/N_t^{HS})$, a polynomial in time, $f(t)$, (which is meant in the SBTC literature to to pick up changes in relative demand for skill, although in our context it

could also pick up unmeasured institutional factors driving the college premium), other time-series controls X_t, which we vary to probe robustness, and, importantly, $UnionDensity_t$.[38]

We choose time-series controls X_t both to follow past literature as well as to capture the most obvious confounds in estimating the effect of unions on inequality. Specifically, following [12] we include the real value of the federal minimum wage and the civilian unemployment rate and following [112] we include the top marginal tax rate in the federal individual income tax schedule. We adjust for heteroskedasticity and AR(1) serial correlation in the error ϵ_t using Newey-West standard errors.

The first two columns of Table 3.2 show the results from this exercise. Col. (1) does not include additional controls X_t, whereas col. (2) does. While we will defer a more detailed discussion of magnitudes until Section 3.7.4, we note here that in both specifications, the coefficient on union density is negative and highly significant (and very similar to each other in magnitude). Appendix Table C.1.9 shows a variety of alternative specifications (e.g., using the Gallup series alone or the BLS series alone to calculate $UnionDensity_t$, instead of averaging the two together, substituting either a quartic or a quadratic for the cubic time polynomial) and also reports more of the coefficients, which we suppress in the main tables in the interest of space.

While the canonical analysis in [74] and related work focuses on the college premium, we extend our analysis in Table 3.2 by using the same specifications as in cols. (1) and (2) but using other measures of inequality as outcomes. Cols. (3) - (4) of Table 3.2 are identical to Cols. (1) - (2) except that the 90/10 log wage ratio for men (also taken from the IPUMS Census and CPS) is used as the outcome variable. The results are quite similar, with union density again having a negative and significant association with inequality that is robust to adding our vector of controls. Appendix Table C.1.10 is the analogue to C.1.9 and shows the result is robust across alternative specifications.

[38] As we do not have a strong view regarding whether, at the aggregate level, our Gallup-based estimate of early union density is better than the traditional BLS estimate, we take a simple average of the two.

Cols. (5) - (8) of Table 3.2 extend this analysis to inequality measures constructed from administrative data (rather than surveys). These have the advantage of being available annually, instead of just every ten years from the Census in the pre-CPS era.[39] These additional years not only give us more observations, but also allow us to use inter-Census variation (e.g., during World War II).

Cols. (5) and (6) use the Gini coefficient constructed by [90] from Social Security data. Without controls, the relationship is negative and highly significant, whereas adding controls decreases the magnitude by roughly one-fourth and significance is reduced to the ten-percent level. The final two columns uses the top-ten-percent income share from [111]. Without controls, the union coefficient is negative and highly significant. But this result is the most sensitive to controls, with the coefficient falling by half and losing its significant. The usual robustness checks for the Gini and the top-ten-percent-share outcomes are in Appendix Table C.1.11 and C.1.12, respectively.

While there are clear limitations to the time-series analysis (e.g. no exogenous variation in union density, and small samples make inference suspect), all specifications in this section control for a cubic in time (and we show results using a quadratic or a quartic in the Appendix Tables), ensuring that the robust negative relationships we find between unions and our measures of inequality are not simply mirroring the U-shape present in many time-series over the 20th century.

[39] A small complication in using these annual outcomes is that our estimates of the skill shares $log(N_t^{Col}/N_t^{HS})$ in equation (3.6) come from survey data, and thus in principle are only available every ten years in the pre-CPS era. To circumvent this issue, we include two separate education controls: (i) skill shares as measured (annually) in our Gallup data and an annual measure of skill shares equal to that from the CPS when it is available; and (ii) interpolating between Census years in the earlier period. In this sense, we treat education as a nuisance variable and simply try to control flexibly for it, allowing us to continue to estimate the conditional effect of union density.

3.7.3 State-year panel regressions

While the time-series analysis generates summary accounts of the aggregate association of unions on the U.S. economy, a major limitation is that there are many unobserved factors (e.g. macroeconomic policy, trade, outsourcing, industry structure) that are likely correlated with both inequality and union density that are not absorbed by our controls. In this section we replicate the analysis at the state-year level, controlling for state and year fixed effects, which can absorb a considerable amount of unobserved heterogeneity.

The Gallup data always contain state identifiers, so we can construct continuous state-year measures of union density throughout the pre-CPS period, something that was not possible with previous data.[40] One limitation of our survey-based data is that small states get small samples, resulting in noisy estimates for annual variation. We use both winsorized measures as well as a split-sample instrumental variables strategy to mitigate this problem. Although we do not have exogenous variation in union density, we can see if the inverse inequality-density relationship that holds in the aggregate time series hold at the state-year level, conditional on year and state fixed effects.[41] Importantly, as all states have access to the same national technology, the year fixed effects in this design controls for simple variants of SBTC that affect all states the same way.

We combine our Gallup state-year measures with household state-year measures calculated from the CPS. We take a weighted average of Gallup-generated state-year union densities and CPS-generated state-year densities, with weights proportional to the number of observations in each sample (so the CPS gets a much larger weight). This procedure results in a panel of state-year union density measures going back to 1937. Note that such a panel was impossible to construct before the Gallup data, as in most years the BLS and Troy series did not break down their aggregate counts geographically.

[40] [122] presents state breakdowns for 1939 and 1956, and [80] use BLS reports to construct state-year measures of density from 1964 onwards.

[41] Similar regressions are estimated at the cross-country level by [84], though their sample period of 1980-2010 is far shorter than ours.

To examine the effect of unions on inequality, we closely follow equation 3.6 and estimate specifications of the form:

$$y_{st} = \beta UnionDensity_{st} + \gamma log\Big(\frac{N_{st}^{Col}}{N_{st}^{HS}}\Big) + \lambda X_{st} + \mu_{t\times South} + \delta_s + \epsilon_{st} \tag{3.7}$$

where y_{st} is a measure of inequality, for example the college-HS wage gap or the percent of total income accruing to the top ten percent, in state s and year t. As before, we control for skill-shares $\log(N_{st}^{Col}/N_{st}^{HS})$ in all specifications.[42] We include South-by-year and state fixed effects in all regressions, $\mu_{t\times South}$ and δ_s, respectively. Note that we include South-by-year fixed effects because, as we showed earlier, Gallup's sampling of the South improves over time. We cluster the standard errors at the state level.

As before, we will show results with and without X_{st}, a vector of additional state-year controls. We try our best to capture the same covariates as in equation (3.6), though in some cases controls that are available at the annual level in the historical period are not available at the state-year level. To control for economic expansions and contractions, we include in X_{st} state-year log GDP per capita and state-year measures of the share of households subject to the federal income tax, as annual state-level unemployment rates are not consistently available until the 1963 CPS. We do not have information on state tax schedules going back to the 1930s, so to capture the political-economy climate, we control for a state-year level "policy liberalism" index developed by [31]. Manufacturing moving from the unionized Northeast and Midwest to the South and West is often cited as a reason for the decline in density, so we include in X_{st} the one-digit industry employment shares at the state-year level. Finally, to deal with possible unobserved but smooth state-specific changes in technology or other unobservables that may be confounding the estimated relationship, X_{st} also includes state-specific linear and quadratic trends.

[42]The top-ten-percent share of income is available at the *annual* level, so just as in the time-series regressions we include both the interpolated IPUMS-CPS education measure (at the state-year level) as well as the Gallup measure of education for that outcome (at the state-year level).

As mentioned above, because our Gallup sample size will become small for less populous states, our coefficients may be attenuated due to finite-sample bias in our state-year level union density measures. To address this concern, we use a "split-sample" IV strategy.[43] For each state-year, we split the Gallup observations into two random samples s_0 and s_1, and use the union density calculated from s_1 to instrument the union density calculated from s_0. This procedures yields the following first-stage equation:

$$UnionDensity^0_{st} = \eta UnionDensity^1_{st} - \iota \log\left(\frac{N^{Col}_{st}}{N^{HS}_{st}}\right) + \lambda^f X_{st} + \mu_{t\times South} + \delta_s + \nu_{st}. \quad (3.8)$$

The second-stage equation in the split-sample IV is merely equation (3.7) with $UnionDensity_{st}$ replaced by $\widehat{UnionDensity}^0_{st}$, the prediction generated from the first stage. Since $UnionDensity^1$ and $UnionDensity^0$ are calculated from a *random* split of the data, the sampling errors in the two measures will be orthogonal. Omitted variable issues aside, if the only issue is measurement error, the IV estimator β^{IV} will yield a consistent, unattenuated estimate of β.

Table 3.3 shows results from the specification in equation (3.7) across the state-year analogues of the four main inequality outcomes used in Table 3.2.[44] As in the previous subsection, the odd-numbered columns do not include the additional controls X_{st}, while the even-numbered columns do. Cols. (1) and (2) show results when the college-premium is the outcome variable. The coefficient on state-year union density is negative and significant, and the magnitude barely changes whether or not additional controls are included. Indeed,

[43] See [9] for an early description of the methodology, and [83] and [14] for more recent exposition and applications, respectively.

[44] For cols. (1) to (4), we simply calculate the skill premium and the 90/10 ratio at the state-year level instead of the year level from the same survey data sources. For cols. (5) and (6), as the Social Security data are not disaggegrated to the state level, we instead just calculate the household-income Gini from our survey data sources. For cols. (7) and (8), we use recently constructed top income shares at the state-year level from [65], based on internal IRS data (downloaded from the World Wealth and Income Database 8). These data are not adjusted for capital gains. [65] show that when the same methodology is applied to national US data the results are still quite close to [111].

across our three other outcomes (the male 90/10 ratio, the Gini coefficient, and the top-ten-percent shares), the coefficient on state-year density is negative and quite robust to adding additional controls.

In Appendix Tables C.1.13 to C.1.17 we show a variety of specifications that add intermediate sets of controls between the odd and even columns reported in Table 3.3.[45] These tables also contain a set of estimates (column 1) that do not use the split-sample IV for state-year union density. These estimates verify the presence of attenuation bias, with the split-sample IV coefficients roughly fifty percent larger than the OLS coefficients.

A natural concern is that unions' compression of state-level income distributions comes at the cost of slowed economic growth (e.g., lowered net business entry). In fact, union density shows consistently positive, but sometimes insignificant, effects on state GDP per capita, and we can rule out even small negative effects of unions on state-level economic activity (see Appendix Table C.1.17).

3.7.4 Discussion of Estimate Magnitudes

While each of our three approaches involves strong identification assumptions, all regression results point to negative effects of unions on inequality. How large are these effects and why might they differ across approaches?

In Table 3.4 we show results from a simple exercise: taking our various coefficients, how much of the *decline* in inequality from 1940-1960 can be explained by the eleven-percentage-point rise in union density, and how much of the *increase* in inequality between 1970 and 2004 can be explained by the twelve-percentage-point decline in density?[46] We begin with

[45] An additional robustness check we perform but do not report is including state-year variation in the passage of right-to-work laws. Including such controls has no effect on our results, and we are persuing the independent effect of right-to-work laws on union composition in ongoing work. See [58] for recent work on right-to-work laws, mainly using CPS-era data.

[46] We end the recent period in 2004 as it is the last year for which the Social Security earnings Gini coefficient measures is available. While somewhat arbitrary, we choose 1940 and 1970 as starting points as they are Census years and thus we have college premium and 90/10 outcomes for those years.

coefficients from the RIF regressions. As we noted, the RIF regression coefficients are larger in magnitude during unions' heyday relative to the more recent period, because unions were organizing less advantaged households during their Great Compression peak years. We thus consider a "high" and "low" RIF effect, the "high" using the simple average of all the RIF coefficients between 1950 and 1970 (i.e., those plotted in Figure 3.9 and 3.10 for the 90/10 and the Gini, respectively) and the "low" using the simple average of coefficients from 2005-2015. Multiplying the "high" coefficients by the rise (decline) in union density between 1940-1960 (1970-2004) can explain a sizeable share of the decline (rise) in the 90/10 ratio and Gini coefficients, roughly between ten and fifteen percent, comparable with the results in [40] and [27]. By contrast, using coefficient estimates from the recent period would suggest that changes in union density explain less than ten (and in some cases less than five) percent of the changes. An important point that the RIF analysis highlights is that the question "how would inequality change today if union density were to somehow increase by x percent" is not, strictly speaking, well-defined: one would want to know from where in the counterfactual non-union income distribution this hypothetical increase in density would come.

Notwithstanding this point, the rest of the table will assume a constant effect of union density across time. For the aggregate time-series and state-year regressions, we always use point-estimates that include the additional covariates (i.e., the even-numbered columns of Tables 3.2 and 3.3). The implied explanatory power of union density is considerable using the time-series coefficients: we in fact substantially "over-explain" the decline in the skill premium from 1940-1960 (potentially consistent with the observation in 74 that skill shares *under*-explain the fall between 1940 and 1950). Roughly speaking, explanatory power when we use the state-year coefficients is about one-fifth to one-half as large as that from the aggregate time-series. Like the RIF results, results from the time-series and state-year regressions imply that changes in union density have similar power to explain *both* the

decline and then the rise of our inequality measures: in general, the magnitudes of the share-explained estimates are similar in the top and bottom panels of the table.

Why does the implied explanatory power of union density vary across our three econometric approaches? As we have emphasized throughout, each of these approaches requires strong identifying assumptions, and to us the most obvious explanation for why they differ is that omitted-variable and other forms of endogeneity affect some of these estimates more than others (in a manner that is impossible to measure or even sign). But even if each of these exercises were somehow perfectly identified, we would *not* expect them to give the same coefficient estimates. For example, the smaller effects yielded by the RIF compared to the time-series could be due to the fact that only the latter picks up spillover effects such as fairness norms or political economy mechanisms. That the state-year estimates are smaller than the national estimates could suggest that these spillover effects are more powerful at the national level or that some countervailing spillover effects (e.g., creating excess labor supply in uncovered sectors) operate at the more local level. Similarly, note that inequality effects at lower levels of geography do not generally "aggregate up" to the national effect, so we should not expect the results from the aggregate time-series and the state-year panel regressions to be similar even absent spillovers. For example, unions' reducing the top-ten share in the richest states will have a much larger effect on the *national* top-ten share than their reducing the top-pen share in a poor state like Mississippi (whose ninetieth percentile during much of our sample period is not much higher than the national median).

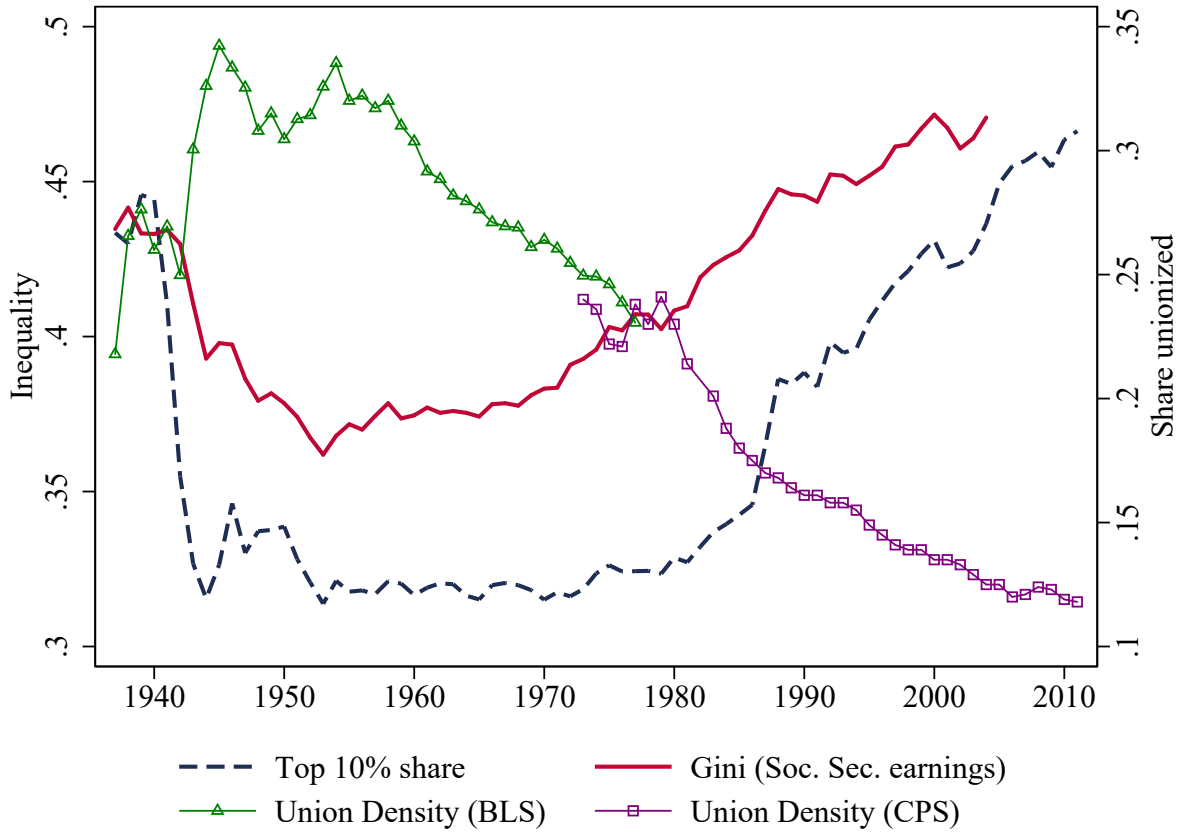
3.8 Conclusion

We leverage historical polling data, allowing us to provide a systematic, representative study of unions' effects on the income distribution over a much longer period than existing work. We show that a combination of low-skill composition, compression, and a large union income premium made mid-century unions a powerful force for equalizing the income distribution.

As unions have receded, it is perhaps surprising—and counter to the predictions of some SBTC models of endogenous unionization—that relatively *skilled* workers are the ones that remain. This pattern mimics the pre-World War II era, when unions were both small and their members relatively skilled. Our results show that over the last nine decades, when unions expand, whether at the national level or the state level, they tend to draw in unskilled workers and raise their relative wages, with significant impacts on inequality.

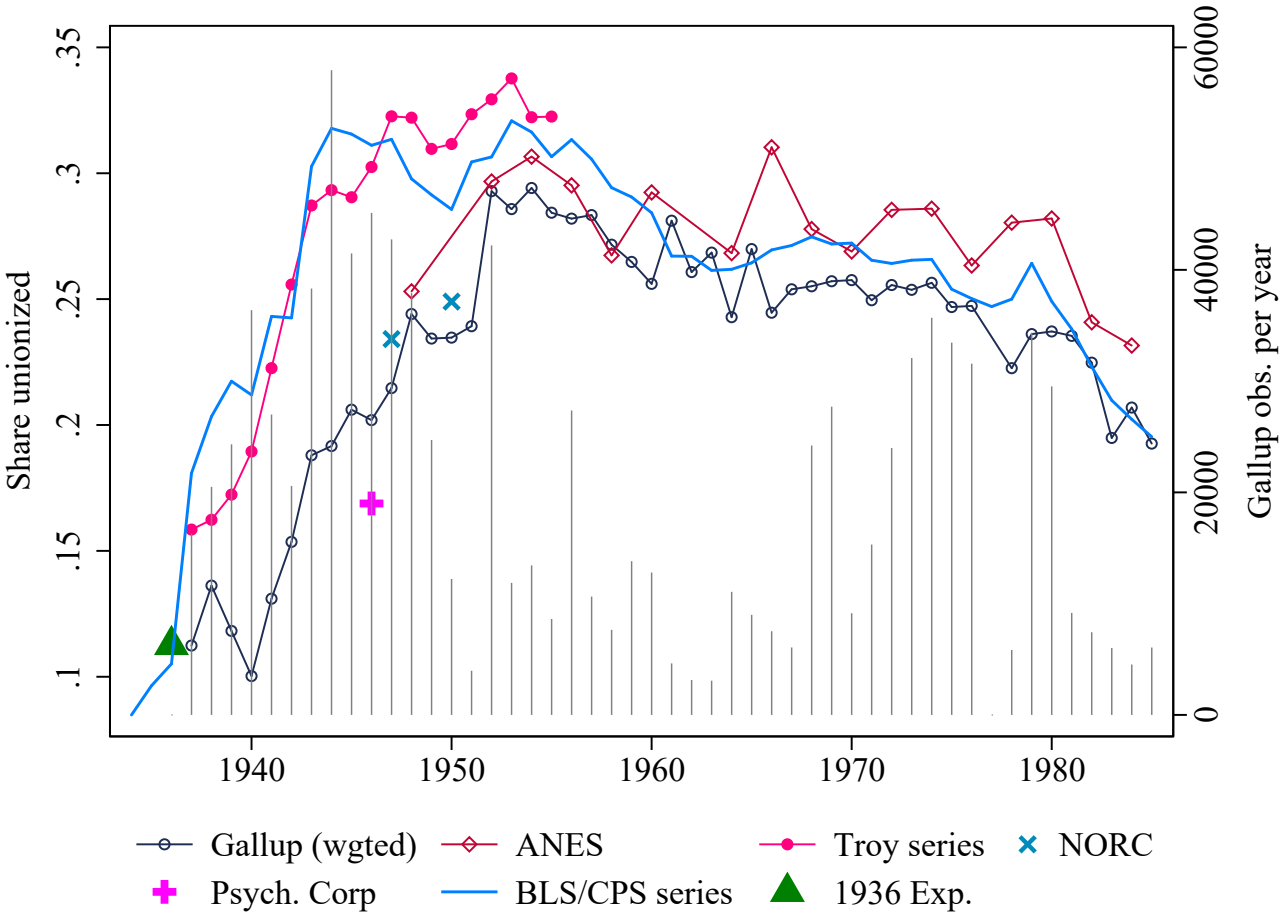
We welcome future work that develops theoretical models explaining the joint evolution of union density, skill composition, premia, and overall inequality we have documented. More work on the effect of unions, perhaps in light of the recent literature documenting pervasive labor market power, would inform whether unions could be an important part of a feasible policy package to lower inequality.

Figure 3.1: Union density and inequality measures, 1917-2011



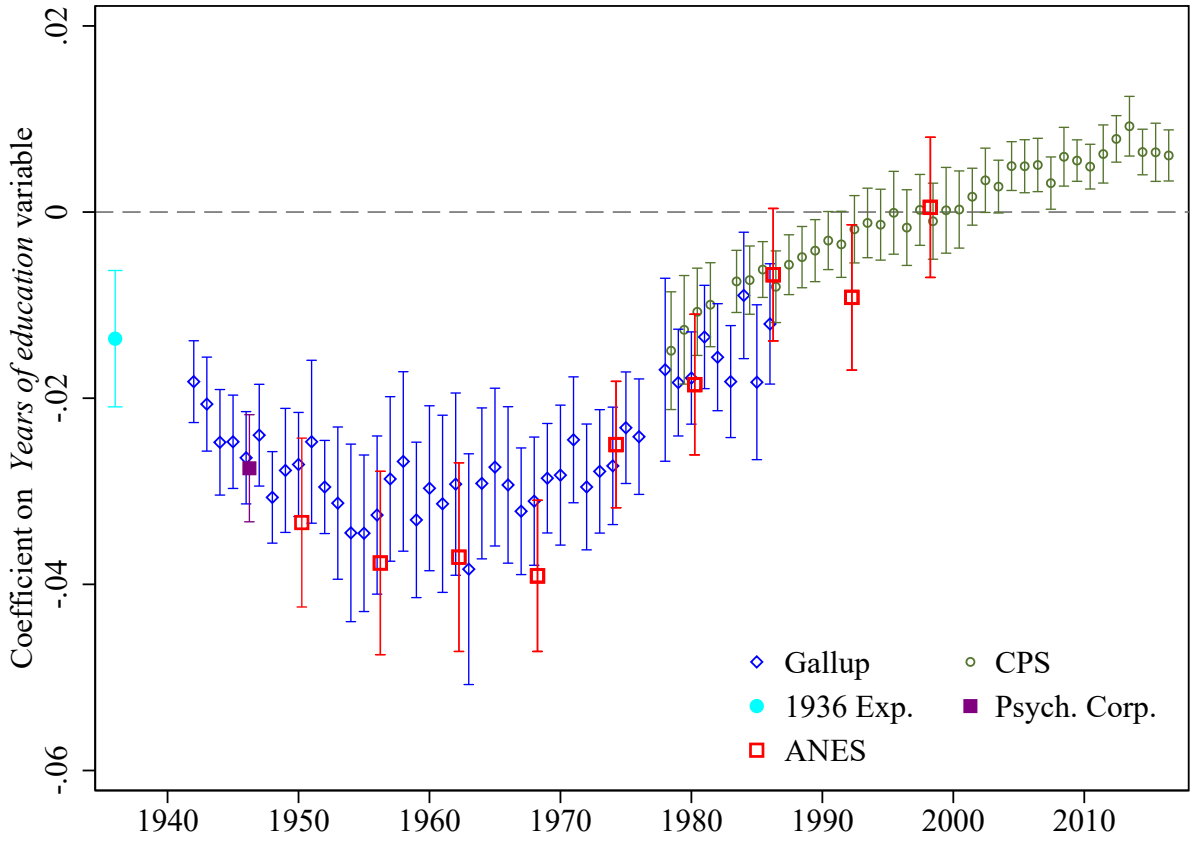
Data sources: Top share inequality from Piketty and Saez (2003, updated 2016). Union density data from Historical Statistics of the United States and the Current Population Survey. We discuss these data sources in detail in Section 3.3 and Appendix C.4.

Figure 3.2: Union density over the twentieth century: Comparing our survey-based measures to existing time-series



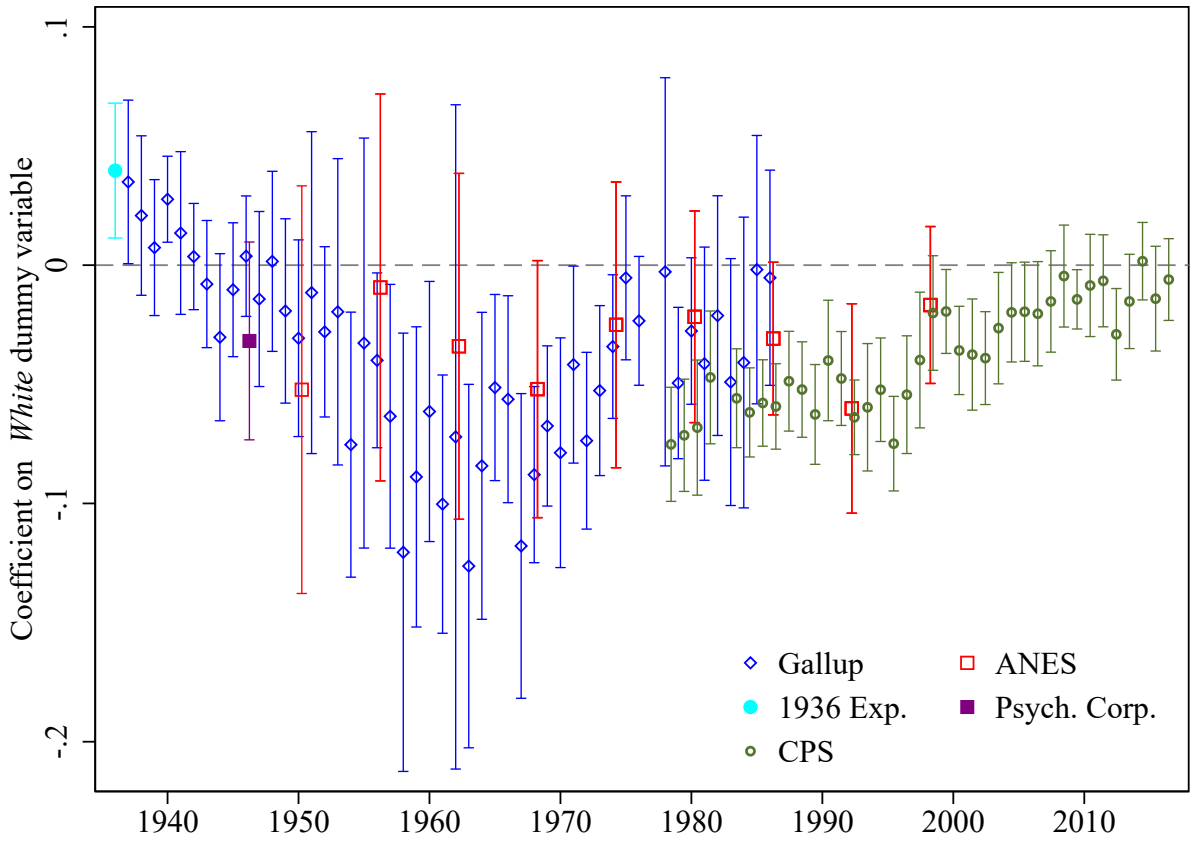
Data sources: See Sections 3.3 and Appendix C.4.
Notes: No sample restrictions are imposed (so farmers and those over age 65 *are* included in this graph). The vertical spikes indicate the number of Gallup observations per year that include the union variable (plotted on the right-hand-side axis).

Figure 3.3: How does educational attainment predict union household status?



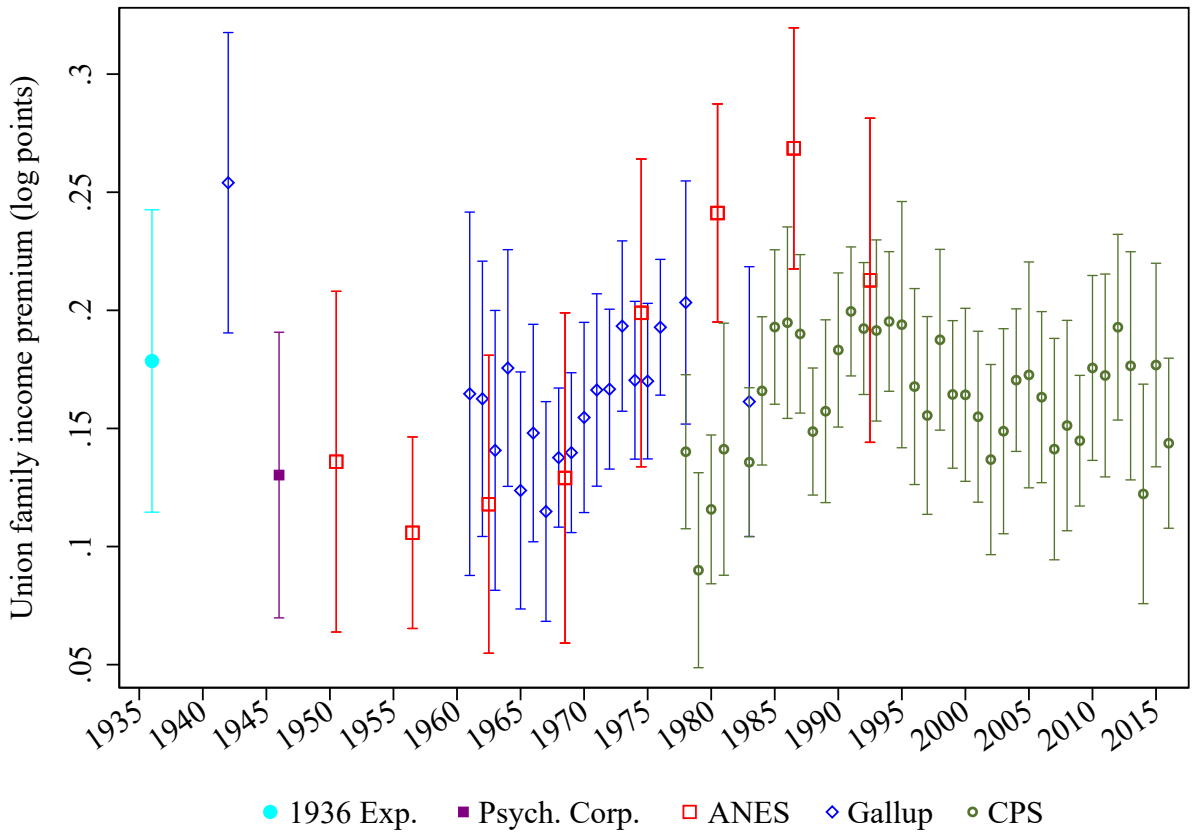
Data sources: Gallup data, 1937–1986; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source.
Notes: We regress household union status on $Years\,of\,education$, state s and survey-date t fixed effects, age and its square, and gender. (The notes to Figure 3.6 describe how we impute years of education if the survey source only gives us categories of educational attainment.) We estimate this equation separately by survey source and by year. The figure plots the coefficient on $Year\,of\,education$. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are based on standard errors clustered by state.

Figure 3.4: How does race predict union household status?



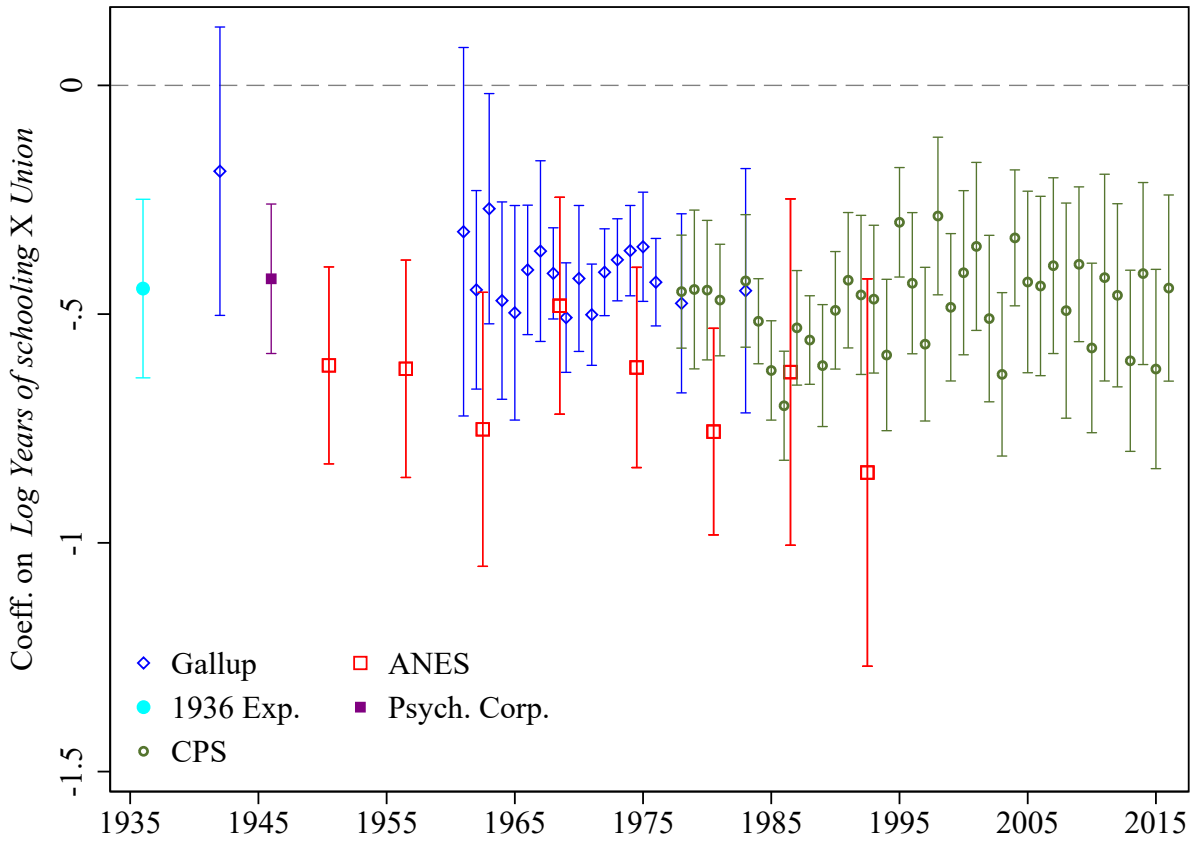
Data sources: Gallup data, 1937–1986; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source.
Notes: For each data source, we estimate, separately by data source and year, household union status on a *White* dummy variable, state s and survey-date t fixed effects, age and its square, and gender. We plot in this graph the coefficients on $White$ from each of these estimations. For the ANES, because the samples are smaller, we group surveys into six-year bins. Confidence intervals are based on standard errors clustered by state.

Figure 3.5: Estimates of the union family income premium



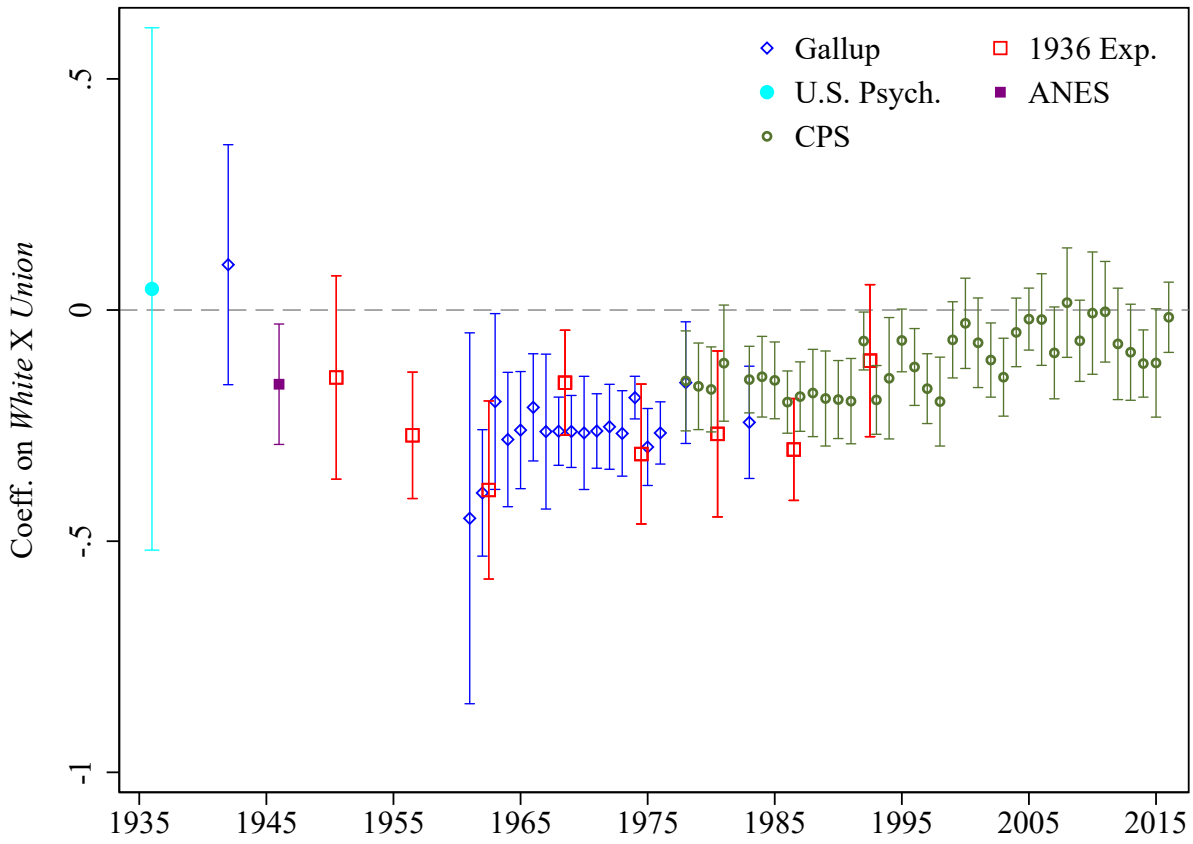
Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source. See Appendix C.2 for details on CPS family income variable construction.
Notes: Each plotted point comes from estimating equation (3.3), which regresses log family income on controls for age, gender, race, state and survey-date fixed effects. Occupation controls are not included. We estimate a separate regression for each survey source and year. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are based on standard errors clustered by state.

Figure 3.6: Differential family union premium by respondent's years of schooling



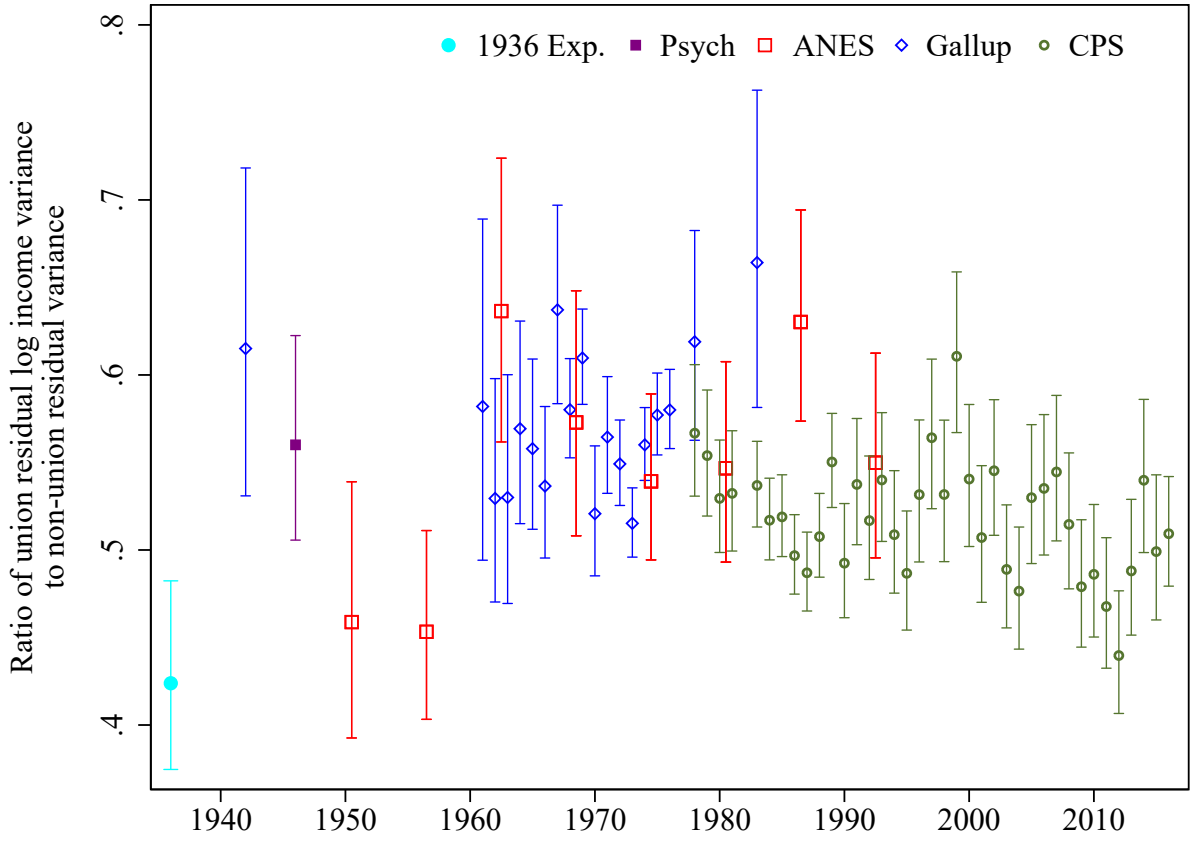
Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source. See Appendix C.2 for details on CPS family income variable construction.
Notes: Each plotted point comes from estimating equation regressing log family income on household union status, its interaction with respondents' log years of schooling, and all other controls in equation (3.3). We estimate this equation separately by survey source and by year. Some survey sources give actual years of schooling. For those that do not, we impute in the following manner: six years for "less than middle school;" eight years for "middle school;" ten years for "some high school;" twelve years for "high school;" fourteen years for "some college" or "vocational training;" sixteen years for "college," eighteen years for "more than college." The figure plots the coefficient on the interaction $Years\,of\,schooling \times Union$. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are based on standard errors clustered by state.

Figure 3.7: Differential family union premium for whites relative to minorities



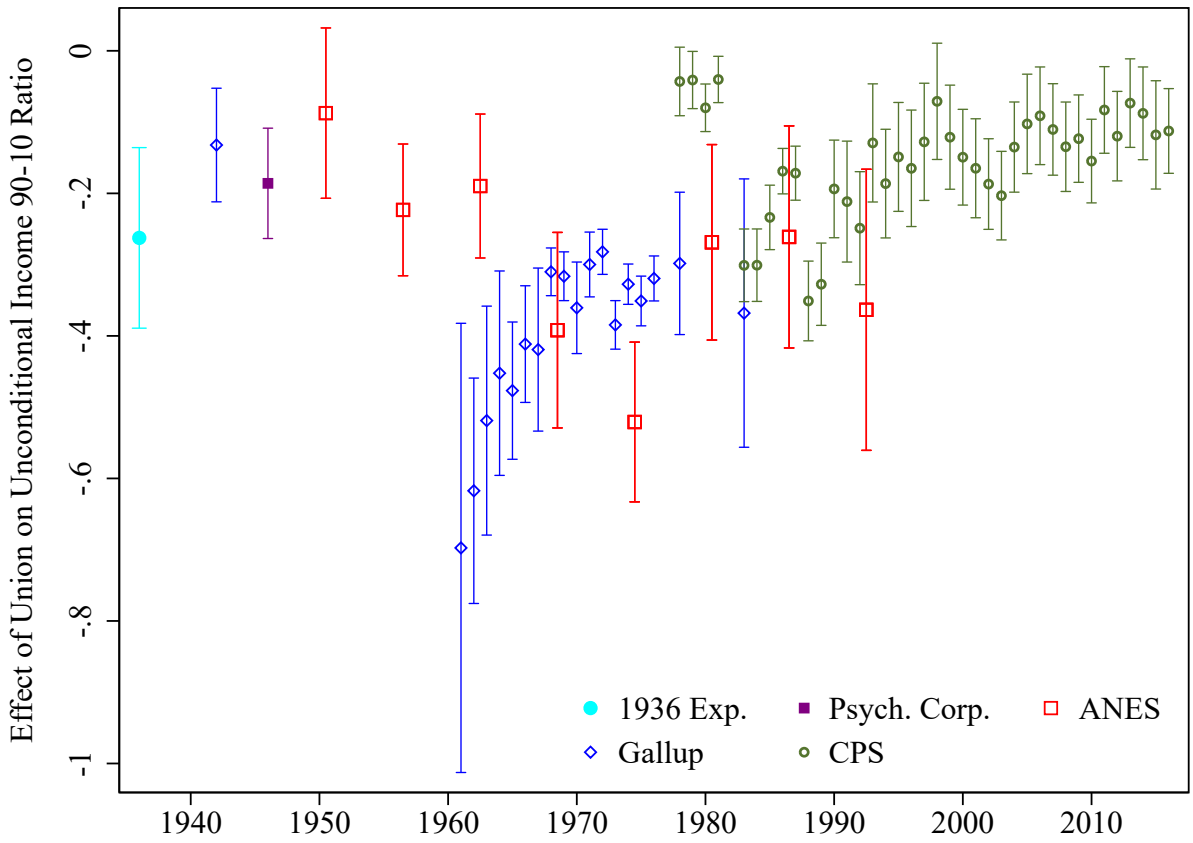
Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source. See Appendix C.2 for details on CPS family income variable construction.
Notes: Each plotted point comes from estimating equation regressing log family income on household union status, its interaction with a White dummy variable, and all other controls in equation (3.3). We estimate this equation separately by survey source and by year. The figure plots the coefficient on the interaction $White \times Union$. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are based on standard errors clustered by state.

Figure 3.8: Ratio of residual variance between union and non-union sectors



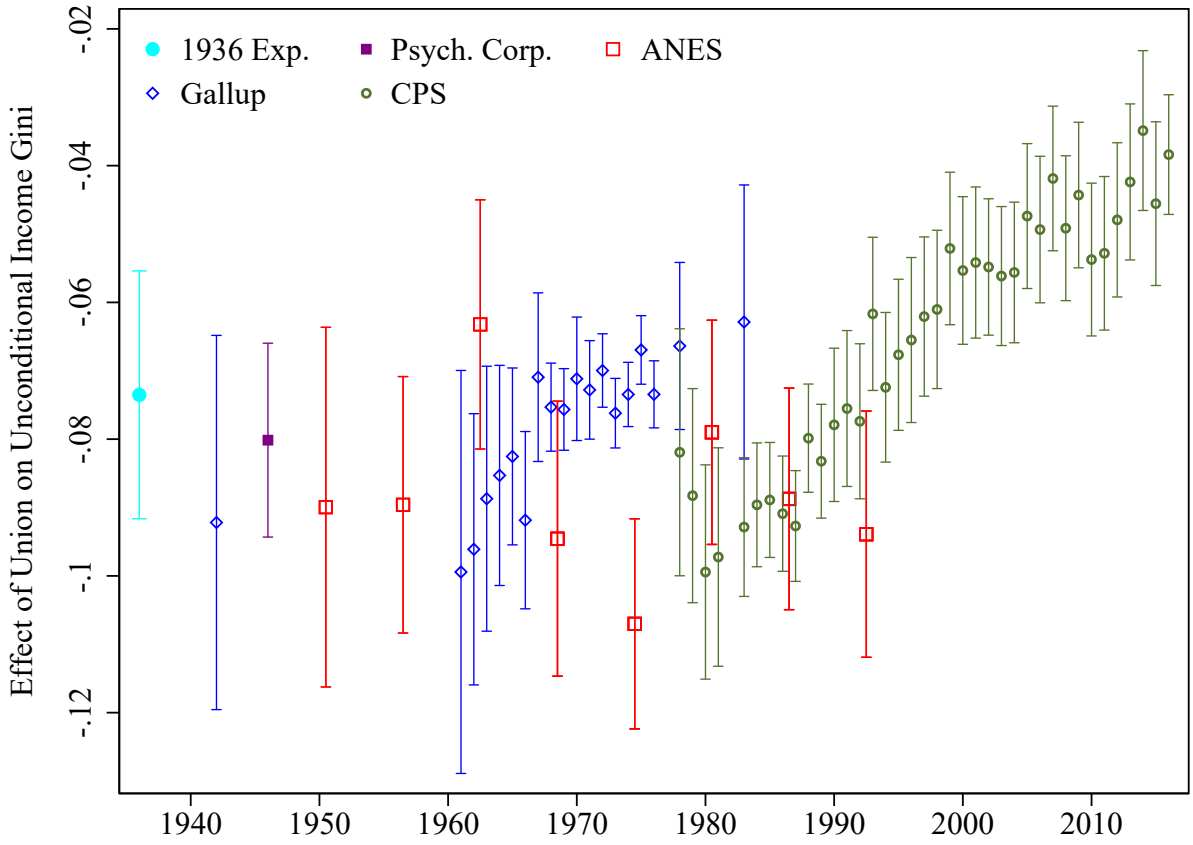
Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source. See Appendix C.2 for details on CPS family income variable construction.
Notes: Each plotted point is the ratio of variance of residuals from regressing log family income on the controls in equation (3.3) separately for union and non-union households, separately by survey source and by year. See Section 3.6.4 for more detail. The figure plots the ratio of the variance of residuals in the union sector to that of the non-union sector (so ratios less than one suggest that residual variance in the union sector is more compressed than in the non-union). The plotted confidence intervals are based on inverting the F-statistic testing the null that the ratio is equal to 1. For the ANES, because the samples are smaller, we group surveys into six-year bins.

Figure 3.9: Effects of union density on the 90/10 ratio of family income



Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section 3.3 for a description of each data source. See Appendix C.2 for details on CPS family income variable construction.
Notes: Each plotted point comes from estimating equation (3.5), which regresses the recentered influence function (RIF) for the difference between the 90th and 10th quantile on controls for age, gender, race, educational attainment fixed effects, household employment status controls, state and survey-date fixed effects. Occupation controls are not included. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are robust to heteroskedasticity.

Figure 3.10: Effects of union density on the Gini coefficient of family income



Data source: See Section 3.3 for a description of each data source. See Appendix C.2 for details on CPS family income variable construction.
Notes: Each plotted point comes from estimating equation (3.5), with regressed the recentered influence function (RIF) for the Gini coefficient on controls for age, gender, race, educational attainment fixed effects, household employment status controls, state and survey-date fixed effects. Occupation controls are not included. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are robust to heteroskedasticity.

Table 3.1: Is selection into unions a function of union density?

	Dep't var.: At least one household member is in a union				
	(1)	(2)	(3)	(4)	(5)
Yrs. of educ x Aggregate union density	-0.280*** [0.0259]	-0.124*** [0.0235]	-0.104*** [0.0231]	-0.122*** [0.0234]	-0.102*** [0.0227]
Yrs. of educ x Year		-0.449*** [0.0272]	-0.266*** [0.0281]	-0.451*** [0.0280]	-0.266*** [0.0292]
Yrs. of educ x Year sq.		0.230*** [0.0136]	0.135*** [0.0138]	0.231*** [0.0140]	0.135*** [0.0143]
White x Aggregate union density				-0.183* [0.0956]	-0.0923 [0.113]
White x Year				0.0842 [0.274]	-0.0303 [0.276]
White x Year sq.				-0.0381 [0.135]	0.0165 [0.136]
Dep't var. mean	0.203	0.203	0.178	0.203	0.178
Year FE?	Yes	Yes	Yes	Yes	Yes
Drop pub. sect. HH?	No	No	Yes	No	Yes
Observations	1143694	1143694	1005363	1143694	1005363

Sources: Gallup, 1937–1976, CPS, 1977-2015.

Notes: These regressions test whether selection into union households by education and by race is a function of union density. All regressions include state and year fixed effects, and controls for age (and its square), race, gender and education. Variables that are interacted with $Year$ ($Year^2$) are divided by 1,000 (1,000,000) for ease of presentation. Note that $Year$ refers to calendar year (e.g., 1942, 1943, etc). Thus, interacting a variable with $Year$ and $Year\ sq.$ allows selection into unions with respect to that variable to follow an arbitrary U-shape in calendar time. Standard errors are clustered by state.

$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.2: Aggregate inequality as a function of union density

	Dependent variable:							
	Coll. premium		90/10 ratio		Gini coeff.		Top 10 share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Union Density	-1.338**	-1.556**	-1.407***	-1.456**	-0.132**	-0.106**	-33.64**	-15.57
	[0.535]	[0.720]	[0.421]	[0.636]	[0.0536]	[0.0525]	[14.41]	[12.50]
Mean, dept. var	0.512	0.512	1.376	1.376	0.410	0.410	35.848	35.848
Gallup edu. control?	No	No	No	No	Yes	Yes	Yes	Yes
Addit. controls?	No	Yes	No	Yes	No	Yes	No	Yes
Cubic polynomial?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49	49	49	49	65	65	70	70

Sources: For cols. (1) - (4), outcome variables generated from Census IPUMS and CPS; for cols (5) and (6) from [90] and for cols (7) and (8) from Piketty and Saez (2003, updated 2016). The union density explanatory variable is the simple average between the Gallup- and BLS-based density measures (see Section C.3 for detail).

Notes: All regressions include controls for the log share of college versus high-school educated workers, calculated in the early years from Census IPUMS and for later years from the CPS. The first four columns use outcome variables calculated from the source (so are only available in Census years until the CPS), but the last four columns use as outcomes *annual* measures, calculated from administrative data. For these measures, we have to control *annually* for skill shares. We include two annual controls: annual skills shares as measured in Gallup and annual skills shares as measured in the Census IPUMS and the CPS (interpolated between Census years in the pre-CPS years). For each outcome variable, the first specification has parsimonious controls (only a time cubic and the skill shares controls) and the second has additional controls (federal minimum wage, the national unemployment rate, and the top marginal tax rate in the federal income tax schedule). Appendix Tables C.1.9, C.1.10, C.1.11, C.1.12 provide additional specifications using the skill premium, the log 90/10 ratio, the Gini coefficient and the top-ten share, respectively, as outcomes. Note that to the log 90/10 is for men only, but all other inequality measures pool both men and women. Standard errors are robust to heteroskedasticity and AR(1) serial correlation. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.3: State-year inequality as a function of union density

	Dependent variable:							
	Coll. premium		90/10 ratio		Gini coeff.		Top 10 share	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Household union share	-0.442*** [0.118]	-0.434*** [0.104]	-0.296** [0.118]	-0.244** [0.0953]	-0.0635*** [0.0235]	-0.0738*** [0.0235]	-5.342*** [2.065]	-3.147** [1.399]
Mean, dept. var.	0.490	0.497	1.386	1.398	0.376	0.378	36.61	36.96
Industry shares	No	Yes	No	Yes	No	Yes	No	Yes
State-spec. quad.	No	Yes	No	Yes	No	Yes	No	Yes
Income covars.	No	Yes	No	Yes	No	Yes	No	Yes
Policy covars.	No	Yes	No	Yes	No	Yes	No	Yes
Observations	1640	1505	1640	1505	1640	1505	3107	2723

Sources: For cols. (1) through (6), dependent variables created using Census and CPS data; for (7) and (8) they are taken from [65]. See Appendix for variable construction. The key explanatory variable comes from state-year average household union share generated from Gallup in the earlier years and the CPS in later years.

Notes: All estimates are from split-sample-IV regressions (see Section 3.7.3 for estimating equations). All regressions include state and year fixed effects; $South \times Year$ fixed effects; and state-year education controls (both from Gallup and CPS at the annual level, and interpolated from the IPUMS Census at the decade level). "Industry shares" controls for state-year share of employment in all one-digit industry categories. "State-spec. quad." indicates that state-specific quadratic time trends are included. "Income covars." indicate that state-year GDP and state-year share of households filing taxes are included. "Policy covars." indicate that state-year minimum wage and a "policy liberalism" index (from 31) are included. Sample size is larger for the top 10 outcome because it is available at the annual level in all years; for the other outcomes, until the CPS in the 1970s, we only have data from the decadal Census.

$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Table 3.4: How much of the variation in inequality is explained by changes in union density?

		Share explained by Δ union density, using β values from...			
	Total Δ	RIF (high)	RIF (low)	Time-series	State panel
—1940-1960					
Union density	0.113				
Skill premium	-0.0926			1.894	0.449
Male 90/10 ratio	-0.326	0.135	0.0379	0.504	0.200
Gini coefficient	-0.0585	0.167	0.0890	0.204	0.116
Top ten percent	-12.77			0.137	0.0188
—1970-2004					
Union density	-0.123				
Skill premium	0.212			0.904	0.214
Male 90/10 ratio	0.465	0.103	0.0290	0.385	0.153
Gini coefficient	0.0875	0.122	0.0649	0.149	0.0843
Top ten percent	12.13			0.158	0.0216

Sources: For cols. (1) through (6), dependent variables created using Census and CPS data; for (5) and (6) they are taken from [65]. See Appendix for variable construction. The key explanatory variable comes from state-year average household union share generated from Gallup in the earlier years and the CPS in later years.

Notes: Col. (1) is the change in each outcome variable j (skill premium, male 90/10 ratio, gini coefficient, and top 10 percent) as well as union density between 1940 and 1960 (1970 and 2004 in lower panel). Values in cols. (2)-(5) are calculated, for column i, row j, as $^{(\Delta_{unionDensity} * \beta_{ij})}/_{\Delta_j}$ where β_{2j} is the average of all RIF coefficients between 1950 and 1970 for outcome j (Figures 3.9 and 3.10), β_{3j} is the average of all RIF coefficients between 2005 and 2015 (Figures 3.9 and 3.10), β_{4j} is the point-estimate in the even-numbered column of Table 3.2 corresponding to outcome j, and β_{5j} is the point-estimate in the even-numbered column of Table 3.3 corresponding to outcome j. Note that 1950-1970 is roughly the period when the RIF coefficients are largest in magnitude (which we term "RIF (high)"), while in the more recent years they are smaller in magnitude ("RIF (low)").

Appendix A

Chapter 1 Appendix

A.1 Additional Tables and Figures

Table A1: Difference-in-Differences Estimates of the Effect of IDR on Loan Repayment Outcomes: Predicted Non-Zero Payments

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean $t = -1$	Mos. 4 to 7	Mos. 8 to 11	Mos. 12 to 15	Mos. 4 to 7	Mos. 8 to 11	Mos. 12 to 15
Minimum Payment	0.55	-0.25^{***}	-0.25^{***}	-0.11^{***}	-0.27^{***}	-0.27^{***}	-0.13^{***}
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Remaining Balance	60.55	-0.13	-0.84^{***}	-0.79^{***}	-0.29^{**}	-1.00^{***}	-0.95^{***}
		(0.17)	(0.18)	(0.21)	(0.15)	(0.16)	(0.20)
Δ Remaining Balance	0.26	-0.07	-0.11^{**}	0.16^{**}	-0.12^{***}	-0.16^{***}	0.11^{**}
		(0.06)	(0.06)	(0.07)	(0.04)	(0.02)	(0.04)
10+ Days Delinquent	0.21	-0.17^{***}	-0.12^{***}	-0.03^{**}	-0.18^{***}	-0.13^{***}	-0.05^{***}
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
90+ Days Delinquent	0.01	-0.05^{***}	-0.04^{***}	-0.03^{***}	-0.06^{***}	-0.05^{***}	-0.05^{***}
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on monthly loan repayment outcomes for those predicted to have non-zero IDR payments. Column 1 reports the dependent variable mean in the month prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive three-month periods, beginning with four months following the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same monthly effect for a regression which omits pre-call month dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 3,338 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A2: Difference-in-Differences Estimates of the Effect of IDR on Loan Repayment Outcomes: High Balance Borrowers

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean $t = -1$	Mos. 4 to 7	Mos. 8 to 11	Mos. 12 to 15	Mos. 4 to 7	Mos. 8 to 11	Mos. 12 to 15
Minimum Payment	0.33	-0.21^{***}	-0.20^{***}	-0.06^{***}	-0.21^{***}	-0.21^{***}	-0.07^{***}
		(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.01)
Remaining Balance	46.46	-0.13^{***}	-0.70^{***}	-0.59^{***}	-0.26^{***}	-0.83^{***}	-0.72^{***}
		(0.05)	(0.06)	(0.08)	(0.04)	(0.06)	(0.07)
Δ Remaining Balance	0.21	-0.15^{***}	-0.14^{***}	0.10^{***}	-0.16^{***}	-0.15^{***}	0.09^{***}
		(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.02)
10+ Days Delinquent	0.22	-0.21^{***}	-0.15^{***}	-0.04^{***}	-0.21^{***}	-0.16^{***}	-0.05^{***}
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
90+ Days Delinquent	0.03	-0.07^{***}	-0.07^{***}	-0.06^{***}	-0.08^{***}	-0.08^{***}	-0.07^{***}
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on monthly loan repayment outcomes for borrowers with balances of $20,000 or more. Column 1 reports the dependent variable mean in the month prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive three-month periods, beginning with four months following the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same monthly effect for a regression which omits pre-call month dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 15,002 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A3: Difference-in-Differences Estimates of the Effect of IDR on Loan Repayment Outcomes: Low Balance Borrowers

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Mean $t = -1$	Mos. 4 to 7	Mos. 8 to 11	Mos. 12 to 15	Mos. 4 to 7	Mos. 8 to 11	Mos. 12 to 15
Minimum Payment	0.11	-0.07^{***}	-0.07^{***}	-0.03^{***}	-0.08^{***}	-0.08^{***}	-0.04^{***}
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Remaining Balance	10.48	-0.06^{***}	-0.20^{***}	-0.19^{***}	-0.10^{***}	-0.23^{***}	-0.23^{***}
		(0.02)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)
Δ Remaining Balance	0.05	-0.04^{***}	-0.02^{***}	0.02^{**}	-0.04^{***}	-0.02^{***}	0.02^{***}
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
10+ Days Delinquent	0.21	-0.22^{***}	-0.17^{***}	-0.03^{***}	-0.23^{***}	-0.18^{***}	-0.04^{***}
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
90+ Days Delinquent	0.02	-0.08^{***}	-0.08^{***}	-0.08^{***}	-0.09^{***}	-0.09^{***}	-0.08^{***}
		(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on monthly loan repayment outcomes for borrowers with balances less than $20,000. Column 1 reports the dependent variable mean in the month prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive three-month periods, beginning with four months following the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same monthly effect for a regression which omits pre-call month dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 16,111 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A4: Difference-in-Differences Estimates of the Effect of IDR on Financial Outcomes: High Balance Borrowers

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Score	597.46	5.26***	8.05***	4.58**	5.17**	7.97***	4.49*
		(1.41)	(1.87)	(2.22)	(2.04)	(2.42)	(2.71)
Bankruptcy	0.20	0.00	−0.01	−0.01	0.00	−0.00	−0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
Any Mortgage	0.25	0.00	0.01	0.02	0.01	0.02	0.02*
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Any Auto Trade	0.75	0.01	−0.00	0.01	0.01	−0.00	0.01
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on yearly financial outcomes for borrowers with balances of $20,000 or more. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same yearly effect for a regression which omits pre-call year dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 60,006 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A5: Difference-in-Differences Estimates of the Effect of IDR on Financial Outcomes: Low Balance Borrowers

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Score	588.95	4.53***	6.83***	7.35***	4.48*	6.78**	7.30**
		(1.61)	(2.13)	(2.38)	(2.31)	(2.77)	(2.89)
Bankruptcy	0.19	−0.01	−0.01	−0.01	−0.01	−0.01	−0.02
		(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)
Any Mortgage	0.15	0.00	0.03***	0.03***	−0.00	0.02**	0.03**
		(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Any Auto Trade	0.66	−0.02*	−0.00	0.00	−0.01	−0.00	0.01
		(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.02)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on yearly financial outcomes for borrowers with balances less than $20,000. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same yearly effect for a regression which omits pre-call year dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 64,441 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A6: Difference-in-Differences Estimates of the Effect of IDR on Credit Cards: High Balance Borrowers

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Card Balances	1.77	0.00	0.26**	0.48***	0.03	0.29**	0.51***
		(0.08)	(0.12)	(0.16)	(0.10)	(0.13)	(0.17)
Log Credit Card Balances	-2.20	0.07	0.33***	0.38***	0.10	0.37***	0.42***
		(0.07)	(0.09)	(0.10)	(0.09)	(0.11)	(0.12)
Number of Credit Cards	3.42	−0.02	0.07	0.17**	−0.02	0.07	0.16*
		(0.04)	(0.06)	(0.08)	(0.05)	(0.07)	(0.08)
Credit Card Limits	3.73	−0.26	0.04	0.23	−0.30	−0.01	0.19
		(0.19)	(0.25)	(0.32)	(0.23)	(0.28)	(0.35)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on yearly credit cards for borrowers with balances of $20,000 or more. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same yearly effect for a regression which omits pre-call year dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 60,006 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A7: Difference-in-Differences Estimates of the Effect of IDR on Credit Cards: Low Balance Borrowers

		Diff-in-Diff			Diff-in-Diff w Linear Trend		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Card Balances	1.06	0.10	0.18*	0.25*	−0.01	0.08	0.14
		(0.08)	(0.11)	(0.13)	(0.10)	(0.13)	(0.15)
Log Credit Card Balances	-2.83	0.07	0.32***	0.36***	0.00	0.26**	0.30**
		(0.07)	(0.10)	(0.11)	(0.09)	(0.11)	(0.13)
Number of Credit Cards	2.57	0.00	0.10	0.13*	−0.04	0.05	0.09
		(0.04)	(0.06)	(0.08)	(0.04)	(0.06)	(0.08)
Credit Card Limits	2.09	0.01	0.24	0.42*	−0.15	0.08	0.25
		(0.11)	(0.17)	(0.23)	(0.14)	(0.19)	(0.25)

Note: This table reports difference-in-differences estimates of the effect of IDR enrollment on yearly credit cards for borrowers with balances less than $20,000. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Columns 5-7 report coefficents on the same yearly effect for a regression which omits pre-call year dummies and includes a linear time trend. The regressions are estimated on the analysis sample as described in the notes to Table 2.1. Both specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Sample size is 64,441 calls. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A8: Difference-in-Differences and Instrumental Variables Estimates of the Effect of IDR on Financial Outcomes

		Difference-in-Differences			Instrumental Variables		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Score	593.26	5.01***	7.53***	5.93***	−8.83	−20.07	−28.65
		(1.06)	(1.40)	(1.63)	(18.29)	(18.40)	(28.09)
Bankruptcy	0.19	−0.00	−0.00	−0.00	0.10	0.14	0.17
		(0.00)	(0.00)	(0.00)	(0.15)	(0.17)	(0.19)
Any Mortgage	0.20	0.00	0.02**	0.02***	0.03	−0.02	−0.07
		(0.00)	(0.01)	(0.01)	(0.09)	(0.12)	(0.13)
Any Auto Trade	0.71	−0.00	−0.00	0.01	0.03	−0.12	−0.16
		(0.01)	(0.01)	(0.01)	(0.16)	(0.16)	(0.15)
Call Time FE		Yes	Yes	Yes	Yes	Yes	Yes
Controls		Yes	Yes	Yes	Yes	Yes	Yes
N	15953	15953	15953	15953	15953	15953	15953

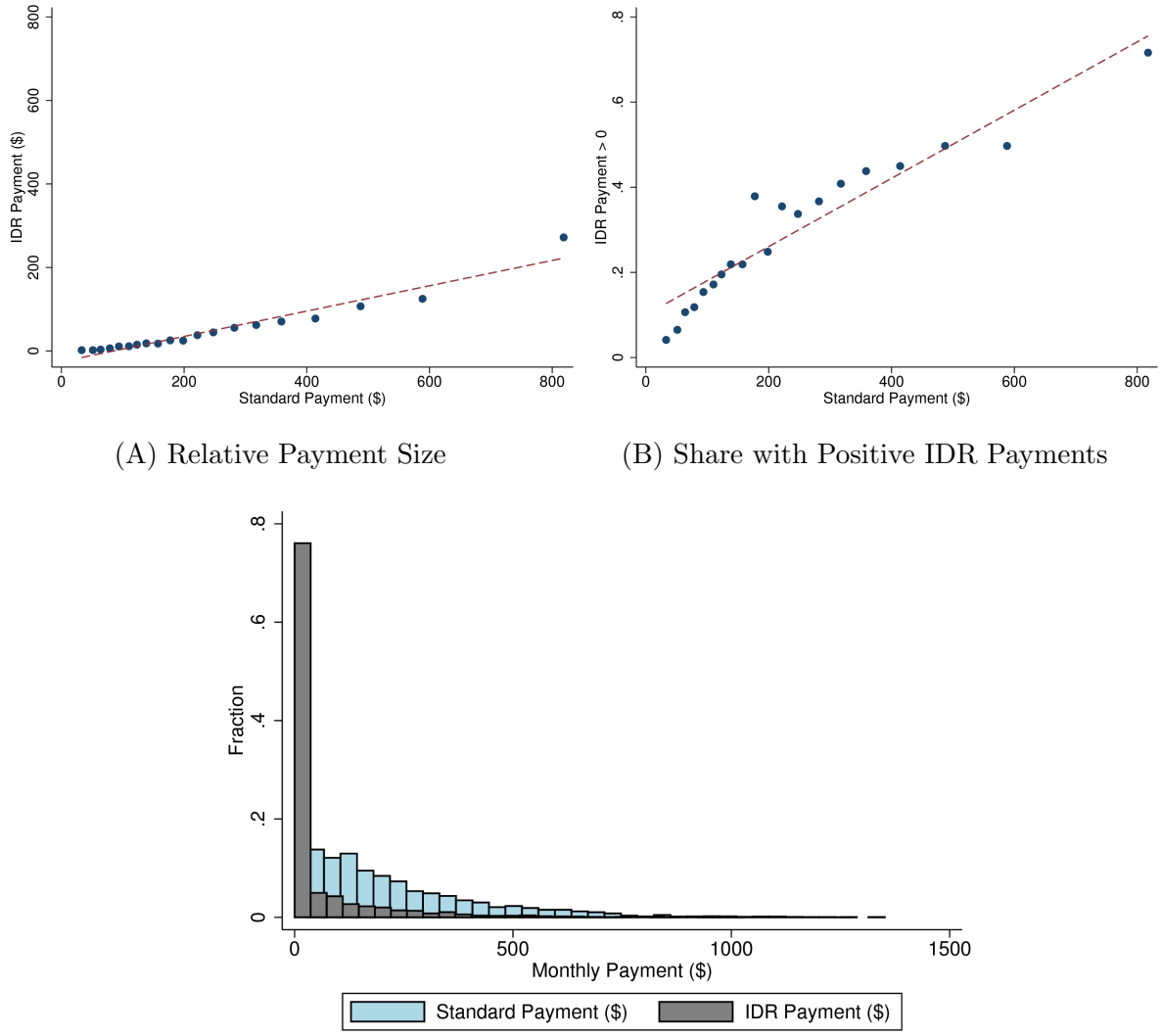
Note: This table reports difference-in-differences and two-stage least squares estimates of the effect of IDR enrollment on yearly financial outcomes. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Each of Columns 5-7 report estimates from separate two-stage least squares regressions on outcomes in the same years. Regressions are estimated on the analysis sample as described in the notes to Table 2.1. Two-stage least squares models instrument for IDR with the agent score calculated using data from other calls made by the agent following the procedure described in Section 1.5.2. All specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Table A9: Difference-in-Differences and Instrumental Variables Estimates of the Effect of IDR on Credit Cards

		Difference-in-Differences			Instrumental Variables		
Dependent Variable	(1) Mean $t = -1$	(2) Year 0	(3) Year 1	(4) Year 2	(5) Year 0	(6) Year 1	(7) Year 2
Credit Card Balances	1.42	0.05	0.24***	0.40***	−1.55	−1.52	−0.44
		(0.05)	(0.08)	(0.11)	(1.13)	(1.65)	(1.86)
Log Credit Card Balances	-2.51	0.07	0.33***	0.38***	−0.53	−0.03	0.47
		(0.05)	(0.07)	(0.07)	(1.05)	(1.32)	(1.42)
Number of Credit Cards	3.00	−0.01	0.07	0.14**	−0.39	−0.36	0.15
		(0.03)	(0.04)	(0.06)	(0.49)	(0.79)	(1.02)
Credit Card Limits	2.92	−0.13	0.14	0.34	−3.72**	−2.58	−1.06
		(0.12)	(0.16)	(0.21)	(1.88)	(2.65)	(3.49)
Call Time FE		Yes	Yes	Yes	Yes	Yes	Yes
Controls		Yes	Yes	Yes	Yes	Yes	Yes
N	15953	15953	15953	15953	15953	15953	15953

Note: This table reports difference-in-differences and two-stage least squares estimates of the effect of IDR enrollment on yearly credit cards. Column 1 reports the dependent variable mean in the year prior to receiving a delinquency call. Columns 2-4 report coefficients on the effect of IDR in consecutive years, beginning with the year of the delinquency call from the pooled OLS regression specified in Equation 1.18. Each of Columns 5-7 report estimates from separate two-stage least squares regressions on outcomes in the same years. Regressions are estimated on the analysis sample as described in the notes to Table 2.1. Two-stage least squares models instrument for IDR with the agent score calculated using data from other calls made by the agent following the procedure described in Section 1.5.2. All specifications include controls from call date and time, as well as amount borrowed, inferred gender, and zip-code median income. Robust standard errors, in parentheses, are clustered at the individual level. *** = significant at 1 percent level, ** = significant at 5 percent level, * = significant at 10 percent level.

Figure A1: Standard versus IDR Payments among IDR-Enrollees



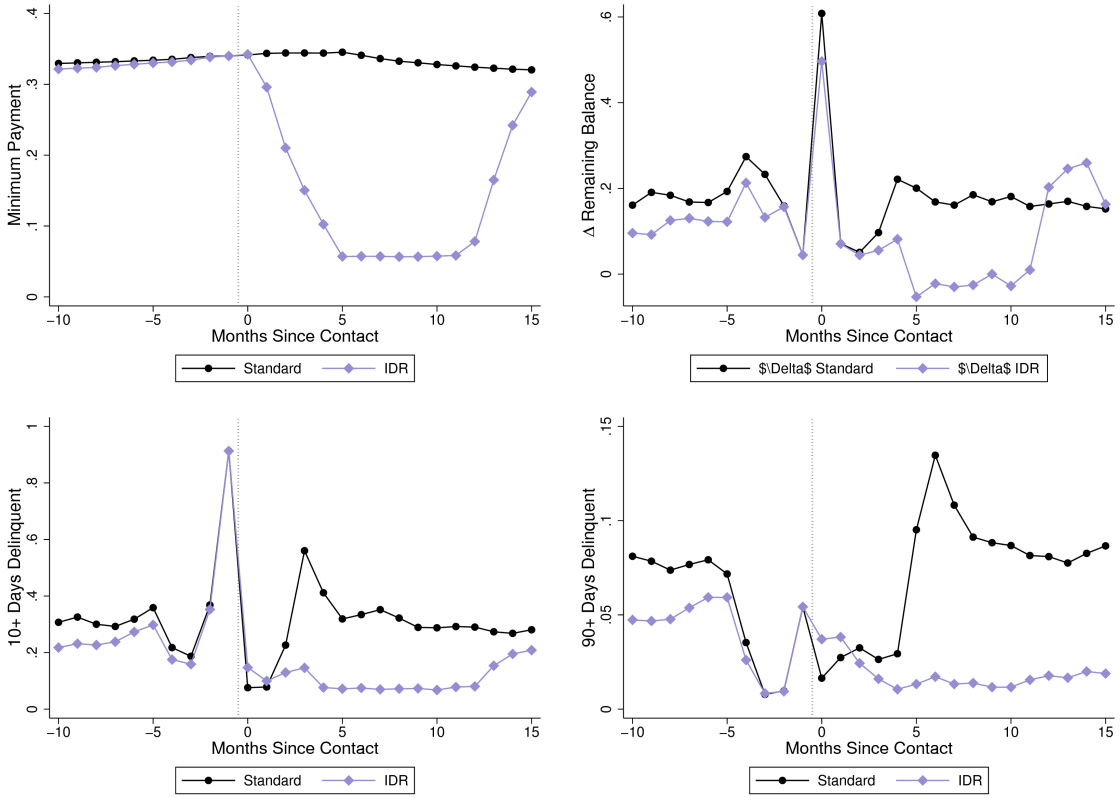
(A) Relative Payment Size

(B) Share with Positive IDR Payments

(C) Distribution of Payments

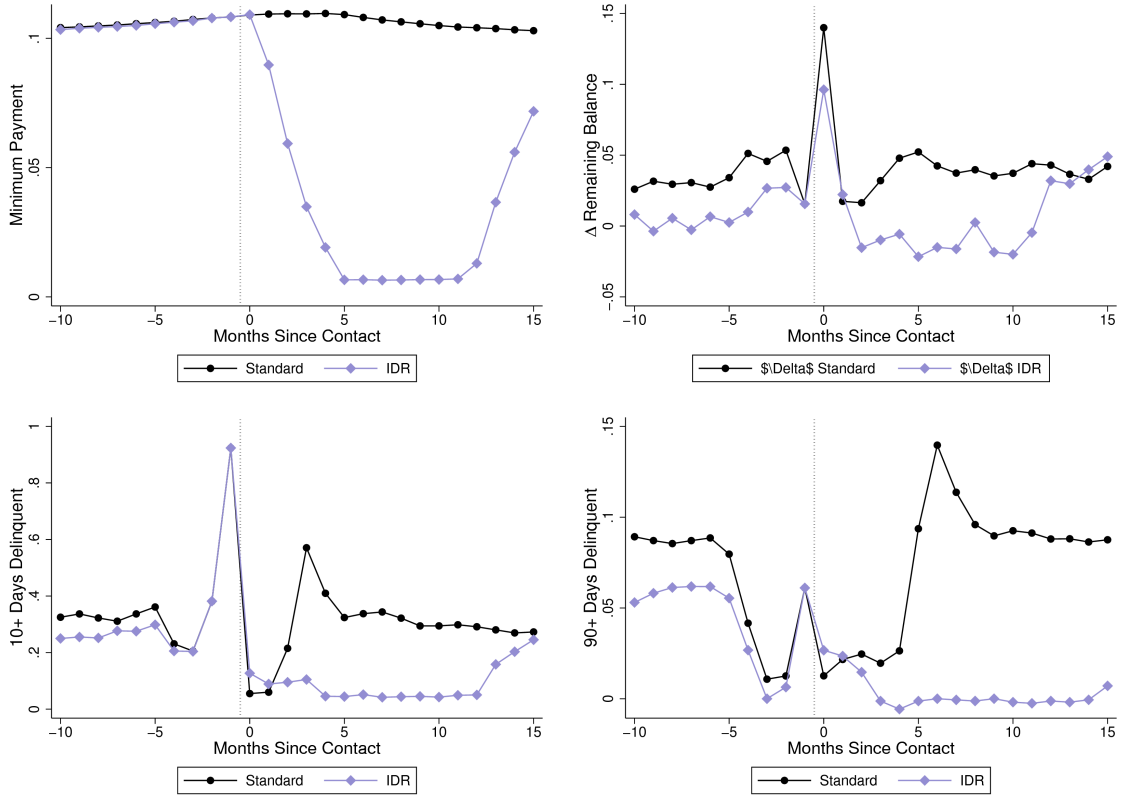
Note: This figure plots the relationship between pre-call standard payments and post-call IDR payments. The binned scatter plot is constructed using payment amounts one month before and six months following the delinquency call for borrowers in the analysis sample who take up IDR. Panel A plots average standard payment size against average IDR payment size. Panel B plots average standard payment size against the share of individuals with IDR payments greater than zero. Panel C plots histograms for standard and IDR payments. See Table 2.1 notes for additional details on the sample.

Figure A2: Pre/Post-Call Trends in Repayment Outcomes: High Balance Borrowers



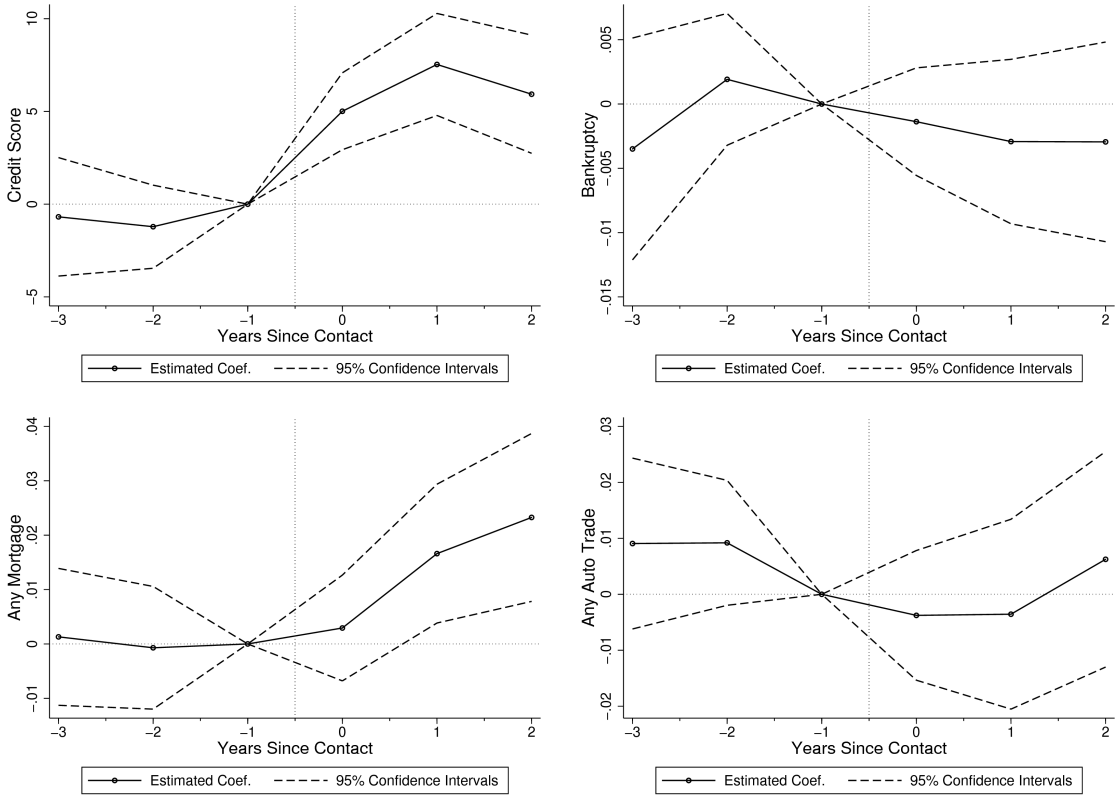
Note: This figure plots the average loan repayment outcomes for IDR enrollees and non-enrollees with $20,000 or more in debt balance in the analysis sample. The horizontal axis denotes time, in months, relative to the month of the loan servicing call. Outcomes are normalized to the average value of the outcome for non-enrollees in the month prior to the call. See Table 2.1 notes for additional details on the outcome measures and sample.

Figure A3: Pre/Post-Call Trends in Repayment Outcomes: Low Balance Borrowers



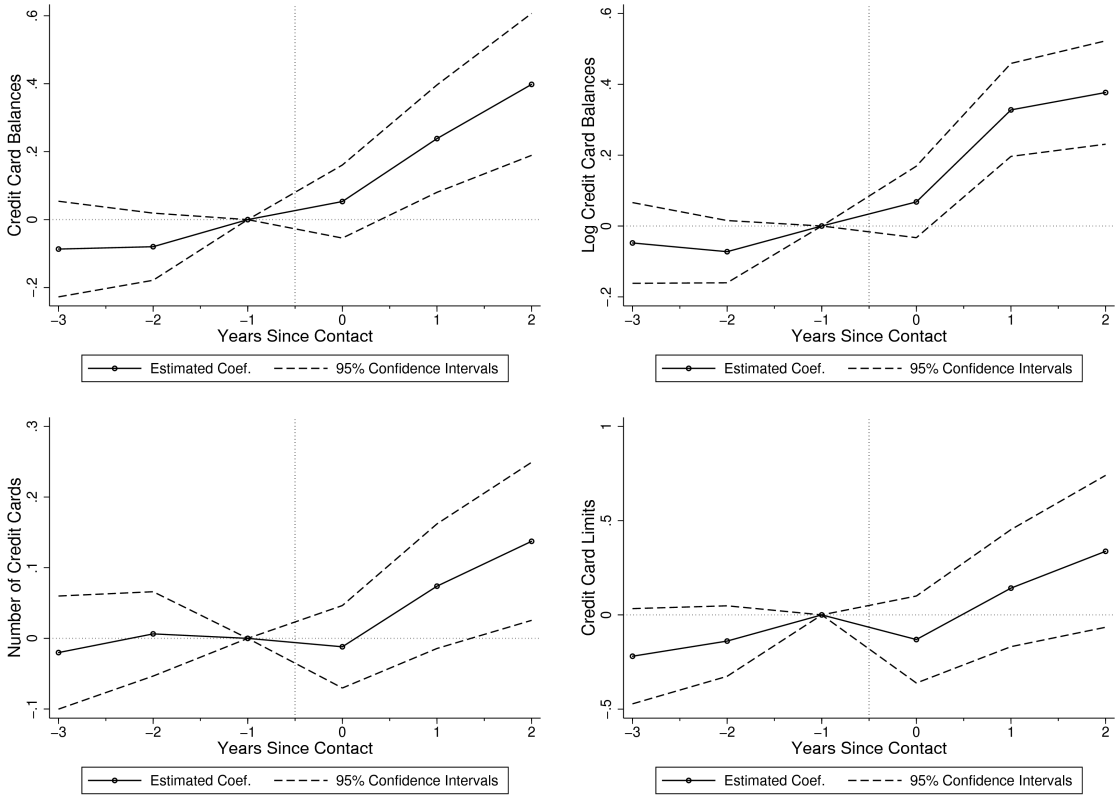
Note: This figure plots the average loan repayment outcomes for IDR enrollees and non-enrollees with less than $20,000 in debt balance in the analysis sample. The horizontal axis denotes time, in months, relative to the month of the loan servicing call. Outcomes are normalized to the average value of the outcome for non-enrollees in the month prior to the call. See Table 2.1 notes for additional details on the outcome measures and sample.

Figure A4: Difference-in-Differences: Financial Outcomes



Note: This figure reports estimated difference-in-differences coefficients for the effect of IDR on financial outcomes. Each point represents the estimated effect of post-call IDR status on the outcome variable at a given time period relative to the date of delinquency call. Relative years are plotted along the x-axis. Dashed lines represent 95% confidence intervals. All regressions include the full set of call date and time fixed effects. Robust standard errors are two-way clustered at the borrower and agent level.

Figure A5: Difference-in-Differences: Credit Cards



Note: This figure reports estimated difference-in-differences coefficients for the effect of IDR on credit card outcomes. Each point represents the estimated effect of post-call IDR status on the outcome variable at a given time period relative to the date of delinquency call. Relative years are plotted along the x-axis. Dashed lines represent 95% confidence intervals. All regressions include the full set of call date and time fixed effects. Robust standard errors are two-way clustered at the borrower and agent level.

A.2 Mathematical Appendix

A.2.1 Model Solution

The agent's problem is given by:

$$V\left(y_t, D_t, A_t\right) = \max_{c_t, A_{t+1}} \left\{u\left(c_t\right) + E\left[V\left(y_{t+1}, D_{t+1}, A_{t+1}\right)\right]\right\} \tag{A.1}$$

$$\begin{aligned} A_{t+1} &\geq L && \text{(A.2)} \\ c_t &= A_t - A_{t+1} + y_t - x_t && \text{(A.3)} \\ D_{t+1} &= D_t - x_t && \text{(A.4)} \end{aligned}$$

The Lagrangian for this problem can be written as:

$$\begin{aligned} \mathcal{L} = {} & u\left(c_t\right) + E\left[V\left(y_{t+1}, D_{t+1}, A_{t+1}\right)\right] \\ & - \lambda_1\left(L - A_{t+1}\right) - \lambda_2\left(A_t - A_{t+1} + y_t - x_t - c_t\right) - \lambda_3\left(D_t - D_{t+1} - x_t\right) \end{aligned} \tag{A.5}$$

where $\lambda_1 \geq 0$, $\lambda_2 > 0$, and $\lambda_3 > 0$. The first-order condition for consumption c_t is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial c_t}\left(y_{t+1}, D_{t+1}, A_{t+1}, A_t, c_t\right) &= 0 \\ u'(c_t) &= -\lambda_2 \end{aligned} \tag{A.6}$$

and the first-order condition for next period's assets A_{t+1} is:

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial A_{t+1}}\left(y_{t+1}, D_{t+1}, A_{t+1}, A_t, c_t\right) &= 0 \\ \frac{dE\left[V\left(y_{t+1}, D_{t+1}, A_{t+1}\right)\right]}{dA_{t+1}} &= -\lambda_1 - \lambda_2 \end{aligned} \tag{A.7}$$

Applying the envelope theorem to Equation A.7 and combining with Equation A.6 yields:

$$u'(c_t) = E\left[\frac{\partial V}{\partial A}(y_{t+1}, A_{t+1}, D_{t+1})\right] + \lambda_1 \tag{A.8}$$

Since $\lambda_1 > 0$ if and only if $A_{t+1} = L$, Equation A.8 implies:

$$u'(c_t) = E\left[\frac{\partial V}{\partial A}(y_{t+1}, A_{t+1}, D_{t+1})\right] \text{ if } A_{t+1} > L \tag{A.9}$$

$$u'(c_t) > E\left[\frac{\partial V}{\partial A}(y_{t+1}, A_{t+1}, D_{t+1})\right] \text{ if } A_{t+1} = L \tag{A.10}$$

which, along with conditions A.2, A.3, and A.4, characterize the solution to the model.

A.2.2 Proof of Liquidity Effects

Note that the solution from the previous section implies:

$$c_t = \begin{cases} E[c_{t+1}] & \text{if } A_{t+1} > L \\ A_t - L + y_t - x_t & \text{if } A_{t+1} = L \end{cases} \tag{A.11}$$

Consider the solution $c_t^* = E[c_{t+1}]$ when the borrowing constraint does not bind. In this case $A_{t+1}^* > L$, so

$$A_t + y_t - x_t - c_t^* > L \tag{A.12}$$

$$c_t^* < A_t - L + y_t - x_t \tag{A.13}$$

$$E[c_{t+1}] < A_t - L + y_t - x_t \tag{A.14}$$

We can therefore rewrite (A.11) as:

$$c_t = \min\{A_t - L + y_t - x_t, E[c_{t+1}]\} \tag{A.15}$$

Iterating forward, we have

$$c_t = \min\left\{A_t - L + y_t - x_t, E\left[\min\left\{A_{t+1} - L + y_{t+1} - x_{t+1}, E\left[c_{t+2}\right]\right\}\right]\right\} \tag{A.16}$$

Smoothing Effect

First I analyze how IDR affects current consumption through a smoothing of the expected net-income profile. Note that for all t, IDR "backloads" minimum payments relative to standard plans:

$$E\left[x_t^I\right] = E\left[\min\left\{\theta y_t, \frac{D_0}{N}\right\}\right] \leq \frac{D_0}{N} = E\left[x_t^S\right] \qquad \forall t \leq N \tag{A.17}$$

$$E\left[x_t^I\right] = E\left[\min\left\{\theta y_t, \frac{D_0}{N}, D_t\right\}\right] \geq 0 = E\left[x_t^S\right] \qquad \forall t > N \tag{A.18}$$

Now consider the effects of a decrease in period-k minimum payments x_k on period-t consumption c_t, for all $k \in [t, N]$. Without loss of generality, let $N = t+1$. Period-t consumption increases through two channels: first, lower x_t means the ceiling on present-day consumption is mechanically lifted by Δx_t. Second, borrowers *expect* a higher ceiling on consumption in period $t+1$. As long as period $t+1$ liquidity constraints held some positive probability of binding under the original payment scheme, reducing minimum payments will increase current-period consumption by lowering precautionary savings; borrowers spend down their assets today because they expect less need to supplement their potentially constrained consumption tomorrow. Formally, suppose that $\frac{dc_t}{dx_t} > 0$, which would require the right-hand side of A.15 to decrease following a drop in x_t. The budget constraint implies this scenario would only be possible if A_{t+1} increased more than $-\Delta x$, which would, in turn, increase $E[c_{t+1}]$ which would violate Equation A.11.

Now consider a commensurate *increase* in period-q minimum payments x_q on period-t consumption c_t, for all $q > N$. Assume that the period-t borrowing constraint does not bind

and expand Equation A.16 an additional period, so:

$$c_t = E\left[\min\left\{A_{t+1} - L + y_{t+1} - x_{t+1}, E\left[\min\left\{A_{t+2} - L + y_{t+2} - x_{t+2}, E\left[c_{t+3}\right]\right\}\right]\right\}\right] \quad \text{(A.19)}$$

By the same logic as above, an increase in period $t+2$ payments x_{t+2} *decreases* period t consumption. However, the net effect of a revenue-neutral "backloading" of payments must be positive. To see why, recall that $E[y_q] \leq E[y_{q'}]$ for all $q < q'$, and compare the effects of payment changes in different periods on period-t consumption in Equation A.19:

$$\frac{dc_t}{dx_q} \geq \frac{dc_t}{dx_{q'}} \quad \forall\ q' > q \quad \text{(A.20)}$$

Any change in payments such that $\Delta x_q < \Delta x_{q'}$ but $\sum_t^\infty \Delta x_q = 0$ implies a net *increase* in c_t. Intuitively, the precautionary response to net income changes in the near future is greater than the response to net income changes in the distant future, as borrowers can gradually accumulate precautionary savings for the latter over many periods.

Insurance Effect

In addition to flattening the net-income profiles, IDR also reduces the per-period variance of net income. Let m denote income net of loan payments, $m \equiv y - x$.

$$\begin{aligned}
\text{Var}(m^I) - \text{Var}(m^S) &= \int_{-\infty}^{\infty} \left(y - x^I\right)^2 dF_y - \int_{-\infty}^{\infty} \left(y - x^S\right)^2 dF_y & \text{(A.21)}\\
&= \int_{-\infty}^{\infty} \left(y - \max\left\{(1-\theta)y, y - x_t^S\right\}\right)^2 dF_y - \int_{-\infty}^{\infty} \left(y - x^S\right)^2 dF_y & \text{(A.22)}\\
&= \int_{-\infty}^{\frac{x^S}{\theta}} \left[(y - \theta y)^2 - \left(y - x^I\right)^2\right] dF_y < 0 & \text{(A.23)}
\end{aligned}$$

Rewriting Equation A.16 in terms of net income, we have:

$$c_t = \min\left\{A_t - L + m_t, E\left[\min\left\{A_{t+1} - L + m_{t+1}, E\left[c_{t+2}\right]\right\}\right]\right\} \quad \text{(A.24)}$$

Note that for any period $k > t$, a mean-preserving contraction in m_t will decrease the likelihood of low realizations of net income and, hence, a binding liquidity constraint; a decrease in $\text{Var}(m_{t+1})$ will increase the value of $E\left[\min\left\{A_{t+1} - L + m_{t+1}, E\left[c_{t+2}\right]\right\}\right]$, inducing less precautionary savings and more consumption in period t.

Appendix B

Chapter 2 Appendix

B.1 Technical Appendix

B.1.1 Proof of Theorem 1

First note that $\mu \neq E[Y|X] \implies \mathcal{F}_\theta(\cdot) \neq F(\cdot|X)$. Under Assumptions 1 and 2,

$$\begin{aligned}\mathbb{E}[\mu|X,Z] &= \mathbb{E}[\mathbb{E}[Y|\Theta]|X,Z] \\ &= \mathbb{E}[Y|X,Z]\end{aligned}$$

So $E[Y|X,Z]$ is a mean-preserving spread of μ, and

$$\begin{aligned} & E[Y|X,Z] && \neq E[Y|X] \\ \implies \quad & \mathbb{E}[\mu|X,Z] && \neq E[Y|X] \\ \implies \quad & \mu_i && \neq E[Y|X] \end{aligned} \tag{B.1}$$

B.1.2 Flexible Contract Model

Before graduating, borrowers can contract with an "insurer" receive lump-sump college payment b in exchange for a commitment to pay income-contingent amount $x(y)$ once income is realized. Borrowers choose contracts $x(y)$ to maximize expected post-graduate utility given by:

$$\int_0^\infty u(y - x(y))dy$$

where $u(\cdot)$ holds standard properties for von Neumann-Morgenstern preferences.

The insurer offers a menu of contracts $\boldsymbol{x}(y)$, each of which is denoted according to the type θ of the intended borrower, i.e. $x(y;\theta)$. The insurer's contracting problem is given by:

$$\max_{\{x(y;\theta)\}} \int_{\theta\in\Theta} \mathbb{E}\left[x(y;\theta) - D|\theta\right] d\boldsymbol{G}(\theta)$$

$$\text{s.t.} \quad \begin{aligned} \mathbb{E}\left[u(y - x(y;\theta))|\theta\right] &\geq \mathbb{E}\left[u(y - x(y;\hat{\theta})|\theta\right] \ \forall\theta,\hat{\theta} \quad \text{(IC)} \\ \mathbb{E}\left[u(y - x(y;\theta))|\theta\right] &\geq \mathbb{E}\left[u(y - D)|\theta\right] \ \forall\theta \quad \text{(IR)} \end{aligned}$$

The risk-neutral insurer offers each contract $x(y,\theta)$ to maximize expected profits subject to the constraints imposed by borrower behavior. Specifically, the incentive-compatibility constraint (IC) ensures that $x(y;\theta)$ is the optimal choice for borrower type θ given the complete menu of insurance options, and the individual rationality constraint (IR) requires that no borrower's contract be worse than her outside option of outright debt repayment.

Appendix C

Chapter 3 Appendix

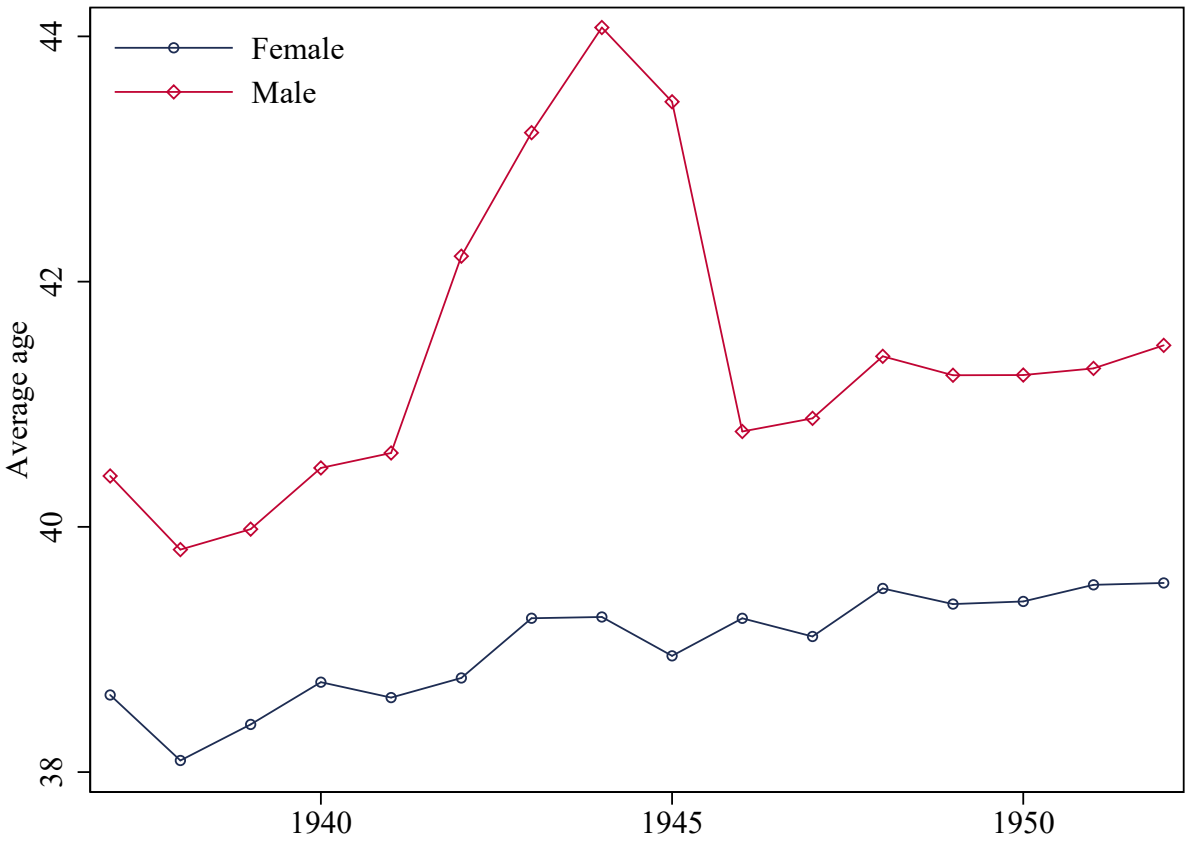
C.1 Supplementary Figures and Tables Noted in the Text

Appendix Figure C.1.1: Comparing unemployment rates in Gallup and the HSUS



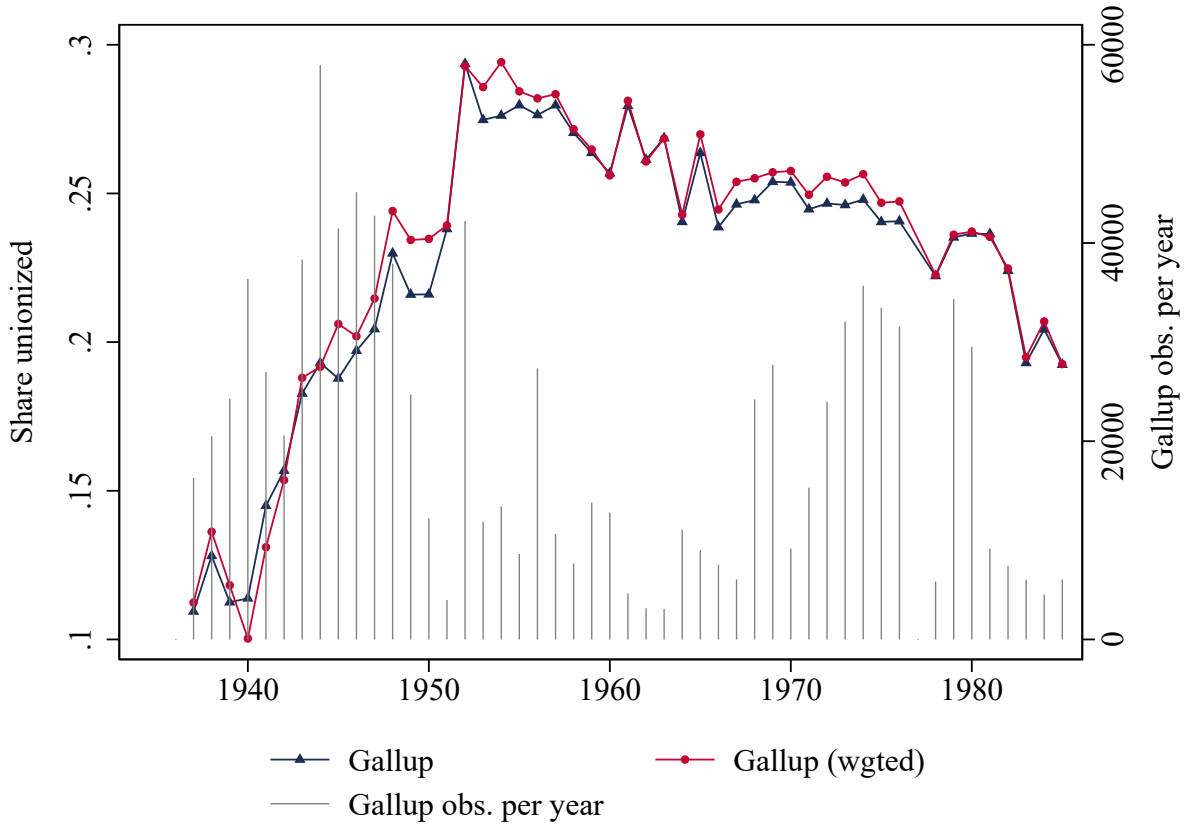
Data sources: Gallup and Historical Statistics of the United States (HSUS)
Notes: Sample in Gallup includes farmers

Appendix Figure C.1.2: Age distribution in Gallup, by gender, 1937-1952



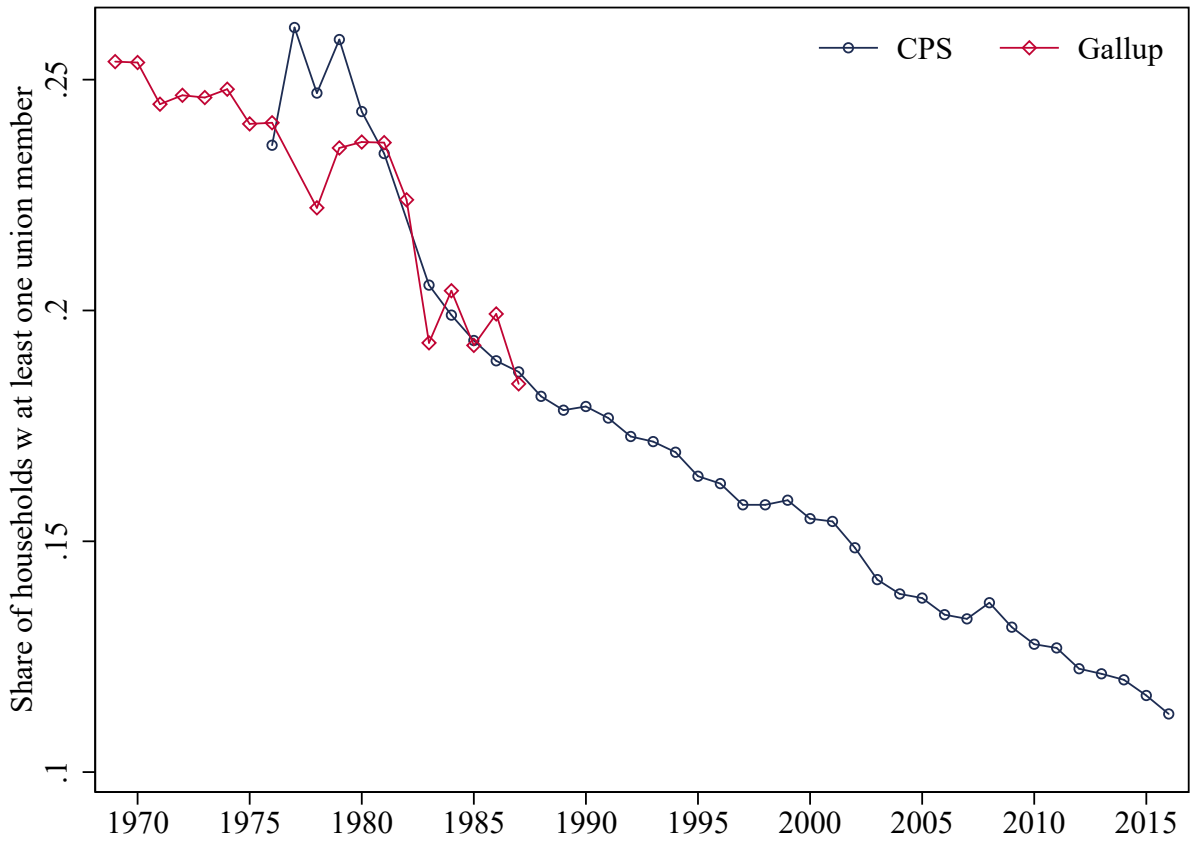
Data sources: Gallup microdata.

Appendix Figure C.1.3: Union share of households in the Gallup data (weighted vs. unweighted)



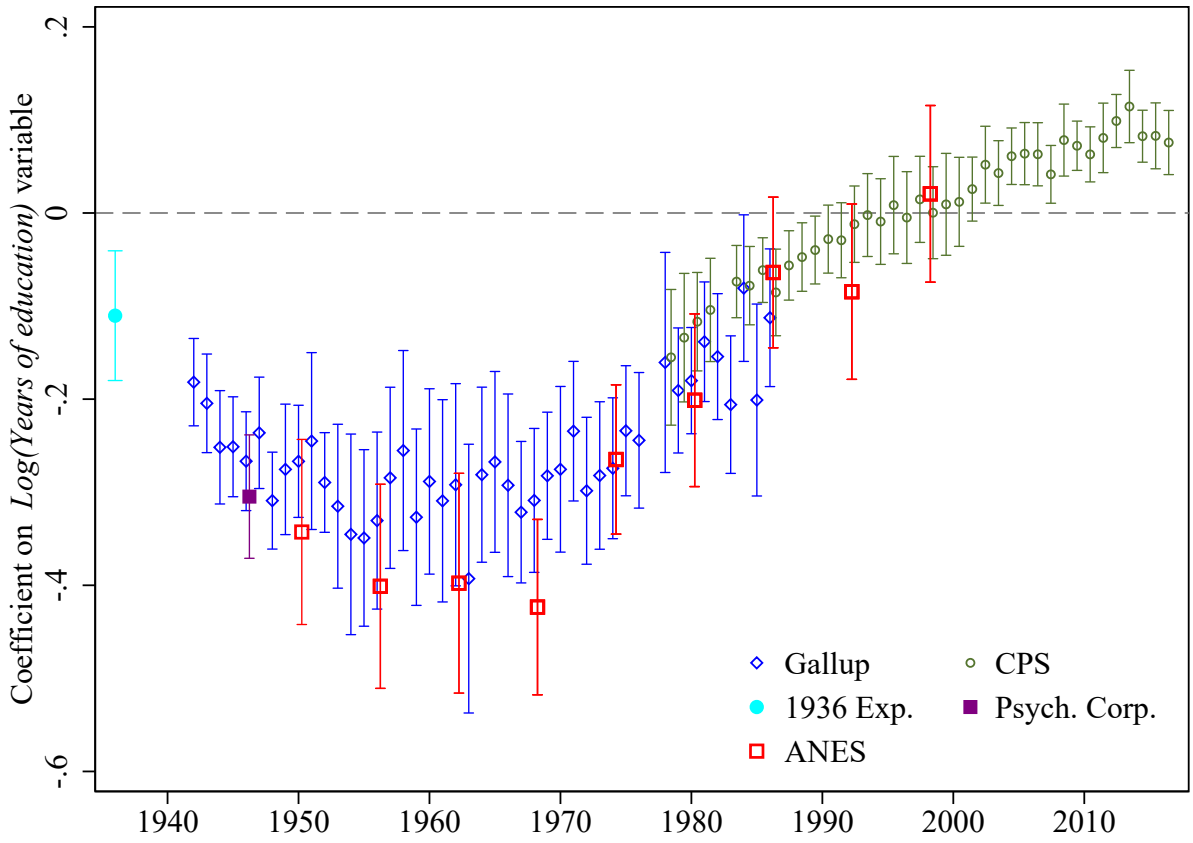
Data sources: Gallup. See Section 3.3 and Appendix C.3 for more detail on data and weight construction.

Appendix Figure C.1.4: Comparing union density in Gallup and CPS, 1970–present



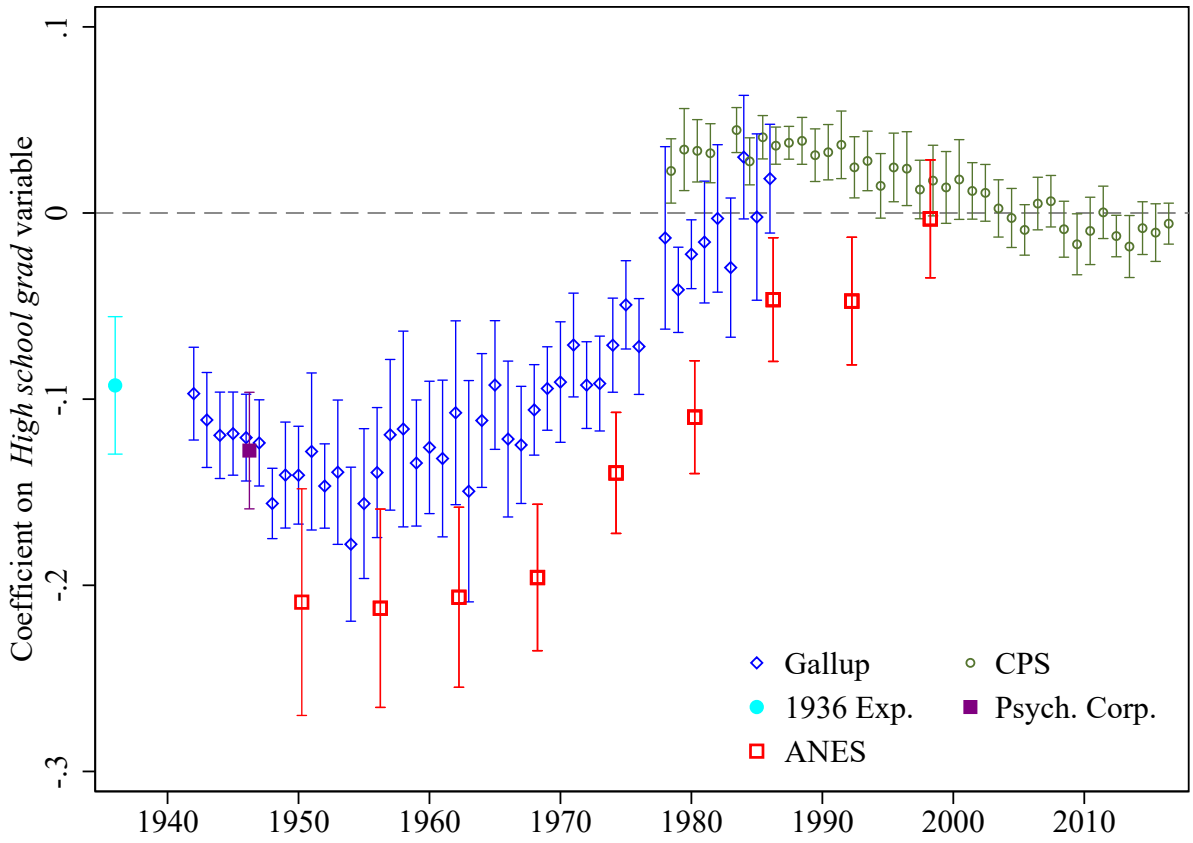
Data sources: Gallup and Current Population Survey

Appendix Figure C.1.5: Selection of union households by log years schooling



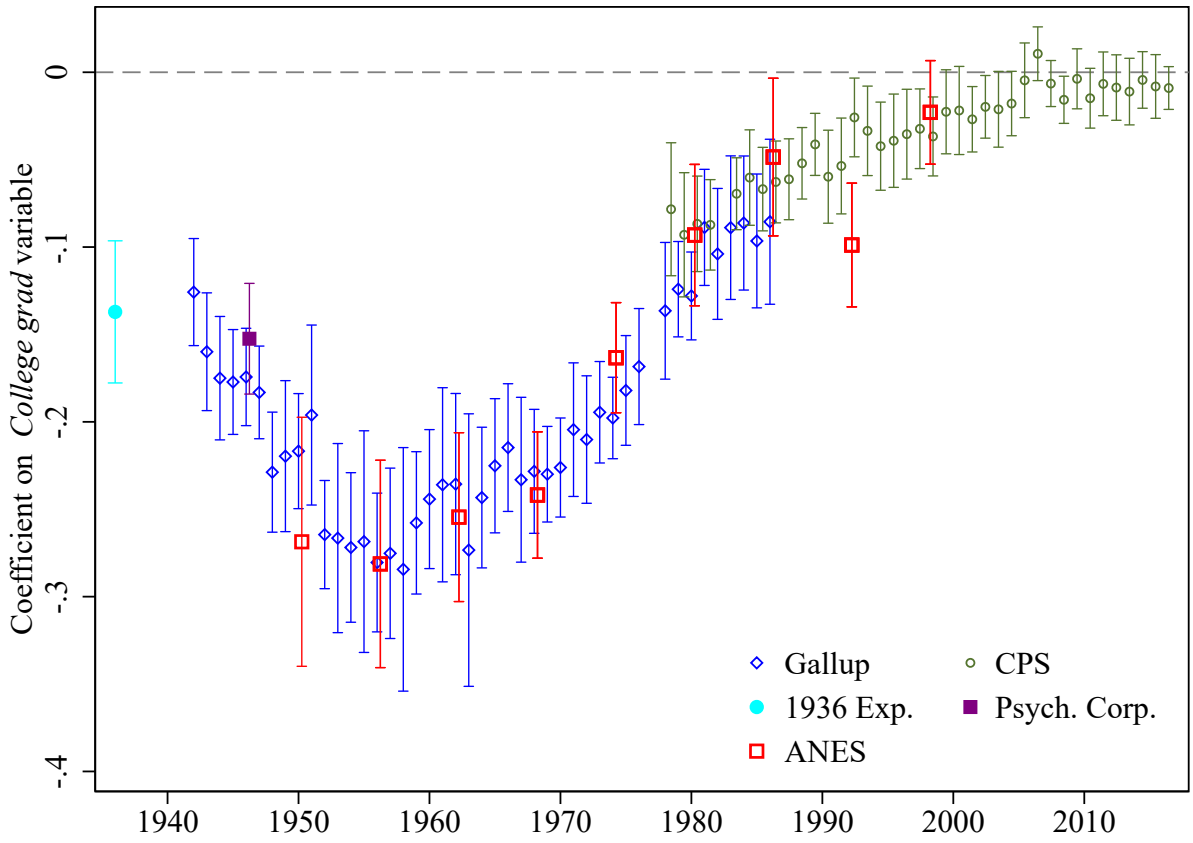
Sources: Gallup, 1937–1986. CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946.
Notes: For each data source, we estimate, separately by data source and year, household union status on *Log Years Education*, state s and survey-date t fixed effects, age and its square, and gender. We plot in this graph the coefficients on *Log Years Education* from each of these estimations. We control for survey-date and state fixed effects, age and its square, and gender. For the ANES, because the samples are smaller, we group surveys into six-year bins. Standard errors are clustered by state.

Appendix Figure C.1.6: Selection of union households by high-school graduation



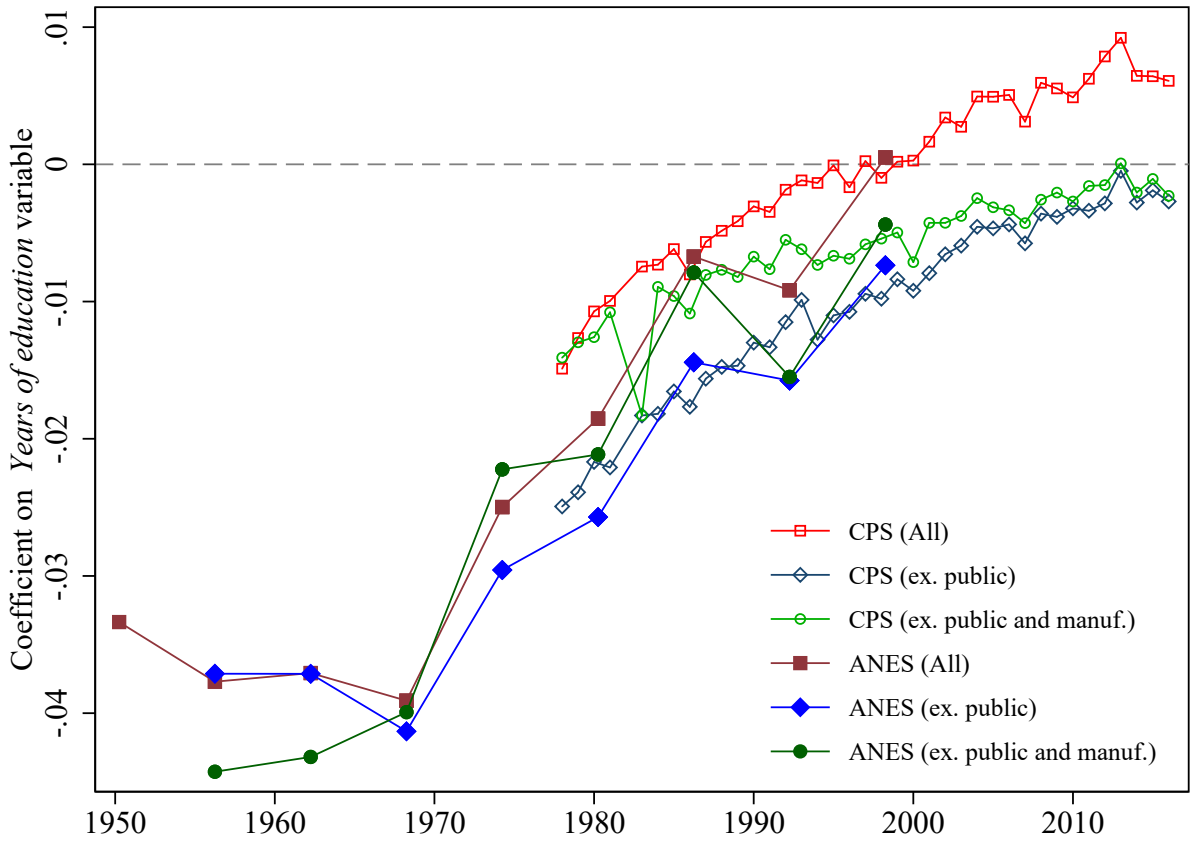
Sources: Gallup, 1937–1986. CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946.
Notes: For each data source, we estimate, separately by data source and year, household union status on a *High School Grad* dummy variable, state s and survey-date t fixed effects, age and its square, and gender. We plot in this graph the coefficients on *High School Grad* from each of these estimations. We control for survey-date and state fixed effects, age and its square, and gender. For the ANES, because the samples are smaller, we group surveys into six-year bins. Standard errors are clustered by state.

Appendix Figure C.1.7: Selection of union households by college graduation



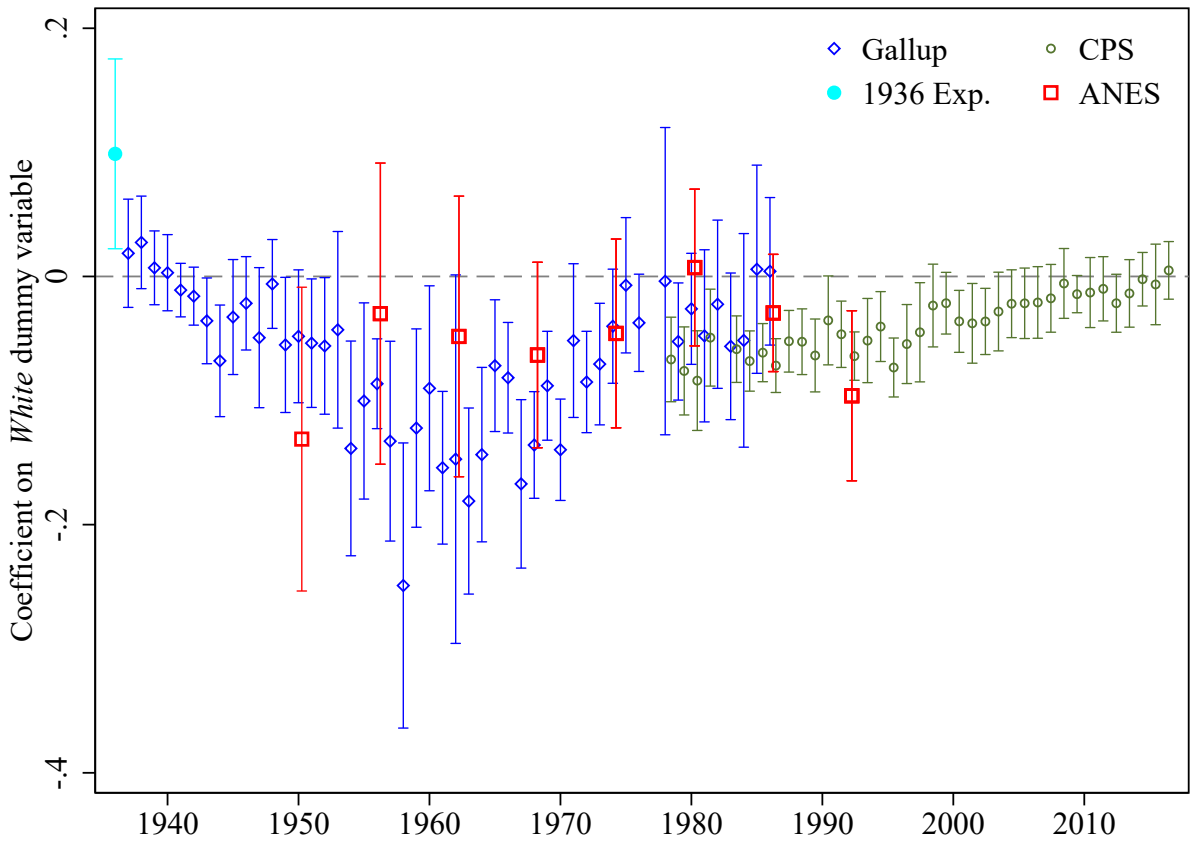
Sources: Gallup, 1937–1986. CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946.
Notes: For each data source, we estimate, separately by data source and year, household union status on a *College Grad* dummy variable, state s and survey-date t fixed effects, age and its square, and gender. We plot in this graph the coefficients on *College Grad* from each of these estimations. We control for survey-date and state fixed effects, age and its square, and gender. For the ANES, because the samples are smaller, we group surveys into six-year bins. Standard errors are clustered by state.

Appendix Figure C.1.8: Selection of union households by education in the ANES and CPS (dropping households with a public- or manufacturing-sector worker)



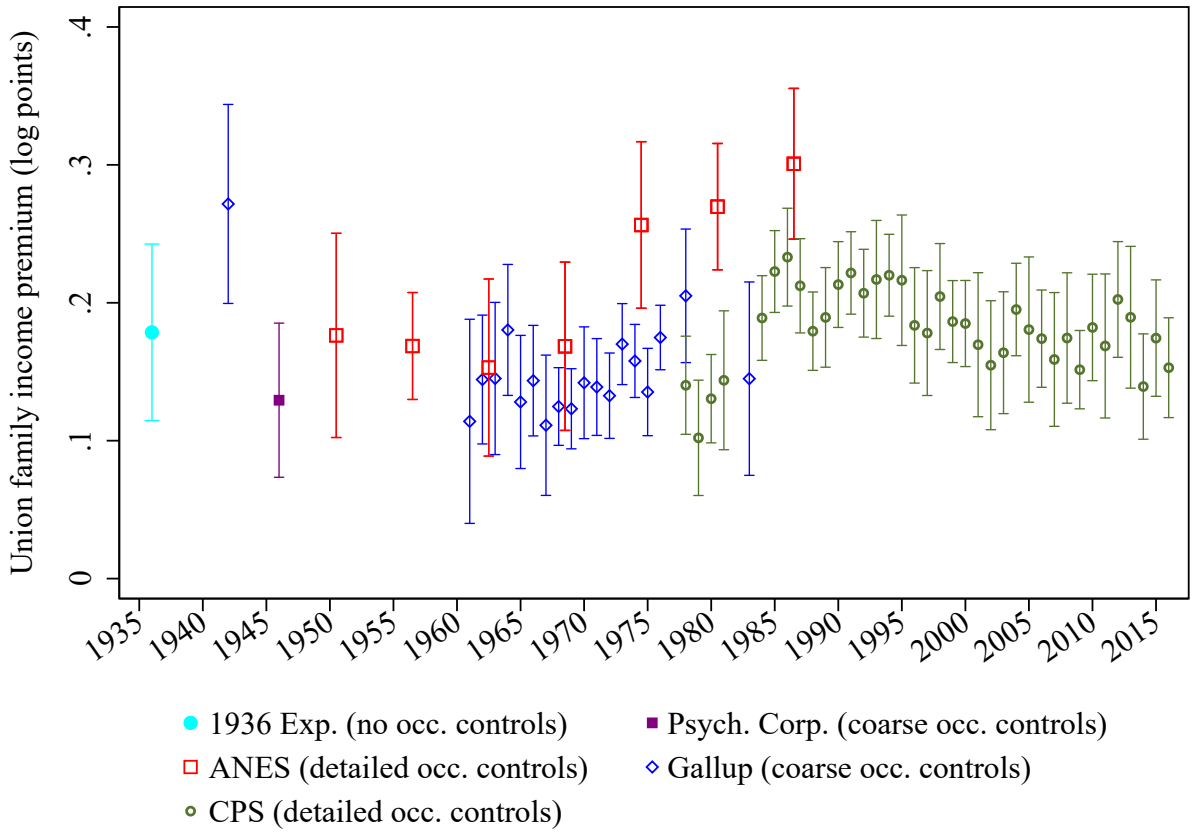
Sources: CPS, 1978–2016; ANES, 1952–1996.
Notes: For each data source, we estimate, separately by data source and year, household union status on a *Years of education* variable, state s and survey-date t fixed effects, age and its square, and gender. We plot in this graph the coefficients on *Years of education* from each of these estimations. We control for survey-date and state fixed effects, age and its square, and gender. For the ANES, because the samples are smaller, we group surveys into six-year bins. Note that we only include ANES and CPS in this graph, because other data sources do not allow us to identify industrial sectors of workers in the household.

Appendix Figure C.1.9: Selection of union households by race (dropping Southern states)



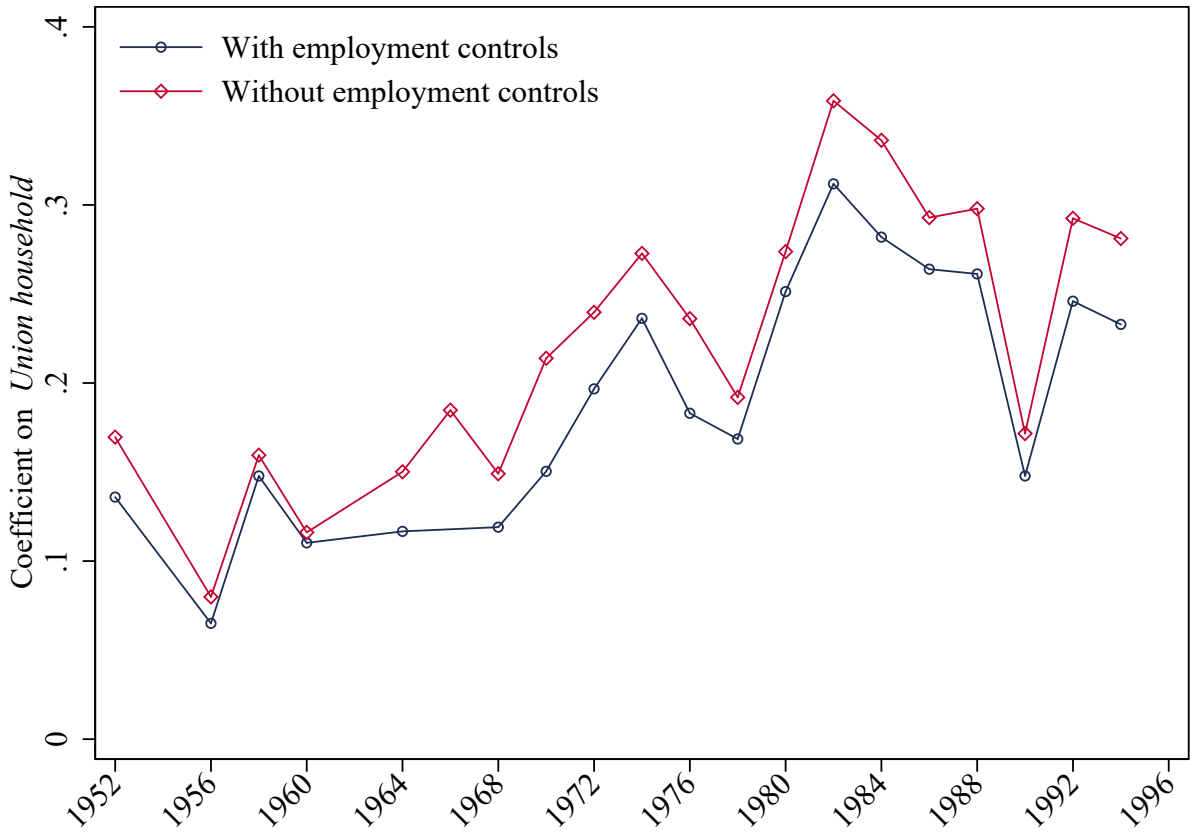
Data sources: Gallup data, 1937–1986; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996. See Section C.3 for a description of each data source.
Notes: For each data source, we estimate, separately by data source and year, household union status on a *White* dummy variable, state s and survey-date t fixed effects, age and its square, and gender. We plot in this graph the coefficients on *White* from each of these estimations. For the ANES, because the samples are smaller, we group surveys into six-year bins. Note that we cannot use the U.S. Psychological Corporation survey in this figure because, while it has state identifiers (thus we can thus control for state fixed effects) the codebook does *not* provide the state names that correspond to the codes (so we cannot drop the South). Confidence intervals are based on standard errors clustered by state.

Appendix Figure C.1.10: Estimates of the union family income premium (including occupation controls when available)



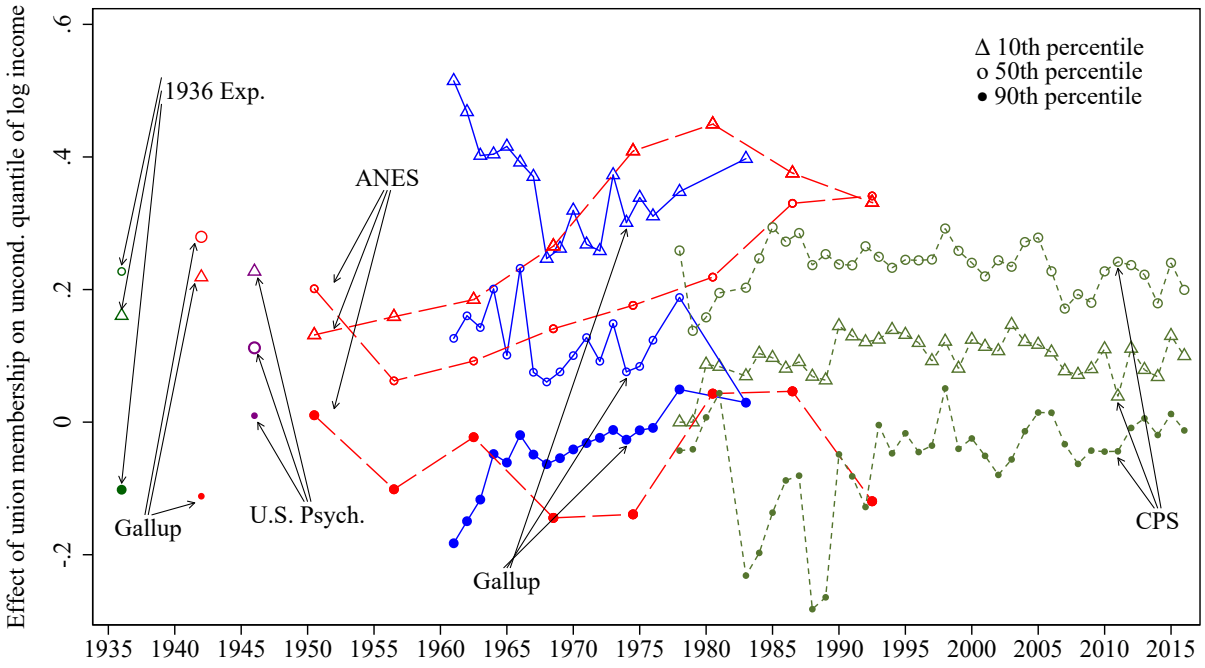
Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section C.3 for a description of each data source. See Appendix C.2 for details on family income variable construction.
Notes: Each plotted point comes from estimating equation (3.3), which regressed log family income on controls for age, gender, race, state and survey-date fixed effects and (in most cases) fixed effects for the occupation of the head. For the ANES, because the samples are smaller, we group surveys into six-year bins. The plotted confidence intervals are based on standard errors clustered by state.

Appendix Figure C.1.11: Estimates of the union family income premium from ANES (with and without employment status controls)



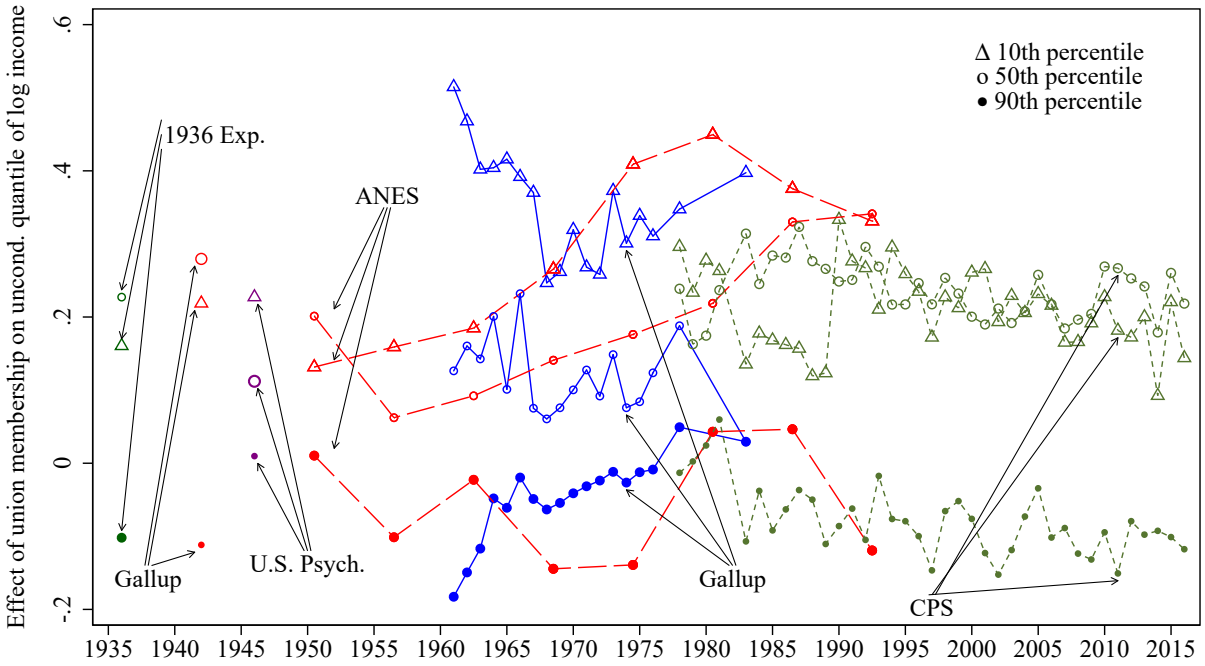
Data source: See Section 3.3 for a description of ANES data.
Notes: Each plotted point comes from estimating equation (3.3), which regresses log family income on controls for age, gender, race, state and survey-date fixed effects. In addition, the first series also includes an indicator for the household head being employed and a separate indicator for the respondent being employed. See Section 3.6.1 for more detail.

Appendix Figure C.1.12: Effects of union density on family income quantiles



Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section C.3 for a description of each data source. See Appendix C.2 for details on family income variable construction.
Notes: Each plotted point comes from estimating equation (3.5), which regresses the recentered influence function (RIF) for the specified quantile on controls for age, gender, race, educational attainment fixed effects, household employment status controls, state and survey-date fixed effects. Occupation controls are not included. The plotted confidence intervals are robust to heteroskedasticity.

Appendix Figure C.1.13: Effects of union density on family income quantiles (using *unbinned* CPS family income)



Data source: Gallup data, 1942, 1961–1974; CPS, 1978–2016; BLS Expenditure Survey, 1936; ANES, 1952–1996, U.S. Psych. Corporation, 1946. See Section C.3 for a description of each data source. See Appendix C.2 for details on family income variable construction.
Notes: Each plotted point comes from estimating equation (3.5), which regresses the recentered influence function (RIF) for the specified quantile on controls for age, gender, race, educational attainment fixed effects, household employment status controls, state and survey-date fixed effects. Occupation controls are not included. The plotted confidence intervals are robust to heteroskedasticity.

Appendix Table C.1.1: Comparing Gallup and IPUMS, 1950–1980

	1950		1960		1970		1980	
	Census	Gallup	Census	Gallup	Census	Gallup	Census	Gallup
South Share	0.242	0.117	0.259	0.138	0.271	0.247	0.296	0.256
—*South*								
Female	0.516	0.505	0.521	0.518	0.529	0.507	0.529	0.503
Age	44.61	44.31	45.07	47.64	45.94	46.35	45.20	46.13
Black	0.200	0.0849	0.182	0.147	0.160	0.129	0.159	0.160
HS grad.	0.294	0.373	0.366	0.372	0.473	0.529	0.619	0.635
—*Non-South*								
Female	0.515	0.504	0.517	0.512	0.528	0.506	0.528	0.503
Age	46.67	43.75	45.96	45.87	46.27	45.38	45.28	44.10
Black	0.0530	0.0454	0.0611	0.0586	0.0709	0.0614	0.0782	0.0874
HS grad.	0.385	0.473	0.450	0.531	0.579	0.659	0.710	0.755
Observ.	296223	182171	5388972	95064	2444218	138098	7475162	128507

Sources: Gallup surveys and 1950–1980 IPUMS.
Notes: We use the Gallup definition of the "South": all eleven states of the former Confederacy plus Oklahoma. All Census results use IPUMS person weights.

Appendix Table C.1.2: Comparing Gallup and IPUMS in 1940

	Gallup	Census	Census	Gallup	Census
–Demographics					
Black	0.0290	0.0895	0.0906	0.0325	0.0357
Female	0.338	0.505	0.344	0.341	0.343
Age	40.45	39.61	40.06	40.40	40.55
HS Graduate	0.493	0.278	0.266	0.494	0.290
College Graduate	0.0720	0.0472	0.0499	0.0709	0.0543
–Geography					
Northeast	0.0835	0.0660	0.0629	0.0946	0.0854
Mid Atlantic	0.262	0.253	0.241	0.297	0.327
East Central	0.207	0.187	0.186	0.235	0.252
West Central	0.176	0.127	0.129	0.200	0.175
South	0.118	0.258	0.263	- -	- -
Rocky Mountain	0.0751	0.0284	0.0308	0.0851	0.0418
Pacific Coast	0.0784	0.0754	0.0818	0.0888	0.111
–Occupation					
Professional	0.0780	0.113	0.122	0.0793	0.129
Farmer	0.209	0.156	0.159	0.185	0.109
Propietors, managers, officials	0.0104	0.0928	0.0875	0.0106	0.0933
Clerks (white collar)	0.294	0.0535	0.0539	0.301	0.0609
Skilled workmen and foremen	0.0906	- -	- -	0.0953	- -
Sales workers	- -	0.0462	0.0457	- -	0.0499
Craftsmen	- -	0.142	0.139	- -	0.153
Operatives	- -	0.146	0.147	- -	0.159
Unskilled or semi-skilled labor	0.190	- -	- -	0.200	- -
Laborers	- -	0.0932	0.0973	- -	0.0944
Service workers (priv. HH)	- -	0.0103	0.0105	- -	0.00626
Other service workers	- -	0.0477	0.0468	- -	0.0508
No answer, N/A, etc.	0.0826	0.0999	0.0920	0.0836	0.0949
HH/gender adjustment	N/A	N/A	Yes	N/A	Yes
Ex. S/SW?	No	No	No	Yes	Yes
Observations	See notes	736832	736832	See notes	544375

Sources: Gallup surveys and 1940 IPUMS.

Notes: The Gallup sample size varies substantially by variable during this period. For the col. (1) sample, all demographics except for education and all geographic variables have a sample size around 159,000 (with small variations due to missing observations). The occupation codes have a sample size of roughly 21,000. The high school completion indicator has a sample size of 5,700. In col. (4) each sample size is roughly twelve percent smaller. "HH / gender adjustment" underweights women and people in large households in the IPUMS, to better match Gallup sampling (which only sampled one person per household and had a target female share of one-third). "Ex S/SW" excludes Southern and Southwestern states (all eleven states of the former Confederacy plus Oklahoma). Note that occupation categories are coarser in Gallup than in the Census (but unfortunately, Gallup categories do not nest Census categories). We do our best to match occupation across these different categorizations. All Census results use IPUMS person weights.

Appendix Table C.1.3: Summary statistics from supplementary data sets

	(1) ANES	(2) BLS exp. dataset	(3) U.S. Psych. Corp.	(4) NORC
Union household	0.254	0.141	0.184	0.274
Female	0.548	0.0346	0.507	0.514
White	0.858	0.920	0.879	0.903
Age	39.67	40.85	39.11	39.84
HS graduate	0.360	0.405	0.470	0.403
South	0.277	0.232		
Log fam. inc.	9.380	7.121	3.720	7.913
Sample period	1952-1988	1936	1946	1950
Observations	30757	4058	4956	1106

Notes: See Section 3.3 for details on the data sources.

Appendix Table C.1.4: Gallup selection results through 1950, robustness to weights

	Dependent variable: Union household				
	(1)	(2)	(3)	(4)	(5)
Years of education	-0.0250***	-0.0293***	-0.0265***	-0.0267***	-0.0237***
	[0.00229]	[0.00191]	[0.00233]	[0.00240]	[0.00221]
Dept. var. mean	0.233	0.215	0.200	0.203	0.193
Weighting scheme	Baseline	None	White x Sth	Raking	Schickler
Observations	198257	208986	198257	195565	59275

Sources: See Section 3.3 and Appendix C.3 for details.
Notes: All regressions include state and survey-date fixed effects and include ages 21–64. Baseline weights are those we use throughout the paper (weights to make Gallup match interpolated Census cells for $White \times South \times Education\, categories$ (16 cells)). $White \times Sth$ are analogous, but match only on those four cells. Raking weights are constructed by matching yearly marginal mean population shares by $Black$, $Female$, and $Region$ to interpolated census shares. See [38] for more details. "Schickler weights" are taken from [117], match on $Black$ and whether a residence has as phone, and are only available through 1945. Standard errors in brackets, clustered by state. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

Appendix Table C.1.5: Is selection into unions a function of union density (using state-year variation in union density, instead of annual variation as in Table 3.1)?

	Dep't var.: At least one household member is in a union			
	(1)	(2)	(3)	(4)
Yrs. educ x State-year union density	-0.0803*** [0.0147]	-0.0839*** [0.0123]	-0.0772*** [0.0145]	-0.0812*** [0.0123]
White x State-year union density			-0.293*** [0.0586]	-0.279*** [0.0664]
Dep't var. mean	0.203	0.178	0.203	0.178
Year FE?	Yes	Yes	Yes	Yes
Drop pub. sect. HH?	No	Yes	No	Yes
Observations	1134257	998146	1134257	998146

Sources: Gallup, 1937–1976, CPS, 1977-2015.

Notes: These regressions test whether selection into union households by education and by race is a function of union density. All regressions include state and year fixed effects, and controls for age (and its square), race, gender and education. Variables that are interacted with $Year$ ($Year^2$) are divided by 1,000 (1,000,000) for ease of presentation. Standard errors are clustered by state. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.6: Estimating family union income premium and reporting coefficients on additional covariates, by data source and time period

	Dep't var: Logged family income					
	(1)	(2)	(3)	(4)	(5)	(6)
Union household	0.212***	0.255***	0.0836***	0.163***	0.150***	0.234***
	[0.0372]	[0.0348]	[0.0229]	[0.00289]	[0.0360]	[0.0189]
Years of educ., respondent		0.132***	0.108***	0.0974***	0.107***	0.111***
		[0.00549]	[0.00587]	[0.000612]	[0.00700]	[0.00465]
Years of educ., household head	0.0915***					
	[0.00766]					
White	0.632***	0.406***	0.255***	0.385***	0.394***	0.327***
	[0.0720]	[0.0618]	[0.0469]	[0.00501]	[0.0626]	[0.0381]
Respondent is female		-0.112***	-0.155***	-0.113***	-0.163***	-0.122***
		[0.0300]	[0.0254]	[0.00271]	[0.0329]	[0.0149]
Household head is female	-0.892***					
	[0.168]					
Age	0.0834***	0.0589***		0.0619***	0.0597***	0.0579***
	[0.00820]	[0.00968]		[0.000796]	[0.00765]	[0.00396]
Age squared, divided by 1,000	-0.899***	-0.671***		-0.741***	-0.679***	-0.587***
	[0.0917]	[0.120]		[0.00972]	[0.0927]	[0.0463]
Age 30-39			0.202***			
			[0.0412]			
Age 40-49			0.193***			
			[0.0295]			
Age 50-59			0.199***			
			[0.0426]			
Data source	Exp. survey	Gallup	U.S. Psych.	Gallup	ANES	ANES
Year(s) in sample	1937	1942	1946	1961-1975	1952-1970	1972-1990
Observations	4157	2524	2373	177099	2628	11777

Sources: See Section C.3 for details.

Notes: All regressions include state fixed effects. For Gallup, survey date fixed effects are included and for ANES, year fixed effects. We control for number of employed individuals in the household, except in the Gallup and U.S. Psych. data, where this control is not available. For the U.S. Psych. survey, age is given in categories, not in years, and the omitted age category in the regression is "under 30" (and we drop any observation above age 60). Otherwise, all other samples include ages 21–64. Standard errors in brackets, clustered by state. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.7: Paid vacation as a function of union status (Gallup, 1949)

	Dep't var: Do you (or husband) get paid vacation?				
	(1)	(2)	(3)	(4)	(5)
Union household	0.220***	0.183***	0.288**	0.280*	0.121***
	[0.0332]	[0.0308]	[0.126]	[0.143]	[0.0312]
White x Union household			-0.111		
			[0.127]		
Years educ. x Union household				-0.00964	
				[0.0130]	
Low-skill labor x Union					0.149***
					[0.0493]
Dept. var. mean	0.523	0.526	0.526	0.526	0.526
State FE?	Yes	Yes	Yes	Yes	Yes
Demographic controls?	Yes	Yes	Yes	Yes	Yes
Occupation FE?	No	Yes	Yes	Yes	Yes
Observations	1895	1864	1864	1864	1864

Data sources: Data from a Gallup survey in May 1949. For details on Gallup data, see Section 3.3 and Appendix C.3.

Notes: The dependent variable is a dummy. Demographic controls include respondent's age and its square, education (four fixed effects), gender and race. When occupation controls are added, they refer to the head of the household. Low-skill occupation dummy in the final column refer to the Gallup categories of "unskilled and semi-skilled labor." Standard errors are in brackets and clustered by state.

$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.8: Ease of finding a job "just as good as" the one you have, as function of union status (Gallup, 1939)

	Would be easy to find another job just as good			
	(1)	(2)	(3)	(4)
Union household	-0.122***	-0.123***	-0.0951***	-0.1000***
	[0.0278]	[0.0256]	[0.0288]	[0.0298]
Mean, dept. var.	0.499	0.499	0.499	0.495
State FE	Yes	Yes	Yes	Yes
Demogr. controls	No	Yes	Yes	Yes
Educ. controls	No	No	Yes	Yes
Occup. controls	No	No	Yes	Yes
Ex. South	No	No	No	Yes
Observations	1952	1952	1952	1686

Notes: Data from a Gallup survey in March 1939. The exact wording of the question from which we create the outcome variable is: "If you lost your present job (business, farm), how hard do you think it would be for you to get another job (business, farm) just as good?" We code the responses "impossible" and "quite hard" as zero and "fairly hard" and "easy" as one. Demographic controls include respondent's age and its square, education (four fixed effects), gender and race. When occupation controls are added, they refer to the head of the household. Low-skill occupation dummy in the final column refer to the Gallup categories of "unskilled and semi-skilled labor." Standard errors are in brackets and clustered by state.
$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.9: Aggregate college-high-school premium as a function of union density (additional specifications)

	Dep't var: College High School Premium						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Educ. Share Ratio	-0.382**	-0.287**	-0.334**	-0.307**	-0.239	-0.218	-0.391*
	(0.178)	(0.140)	(0.163)	(0.152)	(0.195)	(0.167)	(0.227)
Union Density (Gallup)		-1.507***					
		(0.481)					
Union Density (BLS)			-0.973*				
			(0.570)				
Union Density (Average)				-1.338**	-1.556**	-1.712***	-1.370**
				(0.535)	(0.720)	(0.581)	(0.575)
Mean, dept. var	0.512	0.512	0.512	0.512	0.512	0.512	0.512
Controls?	No	No	No	No	Yes	Yes	Yes
Time Polynomial?	Cubic	Cubic	Cubic	Cubic	Cubic	Quadratic	Quartic
Observations	49	49	49	49	49	49	49

Sources: The college premium was created using Census and CPS data. See Appendix C.2 for details on variable construction.

Notes: This table shows companion specifications to the regressions reported in cols. (1) and (2) in the main Table 3.2. All regressions include "time polynomial" controls, either up to quadratic, cubic or quartic level. "Controls" include the federal minimum wage, unemployment rate for civilian men, and top marginal tax rates."Union Density (Average)" uses the mean of BLS and Gallup Union Density Series. Standard errors are robust to heteroskedasticity and AR(1) serial correlation. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.10: Aggregate male 90/10 ratio as a function of union density (additional specifications)

	Dep't var: Log Percentile 90-10 Men						
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Educ. Share Ratio	-0.175**	-0.100	-0.103	-0.096	-0.084	0.156	-0.189
	(0.076)	(0.095)	(0.087)	(0.093)	(0.127)	(0.211)	(0.158)
Union Density (Gallup)		-1.185***					
		(0.416)					
Union Density (BLS)			-1.465***				
			(0.395)				
Union Density (Average)				-1.407***	-1.456**	-3.242***	-1.328**
				(0.421)	(0.636)	(0.582)	(0.566)
Mean, dept. var	1.376	1.376	1.376	1.376	1.376	1.376	1.376
Controls?	No	No	No	No	Yes	Yes	Yes
Time Polynomial?	Cubic	Cubic	Cubic	Cubic	Cubic	Quadratic	Quartic
Observations	49	49	49	49	49	49	49

Sources: The Log 90-10 Percentile for Men was created using Census and CPS data. See Appendix C.2 details on variable construction.

Notes: This table shows companion specifications to the regressions reported in cols. (3) and (4) in the main Table 3.2. All regressions include "Time Polynomial" controls, either up to quadratic, cubic or quartic level. "Controls" include the federal minimum wage, unemployment rate for civilian men, and top marginal tax rates."Union Density (Average)" uses the mean of BLS and Gallup Union Density Series. Standard errors are robust to heteroskedasticity and AR(1) serial correlation. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.11: Aggregate Gini coefficient as a function of union density (additional specifications)

	Dep't var: Gini Coefficient					
	(1)	(2)	(3)	(4)	(5)	(6)
Union Density (Gallup)	-0.083					
	(0.053)					
Union Density (BLS)		-0.120***				
		(0.040)				
Union Density (Average)			-0.132**	-0.106**	-0.216***	-0.127**
			(0.054)	(0.053)	(0.046)	(0.050)
Mean, dept. var	0.410	0.410	0.410	0.410	0.410	0.410
Educ. Control	Yes	Yes	Yes	Yes	Yes	Yes
Addit. controls?	No	No	No	Yes	Yes	Yes
Time Polynomial?	Cubic	Cubic	Cubic	Cubic	Quadratic	Quartic
Observations	65	65	65	65	65	65

Sources: The Gini Coefficient is calculated from social security data by [90].

Notes: This table shows companion specifications to the regressions reported in cols. (5) and (6) in the main Table 3.2. All regressions include "Time Polynomial" controls, either up to quadratic, cubic or quartic level. "Controls" include the federal minimum wage, unemployment rate for civilian men, and top marginal tax rates. "Union Density (Average)" uses the mean of BLS and Gallup Union Density Series. "Educ. Controls" include College-High School Share Ratio using CPS, Gallup Data and Census (interpolated) data. Standard errors are robust to heteroskedasticity and AR(1) serial correlation.

$^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.12: Aggregate top-ten income share as a function of union density (additional specifications)

	Dep't var: Top 10 Income Share					
	(1)	(2)	(3)	(4)	(5)	(6)
Union Density (Gallup)	-16.285					
	(10.664)					
Union Density (BLS)		-37.049**				
		(14.165)				
Union Density (Average)			-33.638**	-15.570	-35.340***	-13.626
			(14.413)	(12.497)	(12.292)	(14.119)
Mean, dept. var	35.848	35.848	35.848	35.848	35.848	35.848
Educ. Control	Yes	Yes	Yes	Yes	Yes	Yes
Addit. controls?	No	No	No	Yes	Yes	Yes
Time Polynomial?	Cubic	Cubic	Cubic	Cubic	Quadratic	Quartic
Observations	70	70	70	70	70	70

Sources: The Top 10 percent income share is calculated from IRS data and updated by [111].
Notes: This table shows companion specifications to the regressions reported in cols. (7) and (8) in the main Table 3.2. All regressions include "Time Polynomial" controls, either up to quadratic, cubic or quartic level. "Controls" include the federal minimum wage, unemployment rate for civilian men, and top marginal tax rates. "Union Density (Average)" uses the mean of BLS and Gallup Union Density Series. "Educ. Controls" include College-High School Share Ratio using CPS, Gallup Data and Census (interpolated) data. See Appendix for variable construction. Standard errors are robust to heteroskedasticity and AR(1) serial correlation. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.13: State-Year College High School Premium as a function of union density (additional specification)

	Dep't var: College High School Premium					
	(1)	(2)	(3)	(4)	(5)	(6)
Household union share	-0.166***	-0.442***	-0.435***	-0.510***	-0.438***	-0.434***
	0.046	0.118	0.118	0.122	0.102	0.104
Mean, dept. var.	0.490	0.490	0.490	0.497	0.497	0.497
R-squared	0.707	0.696	0.701	0.678	0.763	0.764
Education Control	Yes	Yes	Yes	Yes	Yes	Yes
Industry Shares	No	No	No	Yes	Yes	Yes
Income covars.	No	No	Yes	Yes	Yes	Yes
Policy covars.	No	No	No	No	No	Yes
Split-Sample IV	No	Yes	Yes	Yes	Yes	Yes
State-spec. quad.	No	No	No	No	Yes	Yes
Min. Year	1940	1940	1940	1940	1940	1940
Max. Year	2009	2009	2009	2009	2009	2009
Observations	1640	1640	1640	1505	1505	1505

Sources: College High School Premium was created using Census and CPS data. See Appendix C.2 details on variable construction.

Notes: This table shows companion specifications to the regressions reported in cols. (1) and (2) in the main Table 3.3. IV estimates are from split-sample-IV regressions (see Section 3.7.3 for estimating equations). All regressions include state and year fixed effects; $South \times Year$ fixed effects; and state-year education controls (both from Gallup and CPS at the annual level, and interpolated from the IPUMS Census at the decade level). "Industry shares" controls for state-year share of employment in all one-digit industry categories. "State-spec. quad." indicates that state-specific quadratic time trends are included. "Income covars." indicate that state-year GDP and state-year share of households filing taxes are included. "Policy covars." indicate that state-year minimum wage and a "policy liberalism" index (from 31) are included. Standard errors are clustered at the state level. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.14: State-Year male 90/10 ratio as a function of union density (additional specification)

	Dep't var: Log Percentile 90-10 Men					
	(1)	(2)	(3)	(4)	(5)	(6)
Household union share	-0.133**	-0.296**	-0.272**	-0.361***	-0.247***	-0.244**
	0.062	0.118	0.119	0.116	0.094	0.095
Mean, dept. var.	1.386	1.386	1.386	1.398	1.398	1.398
R-squared	0.671	0.669	0.685	0.681	0.790	0.790
Education Control	Yes	Yes	Yes	Yes	Yes	Yes
Industry Shares	No	No	No	Yes	Yes	Yes
Income covars.	No	No	Yes	Yes	Yes	Yes
Policy covars.	No	No	No	No	No	Yes
Split-Sample IV	No	Yes	Yes	Yes	Yes	Yes
State-spec. quad.	No	No	No	No	Yes	Yes
Min. Year	1940	1940	1940	1940	1940	1940
Max. Year	2009	2009	2009	2009	2009	2009
Observations	1640	1640	1640	1505	1505	1505

Sources: Log 90-10 Percentile for Men was created using Census and CPS data. See Appendix C.2 details on variable construction.

Notes: This table shows companion specifications to the regressions reported in cols. (3) and (4) in the main Table 3.3. IV estimates are from split-sample-IV regressions (see Section 3.7.3 for estimating equations). All regressions include state and year fixed effects; $South \times Year$ fixed effects; and state-year education controls (both from Gallup and CPS at the annual level, and interpolated from the IPUMS Census at the decade level). "Industry shares" controls for state-year share of employment in all one-digit industry categories. "State-spec. quad." indicates that state-specific quadratic time trends are included. "Income covars." indicate that state-year GDP and state-year share of households filing taxes are included. "Policy covars." indicate that state-year minimum wage and a "policy liberalism" index (from 31) are included.Standard errors are clustered at the state level. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.15: State-Year Gini coefficient as a function of union density (additional specification)

	Dep't var: Gini Coefficient					
	(1)	(2)	(3)	(4)	(5)	(6)
Household union share	-0.027**	-0.063***	-0.058**	-0.076***	-0.074***	-0.074***
	0.011	0.024	0.023	0.025	0.023	0.024
Mean, dept. var.	0.376	0.376	0.376	0.378	0.378	0.378
R-squared	0.712	0.710	0.720	0.718	0.783	0.783
Education Control	Yes	Yes	Yes	Yes	Yes	Yes
Industry Shares	No	No	No	Yes	Yes	Yes
Income covars.	No	No	Yes	Yes	Yes	Yes
Policy covars.	No	No	No	No	No	Yes
Split-Sample IV	No	Yes	Yes	Yes	Yes	Yes
State-spec. quad.	No	No	No	No	Yes	Yes
Min. Year	1940	1940	1940	1940	1940	1940
Max. Year	2009	2009	2009	2009	2009	2009
Observations	1640	1640	1640	1505	1505	1505

Sources: The Gini Coefficient was created using Census and CPS data. See Appendix C.2 details on variable construction.

Notes: This table shows companion specifications to the regressions reported in cols. (5) and (6) in the main Table 3.3. IV estimates are from split-sample-IV regressions (see Section 3.7.3 for estimating equations). All regressions include state and year fixed effects; $South \times Year$ fixed effects; and state-year education controls (both from Gallup and CPS at the annual level, and interpolated from the IPUMS Census at the decade level). "Industry shares" controls for state-year share of employment in all one-digit industry categories. "State-spec. quad." indicates that state-specific quadratic time trends are included. "Income covars." indicate that state-year GDP and state-year share of households filing taxes are included. "Policy covars." indicate that state-year minimum wage and a "policy liberalism" index (from 31) are included. Standard errors are clustered at the state level. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.16: State-year top-ten income share as a function of union density (additional specification)

	Dep't var: Top 10p. Income					
	(1)	(2)	(3)	(4)	(5)	(6)
Household union share	-2.124**	-5.342***	-6.158***	-3.994**	-3.087**	-3.147**
	0.960	2.065	2.103	1.727	1.410	1.399
Mean, dept. var.	36.606	36.608	36.608	36.964	36.964	36.964
R-squared	0.782	0.779	0.805	0.832	0.915	0.915
Education Control	Yes	Yes	Yes	Yes	Yes	Yes
Industry Shares	No	No	No	Yes	Yes	Yes
Income covars.	No	No	Yes	Yes	Yes	Yes
Policy covars.	No	No	No	No	No	Yes
Split-Sample IV	No	Yes	Yes	Yes	Yes	Yes
State-spec. quad.	No	No	No	No	Yes	Yes
Min. Year	1940	1940	1940	1940	1940	1940
Max. Year	2009	2009	2009	2009	2009	2009
Observations	3108	3107	3107	2723	2723	2723

Sources: The Top 10 percent share of income data comes from [65] and is discussed in the text.
Notes: This table shows companion specifications to the regressions reported in cols. (7) and (8) in the main Table 3.3. IV estimates are from split-sample-IV regressions (see Section 3.7.3 for estimating equations). All regressions include state and year fixed effects; $South \times Year$ fixed effects; and state-year education controls (both from Gallup and CPS at the annual level, and interpolated from the IPUMS Census at the decade level). "Industry shares" controls for state-year share of employment in all one-digit industry categories. "State-spec. quad." indicates that state-specific quadratic time trends are included. "Income covars." indicate that state-year GDP and state-year share of households filing taxes are included. "Policy covars." indicate that state-year minimum wage and a "policy liberalism" index (from 31) are included. Standard errors are clustered at the state level. $^{*}p < 0.1,^{**}p < 0.05,^{***}p < 0.01$

Appendix Table C.1.17: Log state-year GDP per capita as a function of union density

	Dep't var: Log GDP/Cap					
	(1)	(2)	(3)	(4)	(5)	(6)
Household union share	0.071***	0.130***	0.142***	0.140***	0.018	0.020
	0.023	0.050	0.050	0.050	0.039	0.040
Mean, dept. var.	-5.379	-5.378	-5.378	-5.322	-5.322	-5.322
R-squared	0.998	0.998	0.998	0.998	0.999	0.999
Education Control	Yes	Yes	Yes	Yes	Yes	Yes
Industry Shares	No	No	No	Yes	Yes	Yes
Share returns.	No	No	Yes	Yes	Yes	Yes
Other covars.	No	No	No	No	No	Yes
Split-Sample IV	No	Yes	Yes	Yes	Yes	Yes
State-spec. quad.	No	No	Yes	No	Yes	Yes
Min. Year	1940	1940	1940	1940	1940	1940
Max. Year	2009	2009	2009	2009	2009	2009
Observations	3108	3107	3107	2723	2723	2723

Sources: Log State GDP/Cap data comes from [65].
Notes: IV estimates are from split-sample-IV regressions (see Section 3.7.3 for estimating equations). All regressions include state and year fixed effects; $South \times Year$ fixed effects; and state-year education controls (both from Gallup and CPS at the annual level, and interpolated from the IPUMS Census at the decade level). "Industry shares" controls for state-year share of employment in all one-digit industry categories. "State-spec. quad." indicates that state-specific quadratic time trends are included. "Income covars." indicate that state-year GDP and state-year share of households filing taxes are included. "Policy covars." indicate that state-year minimum wage and a "policy liberalism" index (from 31) are included. Standard errors are clustered at the state level. $^{*}p < 0.1,^{**} p < 0.05,^{***} p < 0.01$

C.2 Sample Selection and Construction of Key Variables

C.2.1 Sample Selection

To construct our main Gallup sample, we apply the following selection criteria to the population of recorded Gallup survey respondents from years 1937 through 1987. First, we eliminate respondents to surveys in which the union membership question was not asked. Second, we remove any respondents younger than 21 or older than 64. Third, we remove respondents who identify themselves as farmers. Fourth, we remove respondents who live in Alaska, Hawaii, or Washington D.C.

Our CPS sample is taken from the May supplements in years 1976 to 1981 and both May and March supplements in years 1983 to 2015. We excluded armed forces from our sample. For state-year measures, the state identifiers are only available starting in 1979, and so our CPS-based series begins there. We excluded Alaska, DC and Hawaii from time series, additionally we excluded Idaho from the state-year analysis to make it comparable with the Gallup sample.

C.2.2 Variable Construction

Union Density In both CPS and Gallup, union density is calculated as the number of households with at least one reported union member between 21 and 64 years old divided by the total number of households. We use the CPS to construct both household union density at the aggregate time-series level as well as the state-year level.

College-High School Share Ratio To construct college-high school share ratio in both CPS and Gallup samples, we first calculate the share of respondents aged 21 to 64 years in each of three categories: high school or less, some college, and college or more. The college-

high school share ratio is calculated as the ratio of college share plus half of some college share relative to high school share plus half of some college share in each year:

$$CollegeHSShareRatio_t = \frac{ShareCollege_t + 0.5 * ShareSomeCollege_t}{ShareHighSchool_t + 0.5 * ShareSomeCollege_t} \tag{C.1}$$

Family Income Our Gallup measure of family income covers years 1942 and 1961 through 1974. Gallup family income is derived from the responses to survey questions of the following form: "Which best represents the total annual income, before taxes, of all the members of your immediate family living in your household?" Responses are coded into income bins which vary across surveys. We construct a harmonized income measure by calculating the midpoint of each interior binned response and multiplying bounds on top- and bottom-coded responses by factors of 1.25 and 0.75, respectively. Our CPS measure of family income is taken from the May and March supplements in years 1978 through 2015. This measures combines all reported income from household members 15 years and older. To construct this variable in early CPS years (May and March before 1990), we use the family income variable, which is binned into 12 categories. For the following years (CPS March only) we use the family income variable non-binned, which reports the total income for the respondent's family.

College premium, 90-10 wage ratio and Gini coefficient While our RIF analysis constructs measures of aggregate income inequality using family income variable defined above, our time series analysis calculates the college premium, 90-10 wage ratio and Gini coefficient following [74]. Each measure uses the log weekly earnings of 18 to 64 year-old full-time and full-year wage and salary workers setting unemployed and NILF to zero. Estimation controls include a full-time dummy, a female dummy, a non-white dummy, a quartic in experience and the interaction of female with non-white and the quartic in experience.

C.3 Background on Gallup Data

Gallup Sampling

Documentation for Gallup surveys prior to 1950 describe the sampling procedure as follows:

> Prior to 1950, the samples for all Gallup surveys, excluding special surveys, were a combination of what is known as a purposive design for the selection of cities, towns, and rural areas, and the quota method for the selection of individuals within such selected areas. The first step in obtaining the sample was to draw a national sample of places (cities, towns, and rural areas). These were distributed by six regions and five or six city size, urban rural groups or strata in proportion to the distribution of the population of voting age by these regional-city size strata. The distribution of cases between the non-south and south, however, was on the basis of the vote in presidential elections. Within each region the sample of such places was drawn separately for each of the larger states and for groups of smaller states. The places were selected to provide broad geographic distribution within states and at the same time in combination to be politically representative of the state or group of states in terms of three previous elections. Specifically they were selected so that in combination they matched the state vote for three previous elections within small tolerances. Great emphasis was placed on election data as a control in the era from 1935 to 1950. Within the civil divisions in the sample, respondents were selected on the basis of age, sex and socioeconomic quotas. Otherwise, interviewers were given considerable latitude within the sample areas, being permitted to draw their cases from households and from persons on the street anywhere in the community.

Beginning in 1950, Gallup's sampling procedure changed. The following excerpt is taken from post-1950 Gallup survey documentation on sampling:

All Gallup polls since 1950, excluding certain special surveys, have been based on a national probability sample of interviewing areas. Refinements in the sample design have been introduced at various points in time since then. However, over this period the design in its essentials has conformed to the current procedure, as follows:

1. The United States is divided into seven size-of-community strata: cities of population 1,000,000 and over; 250,000 to 999,999; and 50,000 to 249,999; with the urbanized areas of all these cities forming a single stratum; cities of 2,500 to 49,999; rural villages; and farm or open country rural areas.
2. Within each of these strata, the population is further divided into seven regions: New England, Middle Atlantic, East Central, West Central, South, Mountain, and Pacific Coast.
3. Within each size-of-community and regional stratum the population is arrayed in geographic order and zoned into equal-sized groups of sampling units.
4. In each zone, pairs of localities are selected with probability of selection proportional to the size of each locality's population–producing two replicated samples of localities.
5. Within selected cities for which population data are reported by census tracts or enumeration districts, these sample subdivisions are drawn with probability of selection proportional to the size of the population.
6. For other cities, minor civil divisions, and rural areas in the sample for which population data are not reported by census tracts or enumeration districts, small, definable geographic areas are drawn, with the probability of selection proportional to size where available data permit; otherwise with equal probability.

7. Within each subdivision selected for which block statistics are available, a block or block cluster is drawn with probability of selection proportional to the number of dwelling units.

8. In cities and towns for which block statistics are not available, blocks are drawn at random, that is, with equal probability.

9. In subdivisions that are rural or open country in character, segments approximately equal in size of population are delineated and drawn with equal probability.

10. In each cluster of blocks and each segment so selected, a randomly selected starting point is designated on the interviewer's map of the area. Starting at this point, interviewers are required to follow a given direction in the selection of households, taking households in sequence, until their assigned number of interviews has been completed.

11. Within each occupied dwelling unit or household reached, the interviewer asks to speak to the youngest man 18 or older at home, or if no man is at home, the oldest woman 18 or older. This method of selection within the household has been developed empirically to produce an age distribution by men and women separately which compares closely with the age distribution of the population. It increases the probability of selecting younger men, who are at home relatively infrequently, and the probability of reaching older women in the household who tend to be under-represented unless given a disproportionate chance of being drawn from among those at home. The method of selection among those at home within the household is not strictly random, but it is systematic and objective and eliminates interviewer judgement in the selection process.

12. Interviewing is conducted at times when adults are most likely to be at home, which means on weekends or if on weekdays, after 4:00 p.m. for women and after 6:00 p.m. for men.

13. Allowance for persons not at home is made by a "times-at-home" weighting procedure rather than by "call-backs." this procedure is a standard method for reducing the sample bias that would otherwise result from underrepresentation of persons who are difficult to find at home.

14. The pre-stratification by regions is routinely supplemented by fitting each obtained sample to the latest available census bureau estimates of the regional distribution of the population. Also, minor adjustments of the sample are made by educational attainment (by men and women separately), based on the annual estimates of the census bureau derived from their current population survey. The sampling procedure described is designed to produce an approximation of the adult civilian population living in the United States, except for those persons in institutions such as hospitals.

Gallup Weighting Procedure

To construct weights, we use post-stratification methods (i.e., cell-weighting). Specifically, we weight observations in the Gallup data so that the annual proportions of education-race-region cells in Gallup match the corresponding proportions in U.S. Census data. The process involves several steps: First, we construct comparable measures of education (less than high school, high school graduate, some college, college graduate), race (white, non-white), and region (South, non-South) in both Gallup and Census data. Second, we construct annual proportions of each education-race-region cell for each dataset. In the Census data, we apply representative household weights and linearly interpolate values for intercensal years to best approximate the "true" annual proportions of each cell. Third, we generate cell-specific

weights w_{ct} by applying the following formula:

$$w_{ct} = \frac{\pi_{ct}^{C}}{\pi_{ct}^{G}} \tag{C.2}$$

where c denotes a particular education-race-region cell (e.g., white Southerners with a college degree), and π_{ct}^{C} and π_{ct}^{G} denote annual cell proportions for Census and Gallup, respectively. Finally, we let $w_{it} = w_{ct}$ for each respondent i in year t corresponding to cell c in the Gallup data and re-normalize so $\sum_{i}^{N_t} \frac{w_{it}}{N_t} = 1$ for each year t.

We repeat the procedure above for several alternative cell definitions (e.g., education-race-age-state, age-gender-region). Our preferred weights use education-race-region cells because we find this definition makes our sample as representative as possible without compromising comparability across surveys or creating excessively small or "empty" cells.[1] For surveys without education data, we use race-region weights.

Comparing Gallup to Census microdata

We begin with Gallup data from 1950 onward, returning shortly to earlier data. Table C.1.1 compares Gallup data to 1950–1980 Census data. To summarize how the *actual* (unweighted) Gallup observations compare to the full U.S. adult population, we compare unweighted Gallup data to Census IPUMS tabulations. Given Gallup's well-documented under-sampling of the South, we show results separately for Southern and non-Southern states.

In 1950 and 1960, Gallup exhibits some under-sampling of the South, but, by 1970, this bias had disappeared. From 1950 to 1960, Gallup under-sampled blacks in both the South and the Non-South. This bias continued in the South through 1970, to a smaller degree. These biases reflect the substantial disenfranchisement of blacks, particularly in the South

[1] For a more thorough discussion of post-stratification weighting, including optimal cell "fineness," see [19]

during this period. Age and gender appear representative in Gallup in both regions in each decade.

Gallup respondents outside the South are more educated than their Census counterparts, with the largest gap being a high school completion difference of ten percentage points in 1950. In the South, except for 1950, Gallup and IPUMS show similar levels of education. Gallup Southern respondents have higher high school completion rates than those in the Census in 1950, not surprising as it was still under-sampling Southern blacks in that year. Later in the paper we will show results with the Gallup data both unweighted and weighted to match Census characteristics, but Table C.1.1 gives some sense of how much "work" the weights must do.

Table C.1.2 looks separately at 1940, given that Gallup's sampling procedures were quite different during its earlier years. In fact, in 1940, very few Gallup surveys ask about education (the summary statistics we present for that variable are based on only 5,767 observations), so in this table we include occupation categories as supplemental proxies for socio-economic status. The first column shows, again, unweighted Gallup data. Col. (2) presents summary statistics for all adults in the 1940 IPUMS. Perhaps the most striking discrepancy is gender: consistent with their stated methodology at the time, Gallup over-samples men. Col. (3) adjusts the Census sampling so that men are sampled at the Gallup frequencies and also down-weights large households (since Gallup only interviews one person per household). Comparing col. (1) versus (3) shows, as expected, that Gallup significantly under-samples the South.

Consistent with concerns about Gallup over-sampling the affluent, Gallup respondents in 1940 are substantially more educated than their Census counterparts. Unfortunately, given that only in 1942 does Gallup begin to regularly include an education question, the Gallup sample for which we have an education measure in 1940 is quite small (about 5,700 individuals, relative to over 150,000 for the other Gallup variables in 1940). Given the

small education sample in 1940, we use occupational categories to further explore socio-economic status in Gallup versus the 1940 Census. Gallup and IPUMS use different occupation categories—Gallup's are much coarser and unfortunately IPUMS categories do not completely nest Gallup categories—so comparisons are not straightforward. Consistent with the concerns cited earlier that Gallup over-sampled the well-to-do, Gallup respondents appear to have slightly higher-status occupations relative to their Census counterparts. For example, "professionals" and "proprietors, managers, officials" appear more numerous in Gallup (these categories are especially useful because IPUMS categories fully nest these Gallup occupations). Reassuringly, farmers and farm laborers are similarly represented in both samples (these two Gallup categories are also fully nested in IPUMS categories, again easing comparisons across data sources).

For the most part, these patterns hold when we drop Southern states from both samples (the final two columns of Table C.1.2). Importantly, outside of the South, Gallup appears to sample blacks in proportion to their population, even in the very early years of its existence. Also, outside the South, Gallup appears to accurately sample the remaining six regions of the US.[2]

In general, we show results with Gallup data using weights to match (interpolated) Census IPUMS summary statistics, even though the need for weights is not obvious after 1950 or 1960. From 1937 until 1941, we weight so that Gallup matched the IPUMS in terms of $White \times South$ cells, given that the summary statistics show that Gallup sampling along these dimensions appears suspect in the early years. Beginning in 1942 (the first year in which Gallup surveys ask the union and education questions in the same survey) we weight by $White \times Education \times South$, where $Education \in$ {No high school degree, HS degree, Some college, College graduate}, thus giving us $2 \times 4 \times$

[2]We use Gallup-defined geographic regions in this table.

$2 = 16$ cells on which to match. In practice, however, our results are very similar with and without weights.

C.4 Existing Measures of Union Density Pre-Dating the Current Population Survey

The CPS first asks respondents their union status in 1973, and then only in selected months until 1983 from which time information on union status was collected each month in the CPS as part of the outgoing rotation group supplement. Before this survey, the primary sources for union density are the BLS and Troy/NBER historical time series mentioned in the introduction. The data underlying these calculations are union reports of membership and dues revenue when available, and a variety of other sources when not available. Neither of these data sources ever used representative samples of individual workers to calculate union density.

In general, the data derived from union reports likely become more accurate by the 1960s. Post-1959 the BLS collected mandatory financial reports from unions as a condition of the Labor-Management Reporting and Disclosure (Landrum-Griffin) Act, and [123] incorporate these data into their estimates of union density. Beginning in 1964, the BLS disaggregates union membership counts by state, and [80] splice these reports together with the CPS to form state-year union density panel beginning in 1964 and continuing through today.[3]

Before the 1960s, however, union data were far less standardized. In the remainder of this section, we detail the methodology of the two most widely used data sources on aggregate union density: the BLS and Troy series.

C.4.1 The BLS estimate of early union density

The BLS series is based on union-reported membership figures starting in the late 1940s. Prior to 1948, the methodology for calculating union membership does not appear standard-

[3] [68] constructs a time-series of union density from 1880 to 1995, splicing together the official series from the BLS with series constructed from the CPS. Freeman reports alternative series constructed by other scholars ([122], [123],[125], and [71]) in the Appendix to his paper.

ized. For example, the 1945 Monthly Labor Report notes as its sources: "This study is based on an analysis of approximately 15,000 employer-union agreements as well as employment, union membership, and *other data available to the Bureau of Labor Statistics* [emphasis ours]" [24][4]

It is obviously hard to verify information from unspecified "sources available to the BLS" but even in instances where the BLS can rely on union membership reports, concerns arise. A key issue is that unions had important incentives to over-state their membership and until the late 1950s faced no penalty for doing so. In the early and mid-1930s, the main umbrella organization for local unions was the American Federation of Labor (AFL). They were often charged with over-stating their membership, presumably to inflate their political influence. For example, a 1934 *New York Times* story casts doubt on the AFL's claim to represent over six million workers, noting that "complete and authoritative data are lacking" and that the figures provided by the AFL "are not regarded as accurate."[5] Individual unions also had an incentive to inflate the numbers they reported to the AFL. For example, the number of seats each union would receive at the annual convention was based on a formula to which membership was the main input.

If anything, these incentives to over-report likely grew after 1937, when the Committee on Industrial Organization broke away from the AFL to form a rival umbrella organization, the Congress of Industrial Organizations (CIO). Both federations of labor, the AFL and CIO, now competed for local unions to join their umbrella organizations, as well as for sympathies of government officials, tasks that were aided by a public perception that the federation was

[4]For example, one alternative source the BLS used was convention representation formulas. "Convention formulas" specified the number of seats, as a function of membership, each union would have at the umbrella organization convention. Inverting this formula and using the convention records, rough estimates of union membership could be formed.

[5]See, *"Organized Labor is Put at 6,700,0000"*, New York Times, May 1935. reporting that "For one thing, complete and authoritative data are lacking, and this is especially true during times of depression, when some unions drop unemployed workers from the rolls and exempt them from paying dues. . . . . The [AFL] reported an average membership of 2,609,011 for the year ended Aug. 31, 1934. These official figures, which are not regarded as an accurate measure of the movement, are far below the peak figure of 4,078,740 for 1920."

large and growing. Based on our read of *New York Times* articles on unions in the late 1930s and early 1940s, one of the most common if not the most common topic is the conflict between the two federations.[6] Individual unions still had incentives to compete for influence within their given federation, and thus inflate membership.

Membership inflation became such an issue that the federations themselves may not have known how many actual members they had. In fact, the CIO commissioned an *internal* investigation into membership inflation, conducted by then-United Steelworkers of America president Philip Murray. Murray's 1942 report concluded that actual CIO membership was less than fifty percent of the official number the federation was reporting. [71]

C.4.2 The Troy estimates of early union density

In his NBER volumes estimating union density, Troy is well aware of the problems documented above with the BLS estimates. For this reason, he defines membership as "dues-paying members" and proceeds to estimate union membership using unions' financial reports where available, presumably under the assumption that financial reports were less biased than membership reports. For each union, he divides aggregate union dues revenue by average full-time member dues to recover an estimate of union membership. While Troy is cognizant of the limitations of his data and methodology, he believes the biases are largely *understating* union membership (e.g. some groups, such as veterans, pay lower than average or no dues).

But union financial reports, like membership reports, are also not verified until the late 1950s. Nor is it obvious that union revenue data are not similarly inflated (in fact, the AFL accused the CIO of lying about their income data, as we mention in footnote 6). Moreover, revenue data are largely incomplete for the 1930s and 1940s. For example, in his 1940 estimates, [122] notes that the sources for 54.4% of his total is *not* in fact from

[6] As just one example, a 1938 *NYT* headline and subtitles read: "Green Says Lewis Falsified Report; A.F.L. Head Alleges Statement on C.I.O. membership is an 'Amazing Inflation; Questions Income Data," referring to AFL head William Green and CIO head John Lewis, respectively.

financial reports, but instead an "Other" category, which includes personal correspondence with unions, asking their membership.[7] As such, for these early years, the Troy data in fact appears to face the same issue with membership-inflation as does the BLS data.[8]

In addition, Troy imputes the membership of many CIO unions in the late 1930s and 1940s by assigning them the membership of their AFL counterpart in the same sector.[9] This procedure likely over-states CIO membership, given that the AFL was believed to be twice as large as the CIO during this period (we also find this 2:1 ratio in our Gallup data), though obviously that average ratio may vary by sector.

In summary, while a likely improvement over the BLS series, it is difficult to believe that Troy's estimates (or [123]) are without extensive mismeasurement. Given the limitations of the existing pre-CPS data on union density, in the next section we introduce a new source: Gallup and other opinion surveys.

C.4.3 Other Pre-CPS State-Year Measures of Union Density

The only sources of state-year data on union density prior to the CPS we are aware of are measures created by [80] from BLS reports (which begin disaggregating union membership regionally, often by state, in 1964) from 1964-1977, and measures created by [123] for the years 1939 and 1956. Our Gallup measure is quite highly correlated (correlation = .724) with the existing Hirsch-Macpherson measures (individual union density as a fraction of non-farm employment) for the 1964-1986 years, which are where there is overlap. This correlation

[7] "Other" is down to 10% by 1960 ([122]).

[8] [122] also only presents validation exercises for his post-1950 data, comparing reported measurement with that inferred from dues receipts for the Chemical and Rubber Workers in 1953, leaving it open whether the BLS or Troy (or neither) is correct for the pre-1950 series.

[9] From [122] [pp. A53]: "The average membership per local industrial union is arbitrarily estimated to be 300, and this figure is multiplied each year by the number of such unions reported by the ClO. The estimate of an average membership of 300 is deemed a fair one since the average membership of the local trade and federal labor unions of the AFL, a class of unions similar to the local industrial unions of the CIO, varies from a low of 82 in 1937 to a high of 193 in 1948."

increases to .75 when we restriction attention to the CPS years with state identifiers (1978-1986).

The historical Troy measures for 1939 and 1956 are constructed from even more fragmentary records than the annual series we discuss above (as many union reports did not disaggregate either revenue or membership by state), we are also correlated with these data in both cross-sections and changes (1939 correlation = 0.78, 1953 correlation = 0.75, correlation in changes =0.5).

C.5 Details on RIF Regressions

In this appendix section we outline the recentered influence function approach of [61], adapted for our outcome measures. The relevant property of a recentered influence function is that its expectation equals the distributional statistic of interest. For quantile τ denoted Q_τ, the quantile RIF is given by $RIF(y, Q_\tau) = Q_\tau + \frac{\tau - \mathbf{1}(y_h < Q_\tau)}{f_y(Q_\tau)}$, and taking expectations verifies that $E[RIF(y, Q_\tau)] = Q_\tau$.

[61] show that a regression of the RIF on covariates yields the approximate effect of the covariates on the distributional statistic of interest (applied to the unconditional distribution). The intuition is that the marginal effect of an increase in union density to a given statistic is given by the average effect of each union individual on that statistic (each observation's "influence"). The estimate for β in equation (3.5) is therefore the effect of a change in union density on the probability that a household's income is less than the value of the quantile τ, i.e., $\frac{dF(Q_\tau)}{dUnion}$, divided by the density of household income at Q_τ ($\frac{dF(Q_\tau)}{dQ_\tau}$). The resulting coefficient thus measures $\frac{dQ_\tau}{dUnion}$ the marginal change in the value of the quantile at τ in response to a small change in $Union$. The RIF of the Gini is not particularly illustrative and we omit it here. For further examples and exposition of distributional regressions see [77] and [49].

Extending this to the 90-10 ratio immediately suggests the following RIF, and inspection confirms that $E[RIF(y_h, Q_{.9} - Q_{.1}] = Q_{.9} - Q_{.1}$.

$$RIF(y_h, Q_{.9} - Q_{.1}) = Q_{.9} - Q_{.1} + \frac{.9 - \mathbf{1}(y_h < Q_{.9})}{f_y(Q_{.9})} - \frac{.1 - \mathbf{1}(y_h < Q_{.1})}{f_y(Q_{.1})} \tag{C.3}$$

One limitation of our historical data is that the income distribution is binned, sometimes very coarsely, so the kernel density used to estimate f will be biased. We can assess this using the CPS where we have both the unbinned family income (which is still topcoded) as well as the binned version. While the exact magnitude and precision of the RIF estimates are

sensitive to the binning procedure, the qualitative pattern in the 90-10 ratio is not. Appendix Figure C.1.13 shows the percentiles with unbinned data; while the effect on the median is unchanged, the effect on the 90th and the 10th percentiles are both more positive than the binned analogue in Appendix Figure C.1.12, which roughly cancels out in the corresponding 90-10 ratio.

Bibliography

[1] John A Abowd and Thomas Lemieux. The effects of product market competition on collective bargaining agreements: The case of foreign competition in canada. *The Quarterly Journal of Economics*, 108(4):983–1014, 1993.

[2] John M Abowd. The effect of wage bargains on the stock market value of the firm. *The American Economic Review*, pages 774–800, 1989.

[3] Daron Acemoglu, Philippe Aghion, and Giovanni L Violante. Deunionization, technical change and inequality. In *Carnegie-Rochester conference series on public policy*, volume 55, pages 229–264. Elsevier, 2001.

[4] Daron Acemoglu and James A Robinson. Economics versus politics: Pitfalls of policy advice. *The Journal of Economic Perspectives*, 27(2):173–192, 2013.

[5] Ömer Tuğrul Açıkgöz and Barış Kaymak. The rising skill premium and deunionization. *Journal of Monetary Economics*, 63:37–50, 2014.

[6] Anna Aizer and Joseph J Doyle. Juvenile incarceration, human capital, and future crime: Evidence from randomly assigned judges. *The Quarterly Journal of Economics*, 130(2):759–803, 2015.

[7] George A Akerlof. The market for" lemons": Quality uncertainty and the market mechanism. *The quarterly journal of economics*, pages 488–500, 1970.

[8] Facundo Alvaredo, Anthony B Atkinson, Thomas Piketty, Emmanuel Saez, and Gabriel Zucman. The world wealth and income database. *Website: http://www. wid. world*, 2016.

[9] Joshua D Angrist and Alan B Krueger. Split-sample instrumental variables estimates of the return to schooling. *Journal of Business & Economic Statistics*, 13(2):225–235, 1995.

[10] Orley Ashenfelter. Racial discrimination and trade unionism. *Journal of political economy*, 80(3, Part 1):435–464, 1972.

[11] David H Autor and Susan N Houseman. Do temporary-help jobs improve labor market outcomes for low-skilled workers? evidence from "work first". *American Economic Journal: Applied Economics*, pages 96–128, 2010.

[12] David H Autor, Lawrence F Katz, and Melissa S Kearney. Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics*, 90(2):300–323, 2008.

[13] Robert B Avery, Paul S Calem, Glenn B Canner, and Raphael W Bostic. An overview of consumer data and credit reporting. *Fed. Res. Bull.*, 89:47, 2003.

[14] Abdurrahman Aydemir and George J Borjas. Attenuation bias in measuring the wage impact of immigration. *Journal of Labor Economics*, 29(1):69–112, 2011.

[15] Nicholas Barr, Bruce Chapman, Lorraine Dearden, and Susan Dynarski. Getting student financing right in the us: lessons from australia and england. *Centre for Global Higher Education*, 2017.

[16] Erling Barth, Alex Bryson, James C Davis, and Richard Freeman. Its where you work: Increases in the dispersion of earnings across establishments and individuals in the united states. *Journal of Labor Economics*, 34(S2):S67–S97, 2016.

[17] Efraim Benmelech, Nittai Bergman, and Hyunseob Kim. Strong employers and weak employees: How does employer concentration affect wages? Technical report, National Bureau of Economic Research, 2018.

[18] Adam J Berinsky. American public opinion in the 1930s and 1940s: The analysis of quota-controlled sample survey data. *International Journal of Public Opinion Quarterly*, 70(4):499–529, 2006.

[19] Adam J Berinsky. American public opinion in the 1930s and 1940s: The analysis of quota-controlled sample survey data. *International Journal of Public Opinion Quarterly*, 70(4):499–529, 2006.

[20] Jillian Berman. Students get tuition aid for a piece of their future. *The Wall Street Journal*, 2017.

[21] Zachary Bleemer, Meta Brown, Donghoon Lee, Katherine Strair, and Wilbert Van der Klaauw. Echoes of rising tuition in students' borrowing, educational attainment, and homeownership in post-recession america. Working Paper, 2017.

[22] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[23] Thomas C Buchmueller, John DiNardo, and Robert Valletta. A submerging labor market institution? unions and the nonwage aspects of work. In *Emerging labor market institutions for the twenty-first century*, pages 231–264. University of Chicago Press, 2004.

[24] Bureau of Labor Statistics. Extent of collective bargaining and union status, january 1945. Technical report, 1945. Bulletin # 829.

[25] Brantly Callaway and William J. Collins. Unions, workers, and wages at the peak of the american labor movement. 2016.

[26] David Card. The effect of unions on the structure of wages: A longitudinal analysis. *Econometrica*, 64(4):957–979, 1996.

[27] David Card. The effect of unions on wage inequality in the us labor market. *ILR Review*, 54(2):296–315, 2001.

[28] David Card and John E DiNardo. Skill-biased technological change and rising wage inequality: some problems and puzzles. *Journal of Labor Economics*, 20(4):733–783, 2002.

[29] David Card, Jörg Heining, and Patrick Kline. Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, 128(3):967–1015, 2013.

[30] David Card and Alan B Krueger. Does school quality matter? returns to education and the characteristics of public schools in the united states. *Journal of political Economy*, 100(1):1–40, 1992.

[31] Devin Caughey and Christopher Warshaw. The dynamics of state policy liberalism, 1936–2014. *American Journal of Political Science*, 60(4):899–913, 2016.

[32] Raj Chetty. Moral hazard versus liquidity and optimal unemployment insurance. *Journal of political Economy*, 116(2):173–234, 2008.

[33] Pierre-André Chiappori and Bernard Salanie. Testing for asymmetric information in insurance markets. *Journal of political Economy*, 108(1):56–78, 2000.

[34] Kristle Cortés, Gary Wagner, Christopher Vecchio, and Guhan Venkatu. Is there a student loan crisis? not in payments. *Policy*, 6:17, 2017.

[35] Gordon B Dahl, Andreas Ravndal Kostøl, and Magne Mogstad. Family welfare cultures. *The Quarterly Journal of Economics*, 129(4):1711–1752, 2014.

[36] Department of Education. Federal student loan portfolio. Website, 2017. Accessed: 2017-10-19.

[37] Department of Education. New student loan report reveals promising repayment trends. Website, 2017. Accessed: 2017-10-19.

[38] Jean-Claude Deville, Carl-Erik Särndal, and Olivier Sautory. Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, 88(423):1013–1020, 1993.

[39] Marco Di Maggio, Amir Kermani, Benjamin Keys, Tomasz Piskorski, Rodney Ramcharan, Amit Seru, and Vincent Yao. Monetary policy pass-through: Mortgage rates, household consumption and voluntary deleveraging. *American Economic Review*, 2016.

[40] John DiNardo, Nicole M Fortin, and Thomas Lemieux. Labor market institutions and the distribution of wages, 1973-1992: A semiparametric approach. *Econometrica*, 64(5):1001–1044, 1996.

[41] John Dinardo and Kevin F Hallock. When unions mattered: the impact of strikes on financial markets, 1925–1937. *ILR Review*, 55(2):219–233, 2002.

[42] John DiNardo and David S Lee. Economic impacts of new unionization on private sector employers: 1984–2001. *The Quarterly Journal of Economics*, 119(4):1383–1441, 2004.

[43] Emin M Dinlersoz and Jeremy Greenwood. The rise and fall of unions in the us. Technical report, National Bureau of Economic Research, 2012.

[44] Will Dobbie, Jacob Goldin, and Crystal Yang. The effects of pre-trial detention on conviction, future crime, and employment: Evidence from randomly assigned judges. 2016. National Bureau of Economic Research Working Paper.

[45] Will Dobbie, Paul Goldsmith-Pinkham, Neale Mahoney, and Jae Song. Bad credit, no problem? credit and labor market consequences of bad credit reports. National Bureau of Economic Research Working Paper 22711, 2016.

[46] Will Dobbie, Paul Goldsmith-Pinkham, and Crystal Yang. Consumer bankruptcy and financial health. *The Review of Economics and Statistics*, forthcoming.

[47] Will Dobbie and Jae Song. Debt relief and debtor outcomes: Measuring the effects of consumer bankruptcy protection. *The American Economic Review*, 105(3):1272–1311, 2015.

[48] Danielle Douglas-Gabriel. Trump just laid out a pretty radical student debt plan. *The Washington Post*, 2015.

[49] Arindrajit Dube. Minimum wages and the distribution of family incomes. 2017.

[50] William Dudley. Opening remarks at the convening on student loan data conference. 2015.

[51] Susan Dynarski. An economist's perspective on student loans in the united states. Working paper, Education Policy Initiative: Gerald R. Ford School of Public Policy, September 2014.

[52] Susan Dynarski. We're frighteningly in the dark about student debt. *The New York Times*, 2015.

[53] Susan Dynarski and Daniel Kreisman. Loans for educational opportunity: Making borrowing work for today's students. *The Hamilton Project. Discussion Paper*, 2013.

[54] Liran Einav, Amy Finkelstein, and Mark R Cullen. Estimating welfare in insurance markets using variation in prices. *The quarterly journal of economics*, 125(3):877–921, 2010.

[55] Liran Einav, Amy Finkelstein, and Paul Schrimpf. Optimal mandates and the welfare cost of asymmetric information: Evidence from the uk annuity market. *Econometrica*, 78(3):1031–1092, 2010.

[56] Henry Farber. The determination of the union status of workers. *Econometrica*, 51, 1983.

[57] Henry Farber. Nonunion wage rates and the threat of unionization. *ILR Review*, 58(3):335–352, 2005.

[58] James Feigenbaum, Alexander Hertel-Fernandez, and Vanessa Williamson. From the bargaining table to the ballot box: political effects of right to work laws. Technical report, National Bureau of Economic Research, 2018.

[59] Amy Finkelstein and Kathleen McGarry. Multiple dimensions of private information: evidence from the long-term care insurance market. *The American economic review*, 96(4):938–958, 2006.

[60] Amy Finkelstein, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Joseph P Newhouse, Heidi Allen, Katherine Baicker, and Oregon Health Study Group. The oregon health insurance experiment: evidence from the first year. *The Quarterly journal of economics*, 127(3):1057–1106, 2012.

[61] Sergio Firpo, Nicole M Fortin, and Thomas Lemieux. Unconditional quantile regressions. *Econometrica*, 77(3):953–973, 2009.

[62] Philip Sheldon Foner. *Organized labor and the black worker: 1619-1973*, volume 475. International publishers, 1976.

[63] Brigham Frandsen. The surprising impacts of unionization on establishments: Accounting for selection in close union representation elections. *Manuscript. Department of Economics, Brigham Young University*, 2013.

[64] Brigham R Frandsen. Party bias in union representation elections: Testing for manipulation in the regression discontinuity design when the running variable is discrete. In *Regression Discontinuity Designs: Theory and Applications*, pages 281–315. Emerald Publishing Limited, 2017.

[65] Mark Frank. Frank-sommeiller-price series for top income shares by us states since 1917. *WTID Methodological Notes*, 2015.

[66] Richard Freeman and James Medoff. Medoff. what do unions do, 1984.

[67] Richard B. Freeman. The effect of unionism on fringe benefits. *Industrial and Labor Relations Review*, 34(4):489–509, 1981.

[68] Richard B Freeman et al. Spurts in union growth: Defining moments and social processes. *NBER Chapters*, pages 265–296, 1998.

[69] Milton Friedman. *Capitalism and freedom*. University of Chicago press, 1962.

[70] Carola Frydman and Raven E Saks. Executive compensation: A new view from a long-term perspective, 1936–2005. *Review of Financial Studies*, page 120, 2010.

[71] Walter Galenson. *The CIO challenge to the AFL: a history of the American labor movement, 1935-1941*. Harvard University Press, 1960.

[72] Peter Ganong and Pascal Noel. The effect of debt on default and consumption: Evidence from housing policy in the great recession. Working Paper, 2017.

[73] Claudia Goldin and Robert Margo. The great compression: The wage structure in the united states at mid-century. *The Quarterly Journal of Economics*, 107:1–34, February 1992.

[74] Claudia Dale Goldin and Lawrence F Katz. *The race between education and technology*. Harvard University Press, 2009.

[75] David B Gross and Nicholas S Souleles. Do liquidity constraints and interest rates matter for consumer behavior? evidence from credit card data. *The Quarterly journal of economics*, 117(1):149–185, 2002.

[76] Justine Hastings, Christopher A. Neilson, and Seth D. Zimmerman. The effects of earnings disclosure on college enrollment decisions. June 2015. National Bureau of Economic Research Working Paper 21300.

[77] Tarjei Havnes and Magne Mogstad. Is universal child care leveling the playing field? *Journal of Public Economics*, 127:100–114, 2015.

[78] Nathaniel Hendren. Private information and insurance rejections. *Econometrica*, 81(5):1713–1762, 2013.

[79] Nathaniel Hendren. Knowledge of future job loss and implications for unemployment insurance. Technical report, National Bureau of Economic Research, 2016.

[80] Barry T Hirsch, David A Macpherson, and Wayne G Vroman. Estimates of union density by state. *Monthly Labor Review*, 124(7):51–55, 2001.

[81] Boris Hirsch and Steffen Müller. Firm wage premia, industrial relations, and rent sharing in germany. 2018.

[82] Sarah Elizabeth Igo. *The averaged American: Surveys, citizens, and the making of a mass public*. Harvard University Press, 2007.

[83] Atsushi Inoue and Gary Solon. Two-sample instrumental variables estimators. *The Review of Economics and Statistics*, 92(3):557–561, 2010.

[84] Florence Jaumotte and Carolina Osorio. *Inequality and labor market institutions*. International Monetary Fund, 2015.

[85] Lawrence F Katz and Kevin M Murphy. Changes in relative wages, 1963–1987: supply and demand factors. *The Quarterly Journal of Economics*, 107(1):35–78, 1992.

[86] Ira Katznelson. *Fear itself: The new deal and the origins of our time*. WW Norton & Company, 2013.

[87] Benjamin J Keys, Tomasz Piskorski, Amit Seru, and Vincent Yao. Mortgage rates, household balance sheets, and the real economy. National Bureau of Economic Research Working Paper 20561, 2014.

[88] Martin Luther King Jr. If the negro wins, labor wins. *A Testament of Hope: The Essential Writings of Martin Luther King, Jr*, pages 201–207, 1986.

[89] Jeffrey R Kling. Incarceration length, employment, and earnings. *The American Economic Review*, 96(3):863–876, 2006.

[90] Wojciech Kopczuk, Emmanuel Saez, and Jae Song. Earnings inequality and mobility in the united states: evidence from social security data since 1937. *The Quarterly Journal of Economics*, 125(1):91–128, 2010.

[91] Andreas Ravndal Kostøl, Magne Mogstad, Bradley Setzler, et al. Disability benefits, consumption insurance, and household labor supply. National Bureau of Economic Research Working Paper, 2017.

[92] David S Lee. Wage inequality in the united states during the 1980s: Rising dispersion or falling minimum wage? *The Quarterly Journal of Economics*, 114(3):977–1023, 1999.

[93] David S Lee and Alexandre Mas. Long-run impacts of unions on firms: New evidence from financial markets, 1961–1999. *The Quarterly Journal of Economics*, 127(1):333–378, 2012.

[94] Jan E Leighley and Jonathan Nagler. Unions, voter turnout, and class bias in the us electorate, 1964–2004. *Journal of Politics*, 69(2):430–441, 2007.

[95] Thomas Lemieux. Estimating the effects of unions on wage inequality in a panel data model with comparative advantage and nonrandom selection. *Journal of Labor Economics*, 16(2):261–291, 1998.

[96] Thomas Lemieux. Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill? *The American Economic Review*, 96(3):461–498, 2006.

[97] H Gregg Lewis. *Unionism and Relative Wages in the United States*. University of Chicago Press, 1963.

[98] H Gregg Lewis. Union relative wage effects. *Handbook of labor economics*, 2:1139–1181, 1986.

[99] Nelson Lichtenstein. *Labor's War at Home: The CIO in World War II*. Temple University Press, 2003.

[100] Lars Ljungqvist and Thomas J Sargent. *Recursive macroeconomic theory*. MIT press, 2012.

[101] Adam Looney and Constantine Yannelis. A crisis in student loans?: How changes in the characteristics of borrowers and in the institutions they attended contributed to rising loan defaults. *Brookings Papers on Economic Activity*, 2015(2):1–89, 2015.

[102] Deborah Lucas and Damien Moore. Costs and policy options for federal student loan programs. a cbo study. *Congressional Budget Office*, 2010.

[103] Alan Manning. Imperfect competition in the labor market. In *Handbook of labor economics*, volume 4, pages 973–1041. 2011.

[104] Benjamin M Marx and Lesley J Turner. Student loan nudges: Experimental evidence on borrowing and educational attainment. National Bureau of Economic Research Working Paper 24060, November 2017.

[105] Nicolai Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7(Jun):983–999, 2006.

[106] Alvaro Mezza, Daniel R Ringo, Shane Sherland, and Kamila Sommer. On the effect of student loans on access to homeownership. Working paper, 2016.

[107] John Stuart Mill. *Principles of Political Economy*. Oxford University Press, 2008.

[108] Sendhil Mullainathan and Jann Spiess. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2):87–106, 2017.

[109] Suresh Naidu, Eric A Posner, and E Glen Weyl. Antitrust remedies for labor market power. *Harvard Law Review*, 2018.

[110] Herbert Roof Northrup. *Organized labor and the Negro*. Periodicals Service Co, 1971.

[111] Thomas Piketty and Emmanuel Saez. Income inequality in the united states, 1913–1998. *The Quarterly journal of economics*, 118(1):1–41, 2003.

[112] Thomas Piketty, Emmanuel Saez, and Stefanie Stantcheva. Optimal taxation of top labor incomes: A tale of three elasticities. *American economic journal: economic policy*, 6(1):230–271, 2014.

[113] Jörn-Steffen Pischke, John E DiNardo, and Kevin F Hallock. Unions and the labor market for managers. 2000.

[114] Albert Rees. The effects of unions on resource allocation. *The Journal of Law and Economics*, 6:69–78, 1963.

[115] Michael Rothschild and Joseph Stiglitz. Equilibrium in competitive insurance markets: An essay on the economics of imperfect information. *The quarterly journal of economics*, pages 629–649, 1976.

[116] Jesse Rothstein and Cecilia Elena Rouse. Constrained after college: Student loans and early-career occupational choices. *Journal of Public Economics*, 95(1):149–163, 2011.

[117] Eric Schickler and Devin Caughey. Public opinion, organized labor, and the limits of new deal liberalism, 1936–1945. *Studies in American Political Development*, 25(02):162–189, 2011.

[118] Robert Shimer and Iván Werning. Liquidity and insurance for the unemployed. *The American Economic Review*, 98(5):1922–1942, 2008.

[119] Jae Song, David J Price, Fatih Guvenen, Nicholas Bloom, and Till Von Wachter. Firming up inequality. Technical report, National Bureau of Economic Research, 2015.

[120] Cong Tang, Keith Ross, Nitesh Saxena, and Ruichuan Chen. What's in a name: a study of names, gender inference, and gender behavior in facebook. *Database Systems for Adanced Applications*, pages 344–356, 2011.

[121] Mathieu Taschereau-Dumouchel. The union threat. 2015.

[122] Leo Troy. *Trade Union Membership, 1897–1962*. NBER, 1965.

[123] Leo Troy and Neil Sheflin. Union sourcebook: Membership, finances, structure, directory. *West Orange, NJ: Industrial Relations Data and Information Services*, 1985.

[124] Bruce Western and Jake Rosenfeld. Unions, norms, and the rise in us wage inequality. *American Sociological Review*, 76(4):513–537, 2011.

[125] Leo Wolman. *The Growth of American Trade Unions, 1880-1923*. NBER, 1924.

[126] Constantine Yannelis. Asymmetric information in student loans. Working paper, 2016.