Three Essays on Worker-Firm Dynamics

by

Dhiren Patki

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in The University of Michigan
2020

Doctoral Committee:

Professor Matthew D. Shapiro, Chair
Professor Charles Brown
Professor Dana Muir
Professor Melvin Stephens Jr.

Dhiren Patki
dpatki@umich.edu
ORCID ID 0000-0002-7661-141X

ACKNOWLEDGEMENTS

When I first arrived at the University of Michigan in the fall of 2013, I did not realize how challenging and rewarding the next seven years were going to be. What I appreciated least was that research would entail incessant failures interspersed with only the briefest of successes. Although I was never able to smooth out this see-saw pattern of emotions, I do know with certainty that undertaking this dissertation would have been a futile effort were it not for the support of a phenomenal group of faculty mentors, my friends, and my family.

I am grateful to my dissertation committee for their excellent advice, their encouragement, and their patience. Having Matthew Shapiro as my dissertation chair has been a privilege. He has been generous with his time, incredibly incisive with his comments, and has taught me to think rigorously about research in economics. His initiative in arranging data access through the CenHRS project has been instrumental for chapters I and III. Charlie Brown and Mel Stephens have been wonderful mentors from whom I have learned a great deal. Listening to them ask questions in seminars is what first made me want to become a labor economist and I continually find myself wondering how they might react to a new theoretical conjecture or to a new empirical fact. I am thankful to Dana Muir for sharing her incredible legal and institutional knowledge about pensions which greatly improved chapter I.

I thank John Bound, Gábor Kézdi, Maggie Levenstein, Trivellore Raghunathan, and Jeff Smith who, although not formally on my dissertation committee, have provided invaluable guidance and support. I am particularly grateful to John for his excellent advice on research, the job market, and life beyond it. Chichun Fang provided a wealth of information about the HRS pension estimation program which has been a critical input for chapter I. Clint Carter and Kristin McCue have facilitated requests to disclose estimates from the U.S. Census Bureau for chapters I and III. I must add the usual boilerplate disclaimer: Any opinions and conclusions in those chapters are expressed by the authors and do not necessarily represent the views of the U.S. Census Bureau. All results have been reviewed to ensure that no confidential information is disclosed. The disclosure review request numbers are 8002 and 8279 for chapter I and 6961 and 7357 for chapter III.

The graduate student community in Ann Arbor have given me a tremendous group of colleagues and friends. Parag Mahajan, who is a co-author on chapter II, has been a wonderful collaborator who brings precision and clarity to any research question. For their friendship, advice, kindness, and support, I thank Ari Binder, Javed Basu-Kesselman, Blanche Blumenthal, Lia Delaney, Pieter De Vlieger, George Fenton, Jamie Fogel, Mike Gelman, Julie Gilbert, Matt Gross, Max Gross, Xing Guo, Sam Haltenhof, Gretchen Lay, Meera Mahadevan, Chad Milando, Stephanie Owen, Ellen Stuart, and Brenden Timpe.

My family have gamely put up with my long absences by cheerfully keeping up with me through FaceTime and only the briefest of visits home. I am deeply grateful to them for their love and support. They have served as sounding boards and provided life advice when I needed it the most. With graduate school behind me, I hope we will see each other more often.

This dissertation has been supported by the CenHRS project grant to the University of Michigan from the Alfred P. Sloan Foundation, the Michigan Node of the NSF-Census Research Network (NCRN) under National Science Foundation Grant SES-1131500, the Alfred P. Sloan Foundation Pre-doctoral Fellowship on the Economics of an Aging Workforce, awarded through the National Bureau of Economic Research (NBER), the F. Thomas Juster Economic Behavior Research Award, the John E. Parker Memorial Prize in Labor Economics, and through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. Chapter I uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants: SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation. Chapter II uses the weakly anonymous Sample of Integrated Labor Market Biographies (Years 1975-2010). Data access was provided via on-site use at the Research Data Center (FDZ) of the German Federal Employment Agency (BA) at the Institute for Employment Research (IAB) and subsequently remote data access under project number fdz1248.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

<u>Table</u>

LIST OF APPENDICES

Appendix

ABSTRACT

This dissertation uses three different sources of matched employer-employee (MEE) data to study how firm-level heterogeneity influences workers' labor market outcomes. Chapters I and II deal with implicit contracts or long-term unenforceable agreements between firms and workers that govern outcomes such as the split between wage and non-wage compensation, the duration of employment relationships, worker productivity, and risk sharing. Chapter III considers the statistical challenge of creating a MEE dataset by linking a household-level survey with establishment-level administrative data in the absence of unique identifiers.

In chapter I, I study the determinants of retirement behavior by exploiting the recent and widespread elimination of traditional pensions and subsequent adoption of 401(k) plans by U.S. employers. Using thousands of firm-level natural experiments, I show that unexpected losses in future compensation engendered by pension plan transitions induce a 1 percentage point increase in retirement on impact. Affected workers who do not retire immediately lengthen their careers and exhibit a 2 percentage point reduction in retirement 10 years after the pension plan transition. Observed heterogeneity in retirement behavior is indicative of differences in wealth and in preferences for leisure. Using these treatment effects as estimation targets, I fit a structural model of retirement and saving in which workers with different levels of wealth and different preferences for leisure exhibit heterogenous behavior when faced with a pension plan transition. I use simulations from the model to show that a counterfactual policy that eliminates Social Security payroll taxes for workers over age 60 increases the average retirement age by one year and provides substantial welfare gains to older workers.

In chapter II, which is co-authored with Parag Mahajan, we use MEE data from Germany to show that cohorts entering the labor market during a recession experience a 4.9 percent reduction in wages cumulated over the first decade of labor market experience. While 40 percent of the recession-induced wage loss is explained by workers matching with lower paying firms, we use a revealed preference-based algorithm to show that fully three-fourths of the losses in employer-specific pay are compensated for by non-pay amenities. The higher non-pay amenities that we associate with recessionary labor market entrants are consistent with the view that employers that hire during business cycle downturns exhibit less cyclically sensitive labor demand and provide greater long-term job security. Our findings indicate

that the welfare cost of labor market entry during recessions is less severe than pecuniary estimates would imply.

In chapter III, which is co-authored with John Abowd, Joelle Abramowitz, Margaret Levenstein, Kristin McCue, Trivellore Raghunathan, Ann Rodgers, Matthew Shapiro, and Nada Wasi, we illustrate an application of record linkage between a household-level survey and an establishment-level administrative dataset in the absence of unique identifiers. Record linkage in this setting is challenging because the distribution of employment across firms is highly asymmetric. To address these difficulties, we use a supervised machine learning model to probabilistically link survey respondents in the Health and Retirement Study (HRS) with employers and establishments in the Census Business Register (BR) to create a new data source which we call the CenHRS. We use multiple imputation to propagate uncertainty from the linkage step into subsequent analyses of the linked data. The linked data reveal new evidence that survey respondents' misreporting and selective nonresponse about employer characteristics are systematically correlated with wages.

CHAPTER I

Breaking the Implicit Contract: Using Pension Freezes to Study Lifetime Labor Supply

1.1 Introduction

The aging of the baby boom generation poses stark challenges that affects not only the financial future of retirees but also the overall performance of the economy. As a large share of the population permanently withdraws from the labor force, economic growth will slow, more individuals will claim Social Security and Medicare benefits, and fewer workers will pay into these social insurance programs. To mitigate some of these adverse effects, researchers have proposed policies that incentivize delayed retirement by altering the tax code or by changing the structure of public and private pension benefits.

Predicting retirement responses to policy-induced changes in compensation structure is subject to two empirical challenges. The first challenge relates to defining the horizon that workers consider when making retirement decisions. Several studies including Krueger and Pischke (1992), Rogerson and Wallenius (2013), Brown (2013), and Manoli and Weber (2016) model retirement in the context of contemporaneous changes in compensation. In contrast, Stock and Wise (1990) emphasize that the retirement decision is determined not only by contemporaneous compensation but also by the entire path of expected future compensation.[1] The role of long horizons in determining retirement is important and has been validated empirically using both reduced-form and structural methods (see, e.g., Coile and Gruber (2007) and references therein). Nevertheless, a limitation of studies that do adopt a forward-looking view on retirement behavior is that they have been identified either using cross sectional variation or using panel fixed effect designs. These empirical approaches

[1] Implicit contracts that solve agency problems are a powerful force connecting current labor supply decisions to future compensation. See Lazear (1979), Lazear (1981) and Akerlof and Katz (1989) for important theoretical contributions.

can be problematic because within- and between-person differences in future compensation are likely correlated with unobserved determinants of lifetime labor supply. Closely related to this concern, the second challenge stems from the fact that unanticipated changes in compensation generate both wealth and substitution effects, and the identification of the structural parameters governing these effects is difficult when relying on panel data alone (see, e.g., MaCurdy (1981)).

In this paper, I address both challenges. I use a novel source of variation to identify a model of retirement behavior in which workers' retirement decisions depend on both current and expected future compensation. My identification strategy relies on the large-scale restructuring of U.S. private sector defined benefit (DB) pension plans which began in the early 2000's. Since then, many employers have reneged on long-standing promises to their workers by eliminating generous DB pension accruals and replacing them with less lucrative defined contribution (DC) and cash balance (CB) pension plans.[2] These actions, known as pension freezes, allow workers to keep previously earned pension benefits but unexpectedly change the present value of future compensation that workers have been promised but are yet to be paid.[3] From a research perspective, pension freezes are useful because they generate unanticipated shifts in workers' age-compensation profiles which induce both wealth and substitution effects.[4] As I describe below, this source of variation credibly identifies preference parameters governing retirement behavior thereby providing a better way to predict how older workers will react to unexpected policy-induced changes in compensation structure.

To study why employers have increasingly restructured their DB plans and to examine how employees respond to these changes, I create a new dataset that matches information from Internal Revenue Service (IRS) administrative records on the universe of private sector pension plans with longitudinal employer-employee linked data from the Census Bureau. My research design pools together thousands of firm-level natural experiments and compares the labor supply behavior of workers whose employers freeze their DB plans with workers whose

[2]401(k) plans are the most common type of DC plan. Unlike traditional DB pensions which guarantee workers an income stream in retirement, DC plans allow workers to accumulate retirement savings in tax-preferred accounts. Employers typically provide incentives for participation in these plans by matching worker contributions. CB plans are functionally similar to DC plans. More details are provided in Section 1.2.

[3]The abrupt nature of these changes are reflected in the sentiment of one disgruntled Verizon employee who, after hearing of the company's decision to freeze his DB pension, said "Oh, it's outrageous that they ... *change the rules in the middle of the game.*" See "Verizon Unveils Major Changes to Retirement Benefits," National Public Radio, *All Things Considered*, December 6, 2005.

[4]Krueger and Pischke (1992), Brown (2013), and Gelber et al. (2016) also rely on unanticipated changes in pension compensation to study retirement behavior. These studies do not use variation in future compensation to identify a structural model or estimate the effect of counterfactual policies.

employers keep their DB plans intact.

On the firm side, I provide evidence that pension freezes are driven primarily by plan funding deficiencies and are unrelated to mass layoffs or employee age structure. On the worker side, I show that the retirement response to these shocks varies in two important dimensions. First, workers initially affected at different ages exhibit differing retirement responses because they experience compensation losses of varying magnitude. Second, holding age fixed, some workers respond to freezes by retiring early while others respond by retiring later. Early retirements reflect the dominance of substitution effects while delayed retirements reflect the dominance of wealth effects. Workers' dynamic response to pension freezes is therefore indicative of heterogeneity in wealth and/or preferences for leisure.

The administrative data that I use describe employment responses to pension freezes. These data do not include information on the dollar value of compensation changes induced by pension freezes. To estimate the pecuniary effects of pension freezes and to conduct counterfactual simulations, I develop and estimate a structural model of retirement timing and saving that relies on rich survey data from the Health and Retirement Study (HRS). The model incorporates heterogeneity in wealth and in preferences for leisure and allows work decisions to depend not only on current compensation but also on the option of higher earnings and increased pension benefits in the future as in Stock and Wise (1990). Unlike Stock and Wise (1990), however, the model allows for saving in multiple asset types, includes Social Security benefits, and features non-linear taxes.[5] I fit the model using the method of simulated moments (MSM) by matching quasi-experimental employment responses to pension freezes as observed in administrative data.

The model that I estimate highlights the importance of forward-looking behavior in governing the retirement decision and lends itself to evaluating the impact of counterfactual policies aimed at incentivizing delayed retirement through a permanent shift in future compensation trajectories. I examine one such proposal through the lens of the model. In particular, I consider eliminating the Social Security or Old Age and Survivors Insurance (OASI) component of the payroll tax for workers over the age of 60. This reform has been proposed as a way of lengthening lifetime labor supply by removing contribution requirements for workers who are fully vested in Social Security benefits (see, e.g., Burtless and Quinn (2002) and Goda et al. (2009)). Simulations from the estimated model predict that an unexpected elimination of the OASI payroll tax at age 60 causes employment rates to rise by an average of approximately 5 percentage points over a 20 year horizon which equates to a

[5]The model I use is a stochastic dynamic programming model of retirement and not an option value model as pioneered by Stock and Wise (1990). Lumsdaine et al. (1992) provide a comparison of the two approaches.

1.1 year delay in the average retirement age. Equivalent variation from the reform averages \$75,000 per worker.[6]

In this paper, I make three contributions to the literature. First, I provide the only available evidence of how older Americans have re-timed their retirement decisions in response to widespread DB pension freezes. These analyses, based on high-quality administrative data, shed new light on the ongoing transition from DB to DC pension provision. While prior studies examining the incentive effects of DB and DC pension plans have emphasized that DC plans encourage longer careers (see, e.g., Friedberg and Webb (2005)), the findings I present in this paper show that unexpected transitions from DB to DC plans generate a mixed response: some workers choose to shorten their careers whereas others choose to lengthen them. Second, I estimate a model of retirement and saving which is identified by plausibly exogenous variation in future compensation arising from pension freezes. This natural experiment-based identification strategy represents a departure from prior structural models of retirement that rely on cross sectional or panel data to isolate variation in compensation (see, e.g., Stock and Wise (1990), Berkovic and Stern (1991), Rust and Phelan (1997), French (2005), and French and Jones (2011)). Finally, I provide evidence about the effectiveness of a counterfactual reform to payroll taxes designed to lengthen careers. In contrast to prior efforts evaluating the effect of payroll tax sunsets (Laitner and Silverman (2012) and Gustman and Steinmeier (2015)), the model and identification strategy employed in this paper explicitly account for the long-term option value channel of continued employment.

The remainder of this paper is structured as follows. Section 1.2 provides institutional details on the DB pension landscape in the United States and explains important recent changes influencing firms' decisions to freeze their plans. Section 1.3 outlines a model of retirement timing and saving and uses the model to evaluate how workers respond to pension freezes. Section 1.4 discusses three different sources of administrative data used in the analyses. Section 1.5 outlines the empirical framework and presents summary statistics. Section 1.6 provides empirical evidence on workers' labor supply responses to pension freezes. Section 1.7 explains the identification and estimation of the structural model and presents parameter estimates. Section 1.8 evaluates the counterfactual OASI payroll tax sunset. Section 1.9 concludes.

[6]Estimated in 2010 dollars with present values calculated at age 60.

1.2 Why are employers freezing DB plans?

In 1980 DB plans covered 61 percent of pension eligible private sector workers. By 2015 the same statistic had fallen to 16 percent.[7] This overall decline occurred not only because of a surge in new 401(k) DC plans but also because of stagnation in the creation of new DB plans. For most of this period, DB plans continued to operate normally with only rare instances of distressed plan terminations triggered by firm bankruptcy.[8] Starting in the late 1990's, however, several prominent firms began converting their DB plans to CB plans. CB conversions switched pension accruals away from formulas that were based on years of service and earnings, to account based plans that provided participants employer contributions that were proportional to earnings and a market linked rate of return on previous contributions.[9] In the early 2000's this shift was amplified as many employers altogether ended traditional DB pension accruals in actions known as hard freezes.[10] The incidence of CB conversions and hard freezes between 1999 and 2015 is shown in Figure 1.1. By 2015, about half of all private sector single employer DB plans had either been converted to CB plans or hard frozen, affecting about 40 percent of active participants or 4.1 million workers.[11]

1.2.1 Costs of DB plan provision have become increasingly volatile

Firm's decisions to renege on DB promises through CB conversions and hard freezes occurred in a deteriorating financial environment that increased the volatility of DB pension costs. As shown in Figure 1.2, the wake of the dot-com bubble and ensuing recession lowered the asset value of pension funds substantially. In 2000, the private sector DB system had \$1.44 in assets for each dollar of future liabilities. By 2004, the funding ratio had slipped

[7]See Table E7, Private Pension Plan Bulletin Historical Tables and Graphs, U.S. Department of Labor, 2018.

[8]The Omnibus Budget Reconciliation Act of 1987 reduced the ability of DB sponsoring employers to take tax deductions for pension contributions. This change led to a spike in non-distress terminations between 1987 and 1990 (see, e.g., Table A-9 in Pension Insurance Data Book 1996, PBGC Single Employer Program `https://www.pbgc.gov/documents/1996databook.pdf`).

[9]CB plans are functionally very similar to DC plans: they provide participants with individual accounts whose value is tied to earnings levels. There are, however, two differences. First, CB plans allow for larger maximum pre-tax deferrals than DC plans. Second, it is common in CB plans for employers to bear some interest rate risk by promising a minimum rate of return on the value of the account. For these reasons, CB plans are subject to the same funding obligations required of DB plans and are legally treated as DB plans.

[10]Over the same period, employers also began closing their DB plans to new entrants in actions known as soft-freezes. I do not consider soft-freezes in this paper.

[11]Hard freezes are exceedingly rare in the public sector affecting less than 0.5 percent of eligible state and local government workers. The main reason for this difference owes to differing legal protections in the private and public sectors. The Employee Retirement Income Security Act of 1974 (ERISA), which governs private pensions, explicitly allows sponsors to modify future pension rules provided that earned pension benefits are not reduced. In the public sector, which is not subject to the provisions of ERISA, several state governments protect both earned and unearned future pension benefits. See, e.g., Monahan (2012).

to 0.85. As the funding position of DB plans worsened, statutory provisions required firms to increase annual contributions; consequently, aggregate payments into DB pension funds rose five-fold from $26 billion in 2000 to $124 billion in 2003. Low interest rates and stock market losses during the Great Recession weakened firms' funding positions drawing further increases in required pension contributions. These shocks disproportionately affected plans with marginal funding status as they were not buffered against large contributions requirements in the same way that overfunded plans were. The role of worsening plan finances as a key predictor of subsequent pension freezes is demonstrated greater detail in Appendix A.3, which relies on firm-level microdata.[12]

In the midst of major changes to the finances of DB pension funds, the Pension Protection Act (PPA) was signed into law in 2006. The PPA established more conservative standards on the interest rates that sponsors could use to discount future liabilities, reduced the period over which sponsors needed to amortize funding deficits from 15 years to 7 years, and required that plans with funding ratios of 80 percent or below make additional minimum contributions and pay higher insurance premiums to the Pension Benefit Guarantee Corporation (PBGC).[13] These key provisions of the PPA, which were phased-in as of 2008, raised statutory minimum pension contributions and imposed greater cost pressure on DB sponsors with marginal funding status.

1.2.2 Legal constraints to freezing plans have been alleviated

Between 1998 and 2000, a handful of prominent employers had converted traditional DB plans to CB plans. Older employees, who stood to lose substantial future pension accruals as a result these transitions, brought class action lawsuits against their employers claiming that CB conversions violated the age discrimination provisions of the Employee Retirement Income Security Act of 1974 (ERISA) (see, e.g., Zelinksy (2000)). As these cases played out in the court system, the legality of CB conversions remained uncertain. The threat of litigation along with large potential settlement costs for class action lawsuits likely constrained other employers from converting traditional DB plans. In 2006, however, a Federal appeals court ruled that IBM's CB conversion was age-neutral, thereby ending uncertainty surrounding

[12] Munnell and Soto (2007), Bovbjerg et al. (2008), and Rauh et al. (2020) also find that cost savings and funding volatility are important determinants of firms' decisions to freeze their DB plans. A compliance check on frozen DB plans conducted by the IRS in 2012 and obtained through a Freedom of Information Act (FOIA) request for this research project further corroborates this conclusion finding that sponsors typically froze their plans due to funding deficiencies.

[13] Some of the PPA's requirements were relaxed during the Great Recession as pension sponsors sought relief from strict funding targets. For instance, The Preservation of Access to Care for Medicare Beneficiaries and Pension Relief Act of 2010 allowed firms to elect extended amortization periods for any two plan years between 2008 and 2011. The extensions were for 9 or 15 years rather than the 7 year requirement of the PPA.

the legality of what was seen by many as an important test for restructuring traditional DB plans.[14] In the same year the new PPA law provided guidelines that CB conversions needed to meet in order to be considered age-neutral. Together, the appeals court ruling and the PPA's new provisions gave willing employers the legal cover they needed to restructure their DB plans.

1.3 Model of retirement and saving

This section develops a model of retirement timing and saving. Using the model, I show that pension freezes generate both substitution and wealth effects and therefore have an ambiguous impact on retirement behavior. I explain how heterogeneity in wealth and in preferences for leisure implies differential retirement responses for workers faced with freeze-induced losses in future compensation. I summarize these comparative statics into three empirically testable implications.

1.3.1 Defined benefit pension incentives

Before characterizing the model, I illustrate key features of DB plans and their effect on total compensation using data from the HRS. The left panel of Figure 1.3 shows the age-compensation profile for the average, DB eligible, HRS respondent between the age of 50 and 70. The dashed line shows wage and salary compensation. The solid line adds DB and DC pension accruals to wage and salary compensation. Notably, about half of the respondents in the DB eligible sample have non-zero DC balances.[15] Two types of non-linearities typical to DB pensions are evident in the figure. First, DB accruals generate prominent spikes in compensation. At age 55, which is a common early retirement age (ERA) in many DB plans, workers can begin claiming benefits thereby generating a substantial jump in pension wealth. After the ERA, accruals grow at a slower rate as workers approach the normal retirement age (NRA) which is typically 65. After the NRA, it is common for DB accruals to no longer grow in a manner that is actuarially fair and, in fact, become negative in real terms. This downturn occurs because workers who postpone retirement do not obtain sufficiently large increases in pension benefits even though they will obtain those benefits for fewer (mortality adjusted) years.[16] In the figure, pension compensation is positive for

[14] See, e.g., Mary Williams Walsh, "Issues Left Unresolved on Pensions," *New York Times*, Jan. 17, 2007.

[15] Earnings and pension benefits are computed for each individual at each age, so the age-based variation in compensation reflects plan-specific formulas for accruals and not changes in sample composition due to retirement.

[16] This implicit tax on continued work is a common feature of DB pensions specifically designed to encourage retirement (see, e.g., Chapter 12 in Gustman et al. (2000)).

workers over 65 only because DC accruals offset negative DB accruals.

The profiles shown in the left panel of Figure 1.3 illustrate how DB plans generate strong incentives for workers to remain employed through the ERA in order to collect generous retirement benefits. In contrast, after age 65, DB plans generate strong incentives for workers to retire. Underpinning these carefully designed incentives is an implicit contract between firms and workers: firms make long-term unenforceable commitments to workers in which continued employment through retirement age will be rewarded with lucrative pension benefits. Workers on the other hand exert effort over the course of their careers and avoid early termination that would result in the loss of promised pension benefits (see, e.g., Lazear (1981), Kotlikoff and Wise (1985), and Lazear and Moore (1988) for theory and evidence). By freezing DB pensions, firms are unilaterally reneging on this implicit contract.

The following section outlines a model of retirement in the presence of DB pension promises and uses the model to understand how worker behavior changes in response to pension freezes.

1.3.2 Setup

Individuals are heterogenous in age ($a \in \{50, 51, \ldots, 80\}$), non-pension wealth $A_a \in [0, \infty)$, DC pension wealth $W_a^{DC} \in [0, \infty)$, and the extent to which they dislike working, g. For individual ι currently aged a, the disutility of working is modeled as

$$g_{\iota a} = \underbrace{\gamma + \phi a}_{\text{deterministic component}} + \underbrace{f_{\iota a}}_{\text{random component}} \tag{1.3.1}$$

$$f_{\iota a} = \rho f_{\iota, a-1} + v_{\iota a}, \tag{1.3.2}$$

where $v_{\iota a} \sim \mathcal{N}(0, \sigma_v^2)$ is an iid disturbance term. The deterministic component of g captures changes in preferences for leisure that are common to workers of the same age, whereas the random component captures individual specific differences in health, disability status, and non-pay aspects of employer-employee match quality. I assume that utility from consumption exhibits constant relative risk aversion (CRRA) and that consumption and leisure are separable. Under these assumptions, flow utility for an individual ι aged a is given by

$$\frac{c_{\iota a}^{1-\frac{1}{\sigma}}}{1-\frac{1}{\sigma}} - g_{\iota a} \times \mathbf{1}\left\{\iota \text{ is working at age } a\right\}. \tag{1.3.3}$$

Earnings at age a are given by e_a which is taxable in the period that it is earned. Conditional on remaining employed, all workers share the same deterministic age-earnings profile with earnings at age a given by e_a. DC wealth evolves according to

$$W_{\iota a}^{DC} = W_{\iota,a-1}^{DC}(1+r) + e_a(m_{\iota a}^{w} + m^{e}(m_{\iota a}^{w})) \tag{1.3.4}$$
$$\bar{C} \geq e_a(m_{\iota a}^{w} + m^{e}(m_{\iota a}^{w})), \tag{1.3.5}$$

where r is the interest rate and $m_{\iota a}^{w}$ is the fraction of earnings that a worker defers towards her DC account on a pre-tax basis. $m^{e}(m_{\iota a}^{w})$ is the fraction of earnings that the worker's employer contributes to her DC account. m^{e} is expressed as a function of $m_{\iota a}^{w}$ reflecting commonly used employer incentives for participation in DC plans. IRS rules limit total contributions to be less than a threshold value $\bar{C}$.[17]

Workers earn DB pension wealth on the basis of a deterministic formula which is set by firms. The formula takes earnings history, tenure, and age as its arguments and returns the value of DB pension wealth:

$$W_{a}^{DB} = \mathcal{B}(\text{earnings,tenure,age}). \tag{1.3.6}$$

If a worker chooses to retire at age a, she obtains the annuity value of her DB wealth, which is b_a^{DB}. Workers also accrue Social Security benefits based on their earnings history in the form of an annuity whose value at age a is given by b_a^{SS}. Like the DB annuity, Social Security can only be claimed in retirement. $b_a = b_a^{DB} + b_a^{SS}$ is the total annuity income for an individual who chooses to retire at age a.[18] I assume that individuals cannot hold debt or borrow against DC pension assets, so $A_{\iota a} \geq 0$ and $W_{\iota a}^{DC} \geq 0$.

Total compensation for worker ι currently aged a is

$$\chi_{\iota a} = \underbrace{e_a(1-m_{\iota a}^{w})(1-\tau)}_{\text{non-deferred compensation}} + \underbrace{\Delta W_{\iota a}^{DC}}_{\text{DC accrual}} + \underbrace{\Delta W_{a}^{DB}}_{\text{DB accrual}}, \tag{1.3.7}$$

where τ is tax rate applied to income after pre-tax deferrals have been made.[19] Equation (1.3.7) reflects the assumption that workers have passed the age where real earnings growth contributes to higher Social Security benefits. Consequently, there are no increments to Social Security wealth from additional years of employment.

[17]In 2010 the total contribution limit was $49,000 with workers over age 50 being allowed to make an extra $5,500 in catch-up contributions. These values are updated annually by the IRS based on cost of living adjustments. See `https://www.irs.gov/pub/irs-tege/cola_table.pdf`

[18]62 is the earliest age at which workers can claim Social Security benefits. When estimating the model, I assume that workers who choose to retire before age 62 start claiming Social Security benefits at age 62.

[19]Income tax can be reduced by elective deferrals to DC accounts but payroll taxes apply to all earned income up to the relevant earnings caps.

1.3.3 Value of retirement

To conserve notation, denote age a variables by v and age $a+1$ variables by v'. The value of retirement at age a is V_a^R which is written recursively as

$$V_a^R(b, A, W^{DC}) = \max_{A', W^{DC'}} \left\{ u(c) + \beta(1-p_a) V_{a+1}^R(b, A', W^{DC'}) \right\} \tag{1.3.8}$$

$$\text{s.t. } c \leq \left(b + W^{DC} - \frac{W^{DC'}}{1+r} \right)(1-\tau) + A - \frac{A'}{1+r}. \tag{1.3.9}$$

In equation (1.3.8), $u(c)$ is utility from consumption, β is the annual discount factor, and p_a is the subjective probability of death within one year for an individual currently aged a. Annuitized pension income (b) is fixed at the age of retirement which is why b rather than b' appears in the continuation value. Given the state variables and the budget constraint, the retirees consumption choice is recast as the choice of how to decumulate pension and non-pension wealth at each age. Pension income, which is the sum of annuitized income and liquidated DC pension wealth, is taxable while liquidated non-pension wealth is not taxable.

Notably, the value function treats retirement as a self absorbing state which excludes the possibility of short-term bridge employment. The assumption of self-absorbing retirement is important in the context of this paper because it makes future compensation contingent on current labor supply in a very stark way: if a worker enters the retirement state, her future labor market options are entirely foreclosed. I provide empirical evidence in Section 1.6.1 that workers affected by pension freezes tend to follow a once-and-for-all retirement pattern which is consistent with the modeling assumption I make here.

1.3.4 Value of working

Denote the age a vector of state variables for a working individual by $X = (e, g, b, W^{DC}, A)$. The value of working at age a is V_a^W which is written recursively as

$$V_a^W(X) = \max_{A', m^w} \left\{ u(c) - g + \beta(1-p_a) E_{g'} \left[\max \left\{ V_{a+1}^R(b', A', W^{DC'}), V_{a+1}^W(X') \right\} \right] \right\} \tag{1.3.10}$$

$$\text{s.t. } c \leq e(1-m^w)(1-\tau) + A - \frac{A'}{1+r}. \tag{1.3.11}$$

In equation (1.3.10), flow utility from consumption is offset by the disutility of working g. The expectation operator for the continuation value term integrates over the random variable g'. By remaining employed at age a, workers preserve the possibility of obtaining higher earnings and higher pension accruals in the future which is reflected in the continuation value. Given the state variables, the budget constraint (1.3.11), and the DC accumulation constraints

(1.3.4) and (1.3.5), workers choose how much non-pension wealth to accumulate and what fraction of their earnings to defer towards DC pensions. I assume that DC pension wealth is illiquid until retirement.[20]

1.3.5 Optimal retirement decision

A worker retires when $V_a^R \geq V_a^W$. This condition defines a cutoff value $\bar{g}_a$ so that any draw of $g_a \geq \bar{g}_a$ will lead to retirement. Solving for $\bar{g}_a$ yields

$$\bar{g}_a = u(c_a^W) - u(c_a^R) + \beta(1 - p_a) \times \tag{1.3.12}$$
$$\left[\underbrace{\overbrace{E_{g'}\left[\max\left\{V_{a+1}^R(a+1, b', A', W^{DC'}), V_{a+1}^W(X')\right\}\right]}^{\text{1: Continuation value of working}} - \overbrace{V_{a+1}^R(a+1, b, A', W^{DC'})}^{\text{2: Continuation value of retiring}}}_{\text{Option value of working one more period}} \right]$$

where c_a^W and c_a^R are the optimal consumption choices in the work and retirement states at age a. Consider terms 1 and 2 on the right side of equation (1.3.12). Higher future earnings and the potential for higher pension accruals raise term 1 and increase the incentive for continued work. On the other hand, higher levels of retirement wealth raise term 2, thereby generating an incentive to retire. These two offsetting incentives are key determinants of the optimal retirement decision, which is summarized by the cutoff value $\bar{g}_a$.

In this framework, the decision about whether to continue working or to retire is determined by the option value channel which embodies the implicit contract of continued employment with the firm: By remaining employed in the current period, the worker obtains earnings growth and increased pension benefits in the future. Furthermore, because DB pension wealth is backloaded and cannot be ported between employers, the option value of working has a particularly strong firm-specific component. Following Stock and Wise (1990), work decisions in this model vary not only due to period-by-period changes in compensation, but also due to changes in the option value of continued employment which is embedded in term 1.

1.3.6 The effect of a pension freeze is theoretically ambiguous

When employers renege on long-term promises by freezing DB pensions, workers experience changes to the total value of compensation that they had previously expected to earn by retirement. The effect of these shocks is illustrated in the right-hand panel of

[20] IRS rules allow penalty free distributions from DC accounts after age $59\frac{1}{2}$. Distributions taken prior to $59\frac{1}{2}$ are subject to a 10 percent tax. Distributions must start at age $70\frac{1}{2}$. I do not incorporate these institutional features into the model.

Figure 1.3. The solid blue line shows the path of total compensation in a world where DB plans are kept intact (i.e. no pension freezes occur). When a DB plan is frozen, workers keep all previously earned benefits but no longer earn any new DB accruals. Instead, workers who do not already participate in DC plans are offered the chance to do so after the freeze.[21] These changes are shown using dashed red lines which simulate the effect of a DB pension freeze for workers aged 52 and 65. The post-freeze simulations shown in the figure assume that respondents without DC wealth begin actively participating in a hypothetical new DC plan after the freeze.

The figure highlights that pension freezes shift the age-compensation profile in different ways based on the age at which workers experience them. The average 52 year-old experiences large initial losses in compensation followed by gains after age 63. In contrast, the average 65 year-old experiences much smaller losses, followed by immediate gains. In fact, workers over age 65 experience unambiguous gains. The main reason that pension freezes have positive effects on compensation over the long-term is because workers with DC plans can continue to accumulate pension wealth even at ages where traditional DB plans would penalize continued work.[22]

Now consider terms 1 and 2 on the right side of equation (1.3.12) for a worker who is under age 65 when her plan is frozen. Holding all else fixed, the loss in DB accruals generates two opposing effects. First, reductions to b^{DB} lower the return to working (term 1) thereby decreasing the option value of working and generating a substitution effect on labor supply. Second, reductions to b^{DB} make workers poorer in retirement (term 2) thereby increasing the option value of working and generating a wealth effect on labor supply. The same mechanisms work with opposite signs on a worker who is over age 65 when her plan is frozen: freezes reduce the penalty for continued work thereby increasing labor supply through the substitution effect. At the same time, greater pension wealth accumulation raises the continuation value of retirement thereby lowering labor supply through the wealth effect. Since substitution and wealth effects work against each other, the impact of freezes on retirement behavior cannot be signed.

Despite the overall theoretical ambiguity, two sources of heterogeneity allow me to characterize the following testable predictions about how pension freezes affect labor supply.

1. Between-age heterogeneity

[21] In the HRS sample the underlies the figure, about half of all DB eligible respondents have no DC wealth.

[22] In the post-65 age range, there is a second reason that freezes generate positive effects on total compensation. DB formulas typically link benefits to pay earned over the last few years of a worker's career. Because pay falls after age 60, a worker in a non-frozen plan will see her benefit shrink as her pay declines while a worker whose plan is frozen will see her benefit stay fixed.

(a) *The relative strength of wealth and substitution effects varies by age:* Workers under 55 when their plans are frozen have large accruals still outstanding and experience substantial losses in compensation. Workers over 55 have less left to earn from their DB pensions when their plans are frozen and therefore experience smaller losses in compensation. As such, wealth effects are more important for workers under 55. Conversely, substitution effects are more important for workers over 55.

(b) *The effect of freezes on employment is reversed after age 65:* Because workers over the age of 65 experience increases rather than decreases in deferred compensation, their initial labor supply response should be reversed in sign relative to workers under the age of 65.

2. Within-age heterogeneity

(a) *Short-term responses are dominated by the substitution effect:* Short ex-ante work horizons reduce the role for wealth effects since workers are about to retire anyway. Thus, the response of workers near the margin of retirement at the time of the freeze (i.e. those with high values of $g_{\iota a}$, $A_{\iota a}$, and $W_{\iota a}^{DC}$) is dominated by the substitution effect. These workers will choose leisure over work when faced with a freeze.

1.4 Data

In this section I describe three distinct data sources that I use to estimate the effect of freezes on worker behavior. Pension plan data come from IRS Form 5500, employer characteristics come from the Census Longitudinal Business Database (LBD), and matched employer-employee data come from the Longitudinal Employer Household Dynamics (LEHD) dataset. I am able to link 92 percent of plans from Form 5500 to employers in the LBD and 89 percent of those LDB employers to the LEHD dataset.[23] Detailed descriptions of the datasets and the linking procedures that I use are provided in Appendix A.1.

1.4.1 Pension characteristics: Form 5500

Form 5500 (F5500) is an annual plan-specific filing that is collected jointly by the IRS, Department of Labor (DoL), and the PBGC to ensure compliance with ERISA. These publicly available data contain rich information on the universe of privately sponsored

[23] These datasets have previously been used in conjunction to study the role of fringe benefits on employee mobility in Decressin et al. (2009).

pension plans. When DB plans are converted to CB plans or hard frozen, this information is reported on F5500, thereby allowing me to identify plans whose participants are affected by freezes.[24] F5500 data also include important pension plan characteristics such as the number of plan participants, present value of pension wealth for plan participants, the value of pension accruals earned by participants in the filing year, and the typical age at which participants claim benefits from the plan.

F5500 filings are identified by a combination of a Federal Employer Identification Number (EIN) and a employer designated Plan Number (PN) that remain consistent over time. While these identifiers are sufficient to match pension plans to single-unit (i.e. operating only one establishment) firms, they are not sufficient for matching to multi-unit (i.e. operating multiple establishments) firms. This is because payroll tax filings for establishments that are part of a multi-unit firm may be recorded under different EINs than the one used in F5500. Attempting to match the F5500 to establishment level data on EIN alone would therefore generate many false non-matches. To overcome this issue, I turn to the Census Business Register (BR).

1.4.2 Firm characteristics: Census Business Register and Longitudinal Business Database

The Census Business Register (BR) is a database of the universe of establishments in the United States.[25] It includes information on business location, organization, industry, and information on revenue, payroll, and employment that is collected from administrative tax records as well as survey data. The relationship between establishments belonging to multi-unit firms are determined using responses to the company organization survey, the economic census, and the annual survey of manufactures. Establishments that are part of the same multi-unit firm share the same Census assigned firm identification number even if they have different EINs.

I rely on the presence of EINs in both the F5500 and the BR to create an initial link between the two files. Secondary to this link, I use the Census firm identifier to identify *all* the establishments associated with a multi-unit pension plan sponsor.[26] Having matched F5500 records to the BR, I use the Census firm identifier to further match those records to

[24]Plans are considered frozen in F5500 if they meet the following condition: "[a]s of the last day of the plan year, the plan provides that no participant will get any new benefit accrual (whether because of service or compensation)."

[25]Information about the BR is confidential and protected by Title 13 and Title 26, U.S. Code. Information in the following paragraph is drawn from `https://www.census.gov/econ/overview/mu0600.html`

[26]In the absence of the Census firm identifier, I would only be able to identify those multi-unit establishments that shared the same EIN as the one reported on F5500, thereby generating a false non-match problem alluded to earlier.

the LBD. The LBD is a cleaned and research ready version of the BR that is restricted to active employers in the private sector.[27]

1.4.3 Worker characteristics: Longitudinal Employer Household Dynamics

To study outcomes at the individual level, I turn to the LEHD which is a quarterly matched employer-employee dataset constructed from state-level unemployment insurance (UI) records.[28] These data cover almost all wage and salary workers in the United States but exclude individuals who are self-employed. In the LEHD, employers are identified using a state UI account number known as the SEIN. I rely on the crosswalk between the SEIN and the Census firm identifier developed in Haltiwanger et al. (2014) to link pension plans in the matched F5500-LBD data to employers in the LEHD.[29]

An important feature of the LEHD is that states become part of the dataset at different points in time. For example, Maryland enters in 1985:Q2 whereas Mississippi enters only in 2003:Q3. Because of staggered entry, the scope of the data grows continuously over time.[30] As a consequence, when I link an employer from the LBD to the LEHD in a given year, I only capture those individuals who work in a state that has already entered the dataset as of that year.

1.4.4 Sample restrictions and data structure

Appendix Table A.1 describes the results of four data linking and sample restriction procedures. The first row shows the match rate between the universe of DB plans extracted from F5500 database between 1996 and 2014 and the BR. The massive scope of the BR allows for a 92 percent match rate at the plan-year level and a 95 percent match rate at the participant-year level.[31]

The set of plan-years represented in the F5500-BR merge contains a mix of firms that sponsor just one DB plan and firms that sponsor multiple DB plans.[32] I limit my sample to firms that have a single plan within the 1996-2014 window for which I have F5500 data. When firms have multiple plans, I retain only those employers who choose either to never freeze their plans, or freeze them all at the same time. The principle driver of this restriction is that I

[27] See Jarmin and Miranda (2002) for details.

[28] See Abowd et al. (2009) for details.

[29] I rely on the 2014 snapshot of the LEHD, which incorporates UI data from 49 states and the District of Columbia through the first quarter of 2015. Alabama is not included in the version of the data that I use.

[30] A number of populous states enter the data relatively early. Illinois enters in 1990, California and Pennsylvania in 1991, Florida in 1992, and New York and Texas enter in 1995.

[31] F5500 reports the count of active participants — i.e. covered workers — in each plan.

[32] Firms that sponsor multiple DB plans typically do so to cover different types of workers. For example, a firm may sponsor different DB plans for salaried and hourly workers or unionized and non-unionized workers.

cannot observe individual pension plan coverage. Consequently, when firms sponsor multiple plans, there is no way of knowing — using the F5500, LBD, or LEHD data — which plan a worker may be covered by. By imposing this restriction, however, I can ascertain whether workers at a given firm have been affected by a freeze in a given year. This sample restriction allows me to retain 94 percent of firm-years but only about 40 percent of worker-years. The discordance between these two rates reflects the fact that only the very largest employers sponsor multiple DB plans.

Having matched F5500 records to the BR and the LBD, I structure the data as follows. I treat each year from 2001-2014 as an experiment year, which is indexed by l.[33] This terminology reflects the research design wherein each experiment year yields a fresh sample of firm-level pension freezes. Workers employed at freezing firms five years prior to a given experiment year constitute the treated group while workers employed at non-freezing firms five years prior to a given experiment year constitute the comparison group. I impose the restriction that firms file F5500 for 5 calendar years prior to the experiment year, which I refer to as the pre-period. I then match the firm-experiment year data to the LEHD, the results of which are shown in the third row of Table A.1. I recover 89 percent of firm-experiment years and 93 percent of employee-experiment years. Finally, as shown in the fourth row of Table A.1, I restrict the sample to firm-experiment years for which important pension plan data is not missing.[34]

When considering the implications of pension freezes on worker decisions, it is worth reiterating that I do not observe individual information on pension plan coverage. To study worker responses in a way that limits the potential for misclassification error, I restrict the sample to firms where DB eligibility is near universal. I impose this restriction by retaining firms where the DB coverage rate is 80 percent or greater in the pre-period.[35] Within the high-coverage rate firms, I select all workers employed at $l-5$, who have at least two years of tenure as of $l-5$, and who will be between the age of 50 and 70 in year l.[36]

[33] I start with 2001 because it is the first year in which pension freezes are reported in F5500. Only a handful of firms engaged in CB conversions prior to 2001.

[34] Important pension plan information includes plan assets and liabilities, accruals earned during the plan year, and the typical benefit claim age for the plan.

[35] The firm-wide DB coverage rate is the ratio of active participants in the plan as reported in F5500 to the count of total employees in the LBD. The 80 percent average coverage rate requirement is based on years $[l-5, l-2]$, i.e. between 5 and 2 years prior to the experiment year. Restricting the sample this way likely eliminates soft freezes in which the firm's plan is closed to new workers. A firm that imposes a soft freeze is likely to see its DB coverage rate decline as workers quit or retire but are not replaced with new, DB eligible, workers.

[36] The two year tenure restriction ensures that workers are fully vested in their pensions as of the cohort year when they may become subject to a freeze. This calculation is based on the 7 year maximum full vesting period allowed for DB plans by ERISA. Note that tenure measurements are right censored in the LEHD when a worker's employment spell begins prior to the year in which the state that they work in enters the

Having discussed the key sources of data and explained how I link and organize these data for analysis, I turn next to the regression framework and associated identification assumptions that I use to estimate causal effects.

1.5 Empirical framework for estimating treatment effects

This section describes the regression framework I adopt to analyze the impact of pension freezes on a variety of labor market outcomes. The identifying assumption is that pension freezes are independent of unobserved determinants of labor supply, conditional on a set of worker characteristics, firm characteristics, and fixed effects. Summary statistics lend credibility to the identifying assumptions.

1.5.1 Regression specification

Let i index firms, let j index cells that bin together workers with the same firm, state, gender, age, tenure, and experiment year, and let t index calendar years. Let $k = t - l$ index years relative to the experiment year. Within each cell, the firm represents a worker's employer as of $k = -5$, and tenure represents the duration of employment at firm i as of $k = -5$. Consider the following regression framework for a given experiment year, l,

$$y^l_{j(i)t} = \alpha_i + \gamma^l_t + {\boldsymbol{x}^{l}}'_{j(i)t}\boldsymbol{\beta}^l + \sum_{k=-5}^{m(l)} \delta^l_k T^l_{ik} + \varepsilon^l_{j(i)t} \tag{1.5.1}$$

where $y^l_{j(i)t}$ measures a labor supply outcome of interest, α_i is a firm fixed effect, γ^l_t is a calendar year fixed effect, and $\boldsymbol{x}^l_{j(i)t}$ is a vector of controls for age, gender, state, race, education, tenure, and prior earnings.[37] T^l_{ik} is an indicator variable that equals 1 if the firm freezes its plan and the current period is k. $\varepsilon^l_{j(i)t}$ is the error term which represents unobserved determinants of labor supply.[38] The parameters of interest are the δ^l_k coefficients which capture the dynamic treatment effect of pension freezes on worker outcomes.

To maximize the precision of the estimates, I stack data from each of the experiment

data.

[37] State is defined based on the location of the workplace in $k = -5$. I control for prior earnings using two variables: the log of average annual earnings prior to $k = -5$ and the growth rate of earnings prior to $k = -5$.

[38] The upper limit of the sum, $m(l)$, represents the number of available post-period years for experiment year l. The maximum available post-period duration is 13 years (this happens when $l = 2001$ as the data runs out in 2014).

years together and estimate a version of equation (1.5.1) where

$$y^{l}_{j(i)t} = \alpha_i + \gamma_{lt} + \boldsymbol{x}^{l'}{}_{j(i)t}\boldsymbol{\beta} + \sum_{k=-5}^{13} \delta_k T^{l}_{ik} + \varepsilon^{l}_{j(i)t}. \tag{1.5.2}$$

In equation (1.5.2), calendar year fixed effects are replaced by experiment year-by-calendar year fixed effects which allow economy-wide shocks to differentially affect workers in each experiment year. In contrast, the effect of the $\boldsymbol{x}^{l}_{j(i)t}$ and T^{l}_{ik} variables are assumed to be constant across experiment years. This estimation strategy allows workers in the comparison group in a given experiment year to enter the treated group in a subsequent experiment year if their employer freezes pensions in the future. In this implementation, the δ_k coefficients are identified by within-experiment year between-firm variation in worker outcomes as well as within-firm between-experiment year variation in worker outcomes. Standard errors are clustered at the firm level.

1.5.2 Identification

The δ_k parameters in equation (1.5.2) represent causal effects of pension freezes on worker outcomes under the assumption that $E[\varepsilon^{l}_{j(i)t}|\alpha_i, \gamma_{lt}, \boldsymbol{x}^{l}_{j(i)t}, T^{l}_{it}] = 0$. Put differently, unobserved determinants of worker labor supply are assumed to have zero mean conditional on firm fixed effects, experiment year-by-calendar year fixed effects, worker level controls, and the freeze indicators. This assumption might be violated by two important sources of bias. First, firm-specific economic distress in the pre-period may result in a subsequent freeze as well as a reduction in firm-specific labor demand through downsizing.[39] Second, it is possible that freezing and non-freezing firms systematically differ in terms of pension generosity and benefit claiming provisions in the pre-period which can influence post-period differences in labor supply behavior.

To account for time-varying pre-period confounders, I rely on propensity score re-weighting. The main idea behind the use of propensity scores is to make the treatment and comparison group units more comparable in terms of observed pre-period characteristics thereby mitigating concerns that post-period differences in behavior are subject to bias. In this setting, the propensity score is the cell-level probability of experiencing a freeze expressed as a function of pre-period variables which influence both firms' decision to freeze and workers' labor supply responses. To mitigate concerns related to pre-freeze firm distress as an omitted variable, the propensity score model includes the pre-period trend

[39] Along with a broader set of statistics on firm dynamics around the freeze, I show in Appendix A.3 that freezing and non-freezing firms do not differ in terms of their pre-period probability of experiencing distress.

in firm size and in worker compensation. To mitigate concerns that differences in plan- and firm-level characteristics between treatment and comparison groups are responsible for post-period labor supply decisions, the propensity score model also includes pre-period trends in pension wealth, pension accruals, benefit claim ages, the age structure of employment at the firm, retirement rates, employment rates, and employer-to-employer transition (E-E) rates. Appendix A.4 provides more details on the conditioning set and explains how the propensity scores are transformed into weights when estimating equation (1.5.2).

Beyond these key threats to the identification strategy, two other confounding effects are potentially at play. First, it is possible that funding deficiencies that lead firm's to freeze their DB plans also resulted in cut backs to health insurance benefits. These unobserved changes could generate their own income and substitution effects on labor supply choices.[40] While these changes are not directly verifiable in the data that I use, indirect evidence from surveys suggests that firms have not altered health benefits as a consequence of freezes.[41] Second, it is possible that observed changes in labor supply are influenced by network effects within the firm. This breakdown of the so called stable unit treatment value assumption (SUTVA) could occur if freeze-affected workers' retirement decisions are influenced not only by changes in compensation but also changes by the retirement decisions of their peers. I ignore peer retirements as a first order concern when interpreting the results because available evidence on the magnitude of peer effects of this variety indicate that they are extremely small.[42]

1.5.3 Summary statistics

Before showing the impact of freezes on worker outcomes, I present a summary of raw data on pre-period characteristics of workers and their employers. These statistics are based on workers employed at the sample of firms where the pre-period coverage rate is in excess of 80 percent. A table showing firm characteristics for the full sample of DB sponsoring employers is provided in Appendix A.3.

In Table 1.1 I show pre-period summary statistics for workers split into three different

[40]Employer sponsored health insurance is reported on F5500 filings. However, the reporting requirement only exists for employers who cover more than 100 workers. More importantly, changes in health insurance plan characteristics cannot be ascertained from F5500.

[41]There are no reports of changes to employer provided health insurance benefits in a sample of 17 large publicly traded firms that froze their plans between 2004 and 2008 as compiled by the Boston College Retirement Research Center (see `http://crr.bc.edu/uncategorized/fact-sheets/`). Similarly, a Government Accountability Office (GAO) survey of freezing employers indicates no reported changes to health benefits (see Bovbjerg et al. (2008)).

[42]Hamman et al. (2016), who use large-scale linked employer-employee data from Germany to investigate these spillovers on retirement behavior, find that one additional peer retirement (at the establishment level) increases the probability of retirement for men by 0.01 percentage points and produces no detectable effect on women.

groups based on age as of the experiment year.[43] Workers in the sample are employed with DB sponsoring firms as of $l - 5$, and the statistics are computed by averaging over five pre-period years. The top panel shows worker characteristics while the lower panel shows pension plan and firm characteristics. Because the statistics are computed from a worker-level dataset, pension and firm characteristics are worker weighted. For each age-specific panel, the first column shows the propensity score re-weighted comparison group mean, the second column shows the difference between the treatment and the comparison group, and the third column shows the p-value for the null hypothesis that there is no difference between the two groups. Across the three sub-samples, workers are well educated with relatively high earnings and the gender split is close to 50-50. Demographic characteristics, pre-period labor market outcomes, pension plan generosity, and firm-level characteristics are very similar in the treated and control groups.[44]

Secondary to the differences between treatment and comparison groups within each age bin, there are also several notable differences between the three age bins. Workers in the younger two age bins are less likely to be white and male, more likely to have a college degree, and have higher earnings. Tenure, measured five years prior to the experiment year, is approximately equal across the three age groups.[45] Finally, the oldest workers are employed at substantially smaller firms with a higher proportion of workers over age 60 and a lower proportion of workers under age 45. Differences in firm-wide age structure across the three age groups could reflect differences in DB pension formulas or other unobserved workplace characteristics. Some of these differences are reflected in higher average pension wealth and delayed retirement claim ages at firms that employ the oldest workers in the sample. Appendix Table A.5 shows the same statistics without propensity score re-weighting provides additional evidence of broad similarity between workers in the treated and control groups.

1.6 How pension freezes affect labor supply and employer attachment

In this section, I use the regression framework and identification strategy developed earlier to investigate the causal impact of pension freezes on the labor supply and employer

[43]Using 55 and 65 as the modal ERA and NRA in DB plans, 50-55 year-old workers are below the ERA at the time of the freeze, 56-64 year-old workers are between the ERA and the NRA at the time of the freeze, and 65-70 year-old workers are over the NRA at the time of freeze.

[44]It is important to note that the measures of pension wealth are based on plan-wide totals and should not be seen as representative of the workers in each age bin

[45]The absence of variation in tenure between workers of different ages is an artifact of the way that states enter the LEHD dataset. If a state enters the dataset after a given employer-employee relationship is established, then the employer-employee history is left censored and tenure is understated.

attachment. I show how freezes have heterogeneous labor supply effects based on the age at which workers experience them. In addition, I show that freezes have heterogenous effects holding age fixed. The treatment effects I present here are consistent with the testable implications developed in the theoretical model.

1.6.1 Employment, retirement, and earnings

Figure 1.4 plots the δ_k coefficients from specification (1.5.2) using the cell-level employment rate as the outcome variable. To investigate age-specific heterogeneity in labor supply responses, I split the data by age as of the experiment year using the same age groups as in Table 1.1. I then estimate the regression model on each age group separately and show the coefficients in the respective panels of the figure.

Looking first at the far left panel shows that treated workers in the 50-55 year-old age group exhibit a small reduction in employment rates in the first six years of the post-freeze period. Reductions in employment reflect substitution effects (i.e. reduced lifetime labor supply), although the economically small magnitude of the coefficients and their statistical insignificance suggests that offsetting wealth effects (i.e. increased lifetime labor supply) are equally important for workers in this age group. After about 8 years post-freeze, wealth effects start to dominate the labor supply response and workers in the treated group experience a 1.5 - 3.3 percentage point increase in employment relative to the comparison group. Another way of stating this finding is that the employment rates of both treated and comparison groups are declining in this age range, but the decline is markedly slower for treated workers. The muted substitution effect and substantial wealth effect that characterizes labor supply responses for 50-55 year-old workers aligns with the fact that large DB accruals are earned before age 55. As such, workers at or under 55 when first faced with a freeze experience large net losses in compensation thereby eliciting strong wealth effects in favor of continued employment.

Moving next to the central panel of the figure shows that substitution effects play an important role on impact for treated workers in the 56-64 year-old age group. In the first six years of the post-freeze period, employment rates for treated workers fall by 1.0 - 1.7 percentage points relative to the comparison group. Starting about 8 years post-freeze, treated workers' labor supply diverges in the opposite direction from the comparison group as the employment rate differential rises by 1.2 - 2.1 percentage points. As with 50-55 year-old workers, the tendency of freeze affected workers to lengthen their working lives relative to the comparison group is indicative of dominant wealth effects. Appendix Figure A.3 shows that freeze-affected women are more likely to exhibit substitution effect dominant responses whereas freeze-affected men are more likely to exhibit wealth effect dominant responses.

Most DB plans incentivize retirement by making the real value of DB accruals negative after age 65. When DB plans are frozen and replaced with DC plans, this implicit tax on continued employment is reduced as workers can obtain offsetting DC accruals at ages where they would otherwise have been penalized. The right most panel of Figure 1.4 shows labor supply behavior that is consistent with higher returns to work for the treatment group relative to the comparison group. Treated workers over age 65 increase their employment rates by 1.2 - 4.5 percentage points which reflects dominant substitution effects. While these effects are not as precisely estimated, they are economically meaningful and align with theoretical predictions and the institutional design of DB plans.

Empirical evidence shown in Figure 1.4 lines up with the three theoretical predictions outlined in Section 1.3. Substitution effects are larger for 56-64 year-old workers relative to 50-55 year-old workers which is a consequence of the former group losing less total compensation than the latter group (prediction 1(a)). Substitution effects are positive because of increased labor market returns for workers over 65, but negative for workers under 65 because of reduced labor market returns (prediction 1(b)). Finally, looking within the first two age groups, substitution effects play a dominant role in the short-term response of the most marginally attached workers who lower their employment rates (prediction 2(a)).

Figure 1.5 shows the impact of freezes on retirement, which is defined as a permanent departure from paid employment in the LEHD.[46] Freeze induced changes in retirement rates are virtually mirror images of the employment effects shown in Figure 1.4, indicating that non-employment and retirement are essentially equivalent for workers affected by freezes. This finding indicates that bridge jobs appear not to be an important transition phase for freeze-affected workers who cut their careers short. It also substantiates the model's assumption of self-absorbing retirement. On the whole, the treatment effects for retirement reinforce employment-based findings.

Figure 1.6 shows the effect of freezes on the log of annual earnings. All the estimates shown here are conditioned on the sample of individuals with positive earnings in a given year. The far left panel shows that there are no evident changes in earnings for 50-55 year-old workers in the first 6-7 years of the post-freeze period. After 7 years, workers in the treated group exhibit slower age-related earnings declines which manifest as a positive earnings differential. In the final few years of the sample window, treated group workers' earnings exceed those of control group workers by about 15 log points. This difference likely arises from continued full-time work for the treated group relative to transitions into part-time or part-year work for the comparison group. It is consistent with wealth effects of the freeze

[46]See Appendix A.2 for details on how retirement status is measured in the LEHD and how it compares with retirement status for a comparable sample of respondents in the HRS.

inducing longer careers in full-time status.

For 56-64 year-old workers, who are shown in the central panel, the pattern is somewhat different. Treated workers experience a statistically significant earnings dip of 1.5 - 4 log points in the first five post-freeze years. This dip could arise for two reasons. First, treated workers who shorten their careers — i.e. exhibit dominant substitution effects in the short-term — may work less than full-time or less than full-year right after the freeze. These reductions in labor supply would appear as earnings losses. Second, it is possible that these workers experience wage declines in the short-term brought on by firm-specific factors.[47] The fact that 50-55 year-old workers do not experience earnings losses suggests that wage reductions are unlikely to be the only cause for the observed earnings dip that 56-64 year-old workers experience. In contrast, the long-term pattern shows a 20-30 log point increase in earnings relative to the comparison group, provides evidence of dominant wealth effects wherein freeze-affected workers delay retirement and continue in full-time or full-year employment while comparison group workers transition to part-time or part-year employment. As was the case for the 50-55 year-old age group, it is important to reiterate that the positive long-term earnings differential reflects a slower decline rather than an increase earnings levels.

Earnings differences for 65-70 year-old workers, shown in the far right panel, are not precisely estimated over the sample window. However, the post-freeze coefficients indicate a fairly sustained drop in earnings of about 15 log points relative to the comparison group. Thus, while freezes induce workers over 65 to delay retirement, the pattern of earnings changes suggest that continued employment for these workers likely comes in the form of more part-time employment. That the oldest workers in the labor market are willing extend their careers at less than full-time rates complements recent survey-based evidence on the importance of flexible hours in supporting longer working lives (Ameriks et al. (2020)).

1.6.2 Employer attachment

While the treatment effects shown thus far relate to the decision about whether to work and how much to work, they do not address the decision about where to work. E-E transitions are a potentially important margin of adjustment particularly given that pension freezes induce employer-specific rather than worker-specific or market-wide changes in compensation. In this subsection, I exploit the matched employer-employee structure of the LEHD to study differences in worker mobility between the treatment and comparison

[47] Analyses of firm-level payroll data shown in Appendix A.3 indicate that freezes lower firm-wide average earnings by approximately 2.5 log points. These earnings changes could stem from the changing post-freeze age composition of the firm's workforce or from reductions in offered wages.

groups.

Figure 1.7 shows the percentage point change in the probability of leaving one's DB sponsoring employer.[48] In the data, a worker is coded as having experienced an employer change if the EIN associated with their UI record in the LEHD changes. It is worth noting that not all EIN changes reflect employee mobility as some firms change their EINs in the course of a merger or acquisition. The large spike in transitions for treated workers that occurs four years prior to the freeze year in the left panel and the center panel is likely an artifact of firm-level EIN recoding.

One period after the freeze, treated workers in the 50-55 and 56-64 year-old age group appear to respond with small but statistically significant increases in employer transitions. For 50-55 year-old workers the transition rate increases by 1.4 percentage points off a baseline rate of 2.9 percent. For 56-64 year-old workers, the transition rate increases by 0.8 percentage points off a baseline rate of 2.8 percent. The magnitude of these effects indicates that relatively younger workers faced with compensation losses from a pension freeze have better outside options to exercise than workers closer to retirement age. After the first year, there is a statistically significant pattern of reduced E-E transitions among workers in the treated group. Transition rates fall by 1.2 percentage points for 50-55 year-old workers and about 0.6 to 0.9 percentage points for 56-64 year old workers. When considered alongside the earnings results, reduced E-E mobility is consistent with continued employment in career jobs for treated group workers as opposed to the counterfactual transition to part-time or bridge jobs for comparison group workers. E-E mobility for workers over 65 is largely unaffected by freezes.[49]

Reduced E-E transition rates for workers in the younger two age groups lines up with the time window where dominant wealth effects lead to increased employment and higher propensity for full-time work. Ironically, the ability of workers to extend their careers in order to make up for lost compensation comes from extended attachment to the very employers responsible for the pension freeze. Reduced E-E mobility, even in the face of substantial employer-specific compensation shock, indicates that full-time work opportunities are limited outside of workers' long-term employers. As such, robust demand for the labor services of older workers within their long-term employers is a key requirement to accommodate policies aimed at supporting longer working lives.

The results I have presented thus far characterize treatment effects of pension freezes on labor supply outcomes. These estimates provide insights about heterogeneity in workers' preference for long careers and illustrate the importance of the right type of labor demand

[48] I.e. the employer that worker's are attached to in period $l-5$.

[49] Coefficient estimates are suppressed due to small sample sizes for latter part of the estimation window.

in sustaining those longer careers. Nevertheless, because I do not observe detailed pension benefit provisions I cannot estimate the post-freeze value of lost DB accruals or the change in DC accruals. The absence of data on pecuniary costs of pension freezes prevents direct estimation of labor supply elasticities. To estimate elasticities indirectly, I rely on estimating and simulating data from the structural model which I turn to next.

1.7 Solving and estimating the structural model

In this section, I explain how I solve and estimate the structural model numerically using MSM (McFadden (1989), Duffie and Singleton (1993)). I identify the model's parameters by matching model-based simulations of employment responses to pension freezes to those observed in real-world administrative data.

1.7.1 Estimation

I estimate the model in two steps. In the first step, I calibrate several parameters (e.g., earnings, pension accruals, DC match rates, Social Security benefits, tax rules, mortality rates, discount factors, and interest rates) using HRS survey data and other external data sources. Appendix A.5.1 provides a detailed description of the calibrations. In the second step, I estimate the model's preference parameters $\boldsymbol{\theta} = (\sigma, \gamma, \phi, \rho, \sigma_v)$ using MSM. The procedure is as follows:

1. For a given value of $\boldsymbol{\theta}$, I numerically solve the model using value function iteration under two different scenarios. In the first scenario, DB accruals progress normally. In the second scenario, I freeze DB accruals for each age between 56 and 64 and compute an alternative set of decision rules.

2. I simulate initial assets and work disutility draws for 5000 individuals who are initially aged 51 to 59 using information from the HRS. I apply decision rules from the no-freeze scenario to obtain work histories and asset accumulation paths from the initial age up to age 80 (the terminal age) to create a simulated control group.

3. Next, I apply the freeze decision rules for the same population of individuals — i.e. individuals with the same initial assets and work disutility draws —starting five years after the initial age. Individuals in this exercise have the same work and asset accumulation choices as the simulated control group for the first five years, but have different work histories and asset accumulation choices once faced with a DB freeze. I call this sample the simulated treated group.

4. I compute two sets of moments using the simulated control and treated groups. The first set of moments is the difference in average employment rates between the two groups (the simulated treatment effect). I compute these differences for 12 periods, starting from the period of the freeze (12 moments). The second set of moments is the employment rate time trend for the simulated control group. I compute the trend moments starting four periods prior to the freeze and lasting five periods after the freeze (10 moments). I then compare the 22 simulated moments with same moments estimated from LEHD data (real-world moments).

5. I iteratively repeat this procedure for different values of $\boldsymbol{\theta}$ and choose the estimate that minimizes the distance between the simulated moments and the real-world moments.

Further details on the solution algorithm and the estimation procedure are provided in Appendix A.5.2.

1.7.2 Identification

Identification of the model's parameters derives from two key sources of variation in real-world data. The first set of moments, which capture the age-based decline in employment rates in the comparison group, identify the constant (γ) and slope (ϕ) terms of the deterministic component of g.

The second set of moments — i.e. the dynamic treatment effect — is informative about within-age variation in individual preferences for continued work. Equation (1.3.12) helps to explain the identification argument for (σ, ρ, σ_v). Notice that $\bar{g}$ is composed of two separate terms: the first term is the change in utility from consumption at retirement, and the second term is the option value of working. Freezes affect both terms. Changes in $\bar{g}$ driven by post-freeze re-optimization of consumption choices provide information about individuals preference for smooth consumption profiles over the life-cycle. This source of variation in $\bar{g}$ aids in the identification of the intertemporal elasticity of substitution (IES), σ.[50]

The treatment effect estimates reveal that freeze-induced changes in the option value of work induce early retirement for some workers and delayed retirement for others. These differences are informative about the persistence (ρ) and variance (σ_v) of the idiosyncratic component of g. In particular, ρ and σ_v need to be large enough so that workers of the same age exhibit different retirement behavior when affected by a compensation shock of the same magnitude. For workers of a given age, those with high values of g have stronger

[50]Note that consumption data are not directly used to identify σ; rather, variation in $\bar{g}$ induced by freezes indirectly aids in identification.

preferences for leisure and retire early. On the other hand, those with low values of g have weaker preferences for leisure and choose to delay retirement.

1.7.3 Model fit and parameter estimates

Figure 1.8 shows how simulated moments compare with observed moments. The left panel shows the employment rate trend for 56-64 year old workers who are not subject to freezes. The light colored area shows the region where simulated moments are specifically targeted to match real-world moments, whereas the shaded region shows out-of-sample fit. The model fits the targeted moments well but slightly underestimates employment propensity out-of-sample. The right panel compares simulated treatment effects to real-world treatment effects, all of which are used as estimation targets. This panel shows that the model captures the magnitude and the timing of employment rate fluctuations induced by pension freezes.

Table 1.2 shows the estimated parameters. The IES estimate is close to 1, which is equivalent to log utility. Because the 5 preference parameters in $\boldsymbol{\theta}$ are identified using 22 moments, I am able to conduct a χ^2 overidentification test. The model is formally rejected on the basis of this test, which I report in the last row of the table. The main reason that the test statistic is large is that the estimation procedure does not account for variance associated with first-step parameters obtained from the HRS (e.g, earnings profiles, DB pension wealth accruals, DC match functions, etc.) Consequently, the variance of second-step simulated moments is understated and the weight attached to each moment (which is inversely related to variance) is large.[51]

1.7.4 Employment elasticities

In the model, labor supply responses to changes in compensation are governed by the fraction of individuals who are marginal with respect to the decision to work or retire which, in turn, is a function of the preference parameter $\boldsymbol{\theta}$. I compute elasticities non-parametrically by plugging averages from simulated data into the following expression

$$\hat{\eta}_a = \frac{\hat{E}\left[\Delta \text{emp}_{\iota a}\right] / \hat{E}\left[\text{emp}_{\iota a}^{C}\right]}{\hat{E}\left[\Delta \chi_{\iota a}\right] / \hat{E}\left[\chi_{\iota a}^{C}\right]}. \tag{1.7.1}$$

In equation (1.7.1), $\hat{\eta}_a$ is the elasticity of employment with respect to a one period shock in current compensation for workers aged a.[52] I estimate $\hat{\eta}_a$ by simulating a 5 percent increase

[51] Standard errors are understated for the same reason.

[52] Because retirement is self-absorbing in the model, $\hat{\eta}_a$ calculated using model based simulations are not directly comparable to elasticity estimates based on model simulations or real-world data where workers can leave and re-enter the labor market (see, e.g., French (2005), Brown (2013), and Gelber et al. (2017)). With

in pre-tax wage compensation and reading off the model implied change in employment rates for different ages. $\Delta\text{emp}_{\iota a}$ is the difference in a worker's employment status with and without the age a shock to compensation. $\text{emp}^C_{\iota a}$ is a worker's employment status without the compensation shock. Analogous definitions apply to the post-tax current compensation measures $\Delta\chi_{\iota a}$ and $\chi^C_{\iota a}$. $\hat{\eta}_a$ is defined for the sample of workers who are employed at $a-1$.

$\hat{\eta}_a$ is the extensive margin labor supply elasticity governing employment responses to short-lived tax changes. This definition of the labor supply response to compensation changes does not account for the option value of continued work as in Stock and Wise (1990). To account for this forward-looking dimension of retirement behavior, I define $\hat{\eta}^{PV}_a$ as the elasticity of employment with respect to the present value of future compensation:

$$\hat{\eta}^{PV}_a = \frac{\hat{E}\left[\Delta\text{emp}_{\iota a}\right]/\hat{E}\left[\text{emp}^C_{\iota a}\right]}{\hat{E}\left[\Delta\omega_{\iota a,R^*}\right]/\hat{E}\left[\omega^C_{\iota a,R^*}\right]}, \tag{1.7.2}$$

where

$$\omega_{\iota a,R^*} = \underbrace{\sum_{t=a}^{R^*}(1-p_a)(1+r)^{t-a}\chi_{\iota t}}_{\text{PV from } k \text{ to retirement}}. \tag{1.7.3}$$

Conditional on being employed at age $a-1$, $\omega_{\iota a,R^*}$ is the present value of future compensation, R^* is the period in which the worker retires, $(1-p_a)$ is the likelihood of being alive at age a, and r is the real interest rate. Because R^* is the ex-ante or pre-shock expected retirement age, $\Delta\omega_{\iota a,R^*}$ in equation (1.7.2) represents the change to the present value of compensation that the worker was expecting to receive before retirement. $\hat{\eta}^{PV}_a$ re-casts a one period shock to compensation at age a and expresses it in terms of a change in the present value of compensation that the worker was expecting to earn before retirement. This parameter is consistent with the notion that retirement decisions involve a comparison of the value of working and retiring at all future ages. I estimate $\hat{\eta}^{PV}_a$ using the same simulated 5 percent shock to wage compensation at age a but I calculate the present value impact of that shock using equation (1.7.3).

Table 1.3 shows the two types of elasticities for workers who experience compensation shocks at age 58, 60, and 62.[53] The estimates in the table show that while workers

self-absorbing retirement, the value of retirement (V^R) does not nest future re-entry into the labor market. Thus, larger compensation cuts — or reductions in option value — are needed to induce retirement relative to a model where V^R nests future re-entry into the labor market. Consequently, the elasticity estimates that I present are lower bounds relative to a model where retirement is not self absorbing.

[53] In the estimates presented in the table, I estimate $\hat{E}\left[\Delta\text{emp}_{\iota a}\right]$, $\hat{E}\left[\Delta\chi_{\iota a}\right]$, and $\hat{E}\left[\Delta\omega_{\iota a,R^*}\right]$ using within-person variation thereby holding age and work disutility (g) fixed. When computing elasticities,

appear to be only weakly responsive to a single period change in compensation, re-casting the change in terms of its effect on the long-term compensation yields a substantially larger elasticity. At age 60 for instance, the estimated employment elasticity with respect to a one period change in after-tax compensation is 0.15 while the option value-based elasticity is 0.9. This difference in elasticities illustrates how measuring the retirement response to one-period-ahead changes in the reward to working understates the importance of forward-looking behavior in explaining retirement decisions.

In the next section, I exploit the full structure of the estimated model to emphasize this difference by examining how a permanent elimination of OASI payroll taxes for older workers affects their retirement behavior.

1.8 Evaluating the effectiveness of a counterfactual payroll tax reform

In this section, I study the effect of eliminating the OASI component of the payroll tax for older workers. I focus on transition cohorts for whom the policy change is unexpected. The main idea behind this reform proposal is to remove disincentives for longer careers that exist under current law. While payroll taxes are applicable to all years of work, the Social Security benefit formula is based on the 35 years of highest paid work. For the typical older worker over the age of 55 or 60 who no longer experiences real earnings growth, an additional year of work generates no substantial increase in Social Security benefits. Nevertheless, because all workers contribute OASI payroll taxes regardless of age, workers who remain employed beyond the 35 year vesting age experience tax burdens with no offsetting benefit increases. By relieving these fully "paid up" workers from additional payroll tax contributions, the reform has the potential of lengthening careers. Understanding the costs and benefits of the reform requires quantifying behavioral responses which I do here using the estimated model.

The flavor of reform that I consider involves an unexpected elimination of the OASI payroll tax for workers who are over age 60. To simplify the analysis, I assume that workers currently bear the full burden of the tax which implies that the counterfactual reform raises income by 10.6 percent (under current law workers and firms each pay 5.3 percent). I assume that the disability insurance (DI) and hospital insurance/Medicare (HI) components of the payroll tax are unaffected as is the Social Security benefit formula.[54] To more accurately portray incentives affecting the majority of employees in the current and future workforce, I

I assume that R^* is known to the worker with perfect certainty. This assumption ignores the effect of randomness in g.

[54] DI and HI payroll tax components collectively amount to 4.7 percent.

remove DB pensions from the model.

Figure 1.9 shows model simulations of employment rate trends for workers between the age of 56 and 75 under current law and under the reform.[55] The two trends diverge starting at age 60, with employment rate differences peaking at 15.3 percentage points at age 66. Averaging over the entire post-60 age window, the reform increases employment rates by 5.2 percentage points.

Table 1.4 provides additional statistics comparing the two different regimes. The left panel shows moments from the distribution of a variety of outcomes under current law, while the right panel shows the same moments under the reform. The third panel shows the mean difference in outcomes. Reflecting the overall increase in employment rates, the average retirement age under the reform rises by 1.1 years. Longer careers allow workers to accumulate almost \$8,000 more in their DC retirement accounts and provide about \$23,000 more in federal income taxes. These gains accrue against a loss in payroll tax revenue of approximately \$28,000. Finally, as shown in the lower panel of the table, I find that equivalent variation from the reform averages almost \$75,000 per worker implying a substantial welfare gain. This estimate is an order of magnitude larger than the one reported by Laitner and Silverman (2012) which is attributable primarily to the fact that I do not attempt to make the reform revenue neutral in this basic analysis.[56]

Taken together, the counterfactual analysis presented here shows that the relatively elastic labor supply of older workers can be harnessed to extend their careers with an OASI payroll tax sunset at age 60. I find that older workers would obtain non-trivial welfare gains from the reform and have more income in retirement although net revenues collected from workers after age 60 decline by about 4 percent.

1.9 Conclusion

In this paper, I exploit widespread and understudied shifts in employer-sponsored pension benefit programs to better understand how retirement behavior responds to changes in compensation. Over the last 20 years many employers have reneged on long-standing promises to continue supporting traditional retirement benefits in the face of rising costs of provision. Although these changes are substantial, there is no evidence on how they have affected workers' labor market outcomes. Creating a new dataset that brings

[55]Because the reform is unexpected there is no difference in employment rates prior to age 60. Simulated workers are included in the sample if they are employed at age 55.

[56]The way that Laitner and Silverman (2012) build revenue neutrality into their analysis is to require a small increment to payroll tax contributions in the pre-vesting period. This increase in pre-vesting age taxes results in lower consumption over the entire life-cycle and smaller equivalent variations.

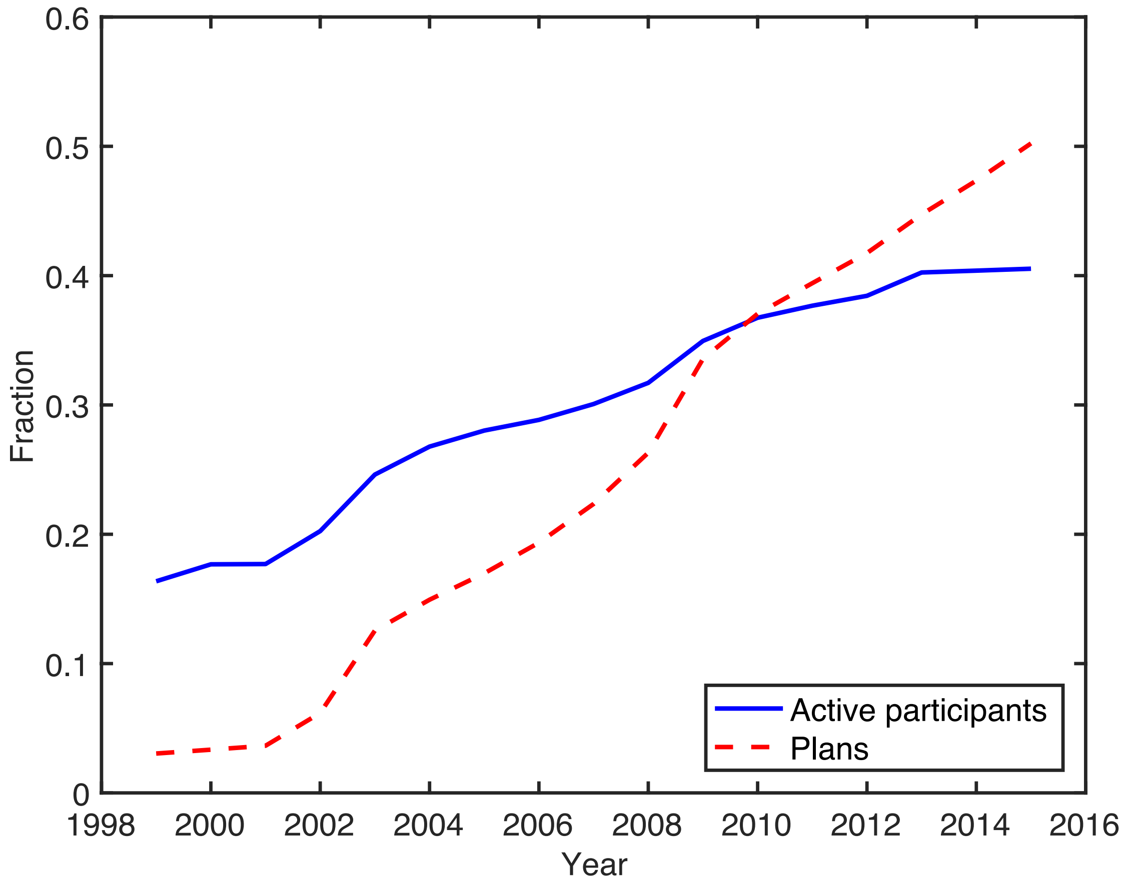
together detailed administrative information on pension plan characteristics with matched employer-employee data, I study the impact of these unexpected shocks on employment, retirement, earnings, and employer attachment for workers between the age of 50 and 70.

I find evidence of substantial heterogeneity among workers, even among those of the same age. When faced with freeze-induced compensation changes, some workers choose to retire early while others choose to delay retirement, thereby illustrating that differences in wealth and preferences for leisure are important in explaining retirement decisions. Using these quasi-experimental treatment effects as targets, I estimate a structural model of retirement and saving that allows for heterogeneity in wealth and leisure preferences. I use the model to simulate the effect of a counterfactual policy that eliminates OASI payroll taxes for workers who are fully vested in their Social Security benefits. Simulations from the estimated model show that eliminating the tax at age 60 induces a 1.1 year delay in the average retirement age and produces large welfare gains.

The empirical setting and the model adopted in this paper highlight the importance of forward-looking behavior in determining labor supply decisions, particularly retirement decisions. It is likely that unobserved determinants of the relationship between employers and employees, such as implicit contracts with a rich set of contingencies, drive the association between current labor supply and expected future compensation (see, e.g., Lazear (1981) and Akerlof and Katz (1989)). These contingencies create tight bonds between employers and employees that many standard models of labor supply do not incorporate. Notably, the durability of these relationships and the impact they have on forward-looking labor supply behavior are not unique to DB-style incentives, which are no longer common in the U.S. private sector. For instance, median tenure for workers between the age of 55 and 64 has remained unchanged for two decades despite major economic transitions including the demise of DB pensions.[57] This fact suggests that a better understanding of the mechanisms driving long-term employer-employee relationships is critical in explaining lifetime labor supply behavior.

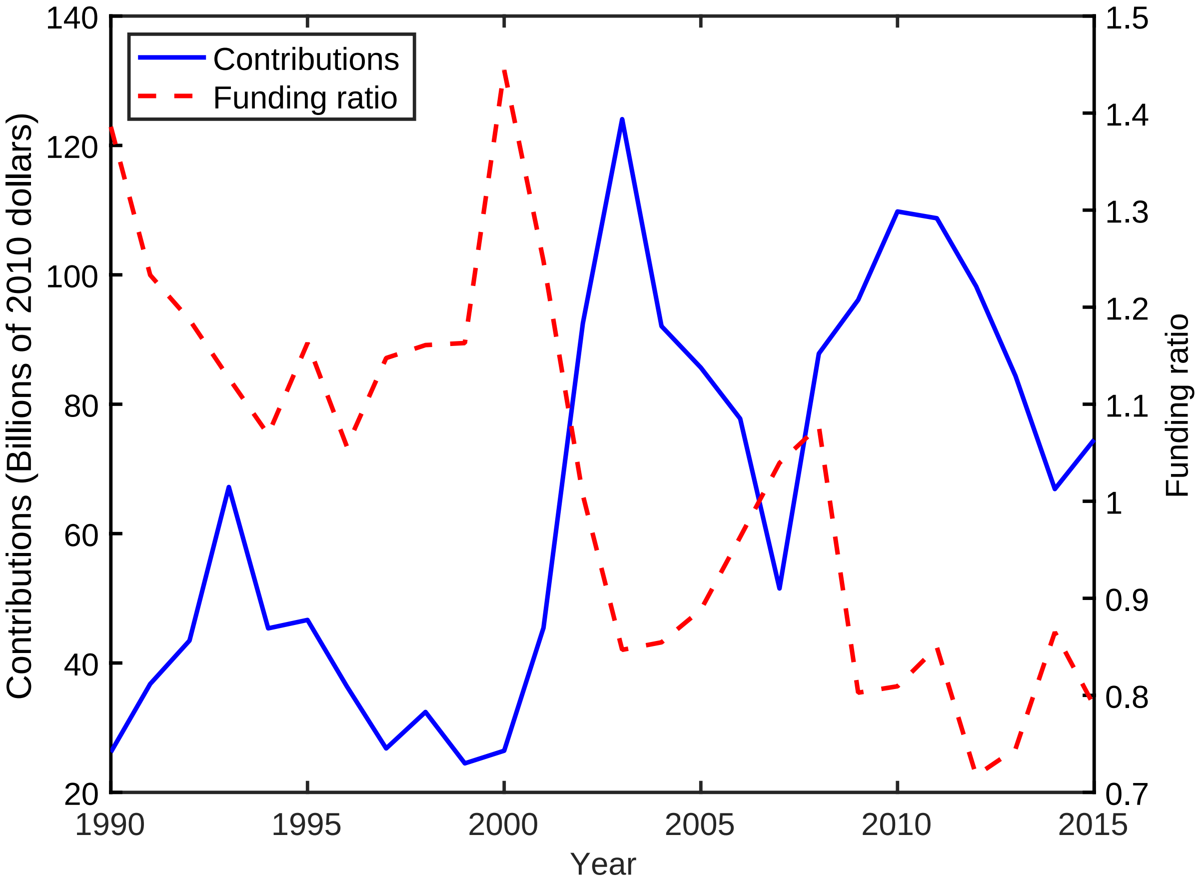
[57]Median tenure, as reported by the Bureau of Labor Statistics, for workers in the 55-64 age range in 1998 and in 2018 was 10.1 years.

Figure 1.1: DB plans hard frozen or converted to cash balance



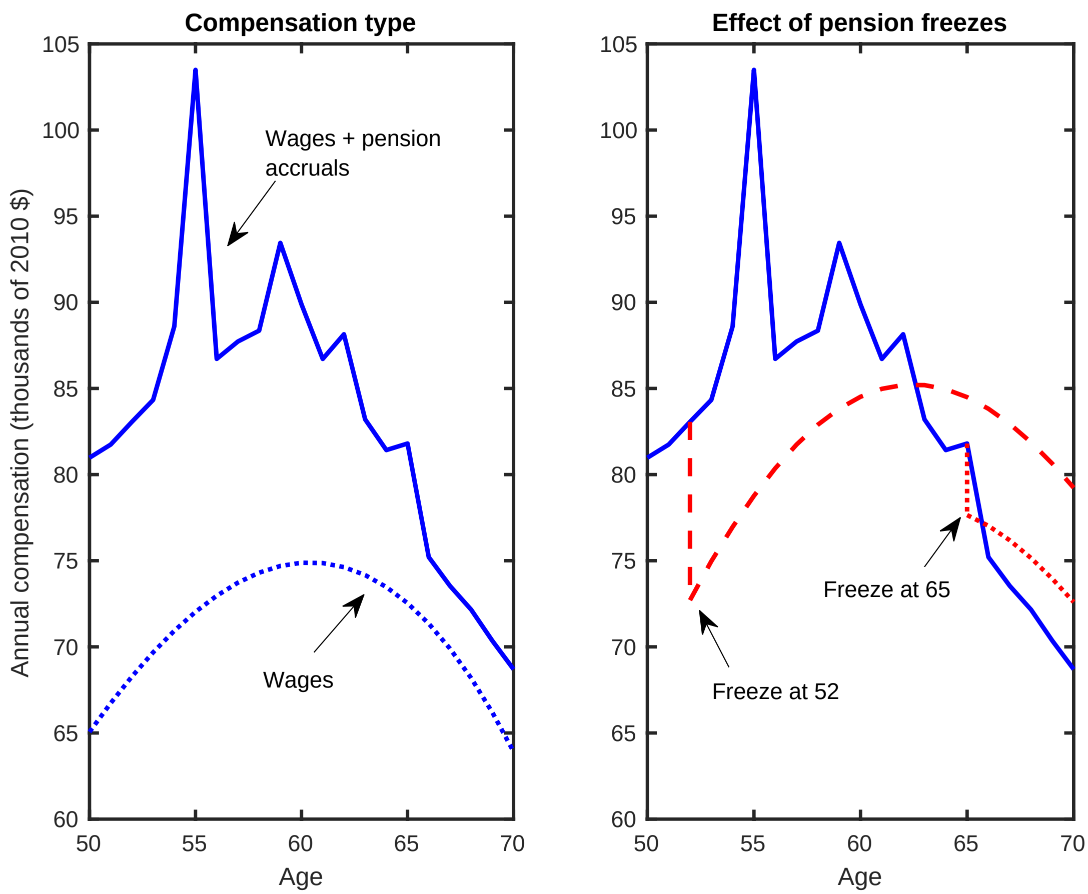
Notes: The dashed red line shows the share of private sector, single employer, DB pension plans that have been hard frozen or converted to cash balance plans. The solid blue line shows the share of active participants in DB plans that have been hard frozen or converted to cash balance plans. Time series are based on 1999-2015 F5500 microdata.

Figure 1.2: Pension costs and funding ratios



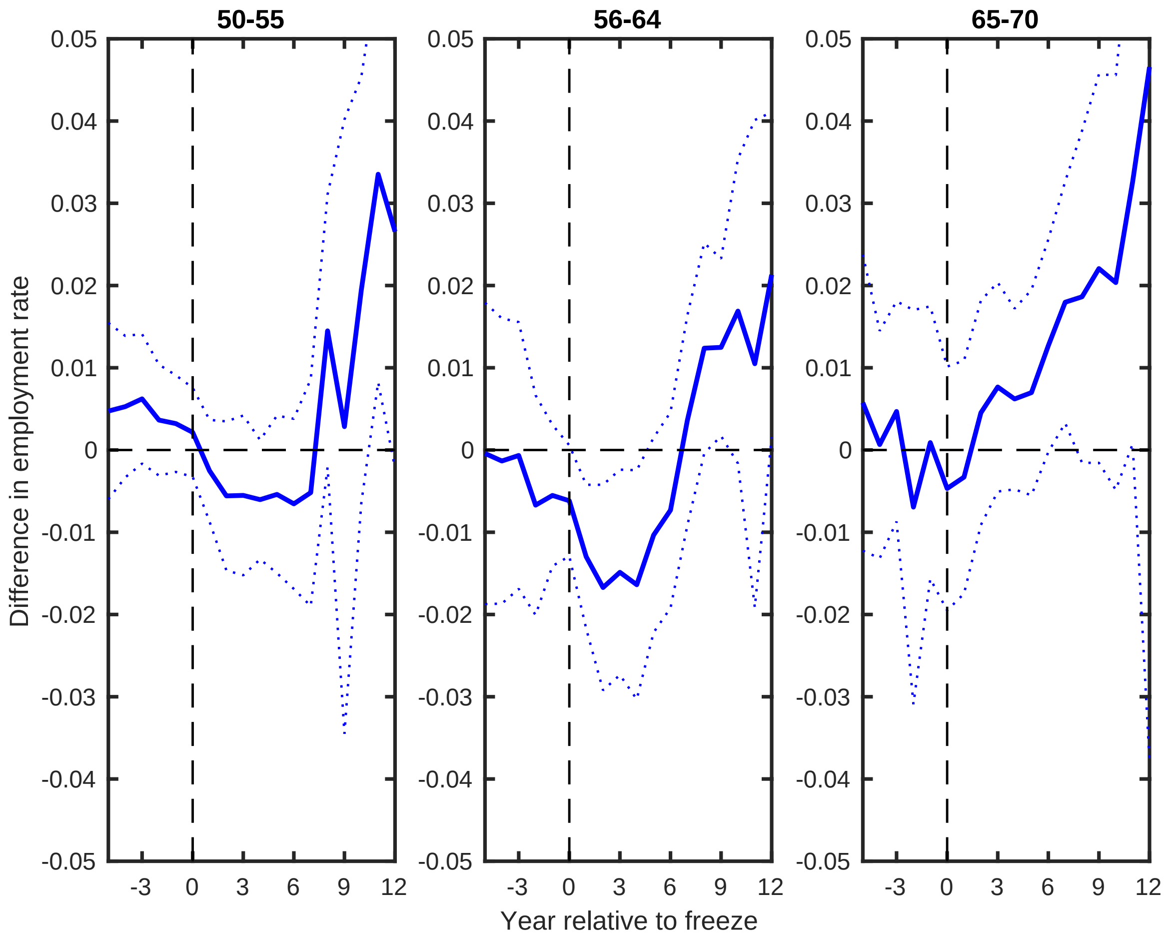
Notes: This figure shows total contributions and aggregate funding ratios for private sector, single employer, DB sponsors. Total contributions (solid blue line) are measured on the left axis. Funding ratios (dashed red line) are measured on the right axis. The funding ratio is the ratio of the market value of assets to the present value of future pension liabilities. Contributions data are drawn from the Department of Labor's Private Pension Plan Bulletin Historical Tables and Graphs 1975-2015. Funding ratios are drawn from the PBGC Pension Insurance Data Book, 2016.

Figure 1.3: Simulated effect of pension freezes on total compensation



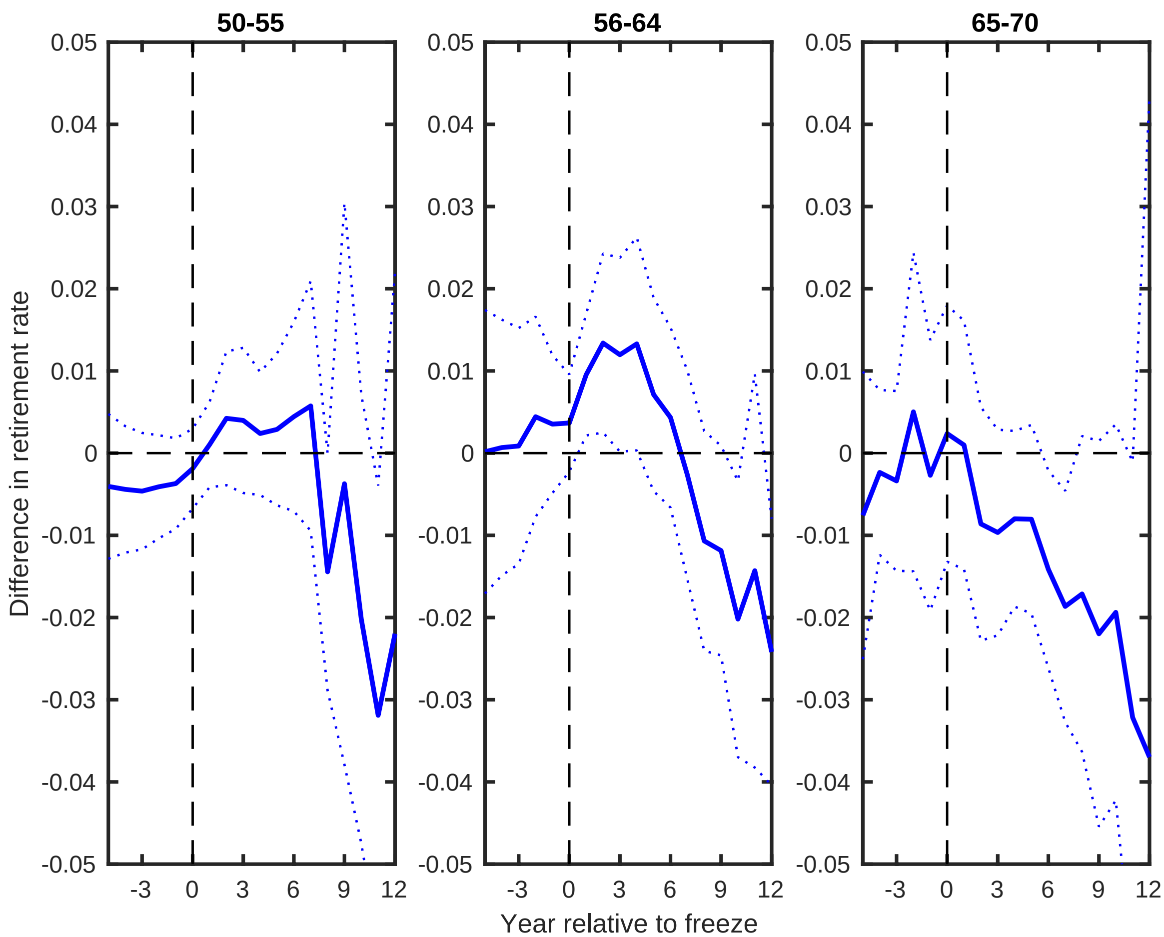
Notes: This figure is based on data from DB eligible HRS respondents in the 2010 wave who are employed in the private sector. For simulated freeze compensation paths, post-freeze DB accruals are set to zero. Respondents with DC plans are assumed to continue participating in them at the same rate; post-freeze contribution rates for respondents without DC plans are imputed using the sample average rate for actively contributing respondents. See Appendix A.6 for details on the sample and simulation of post-freeze compensation.

Figure 1.4: Impact of freezes on employment



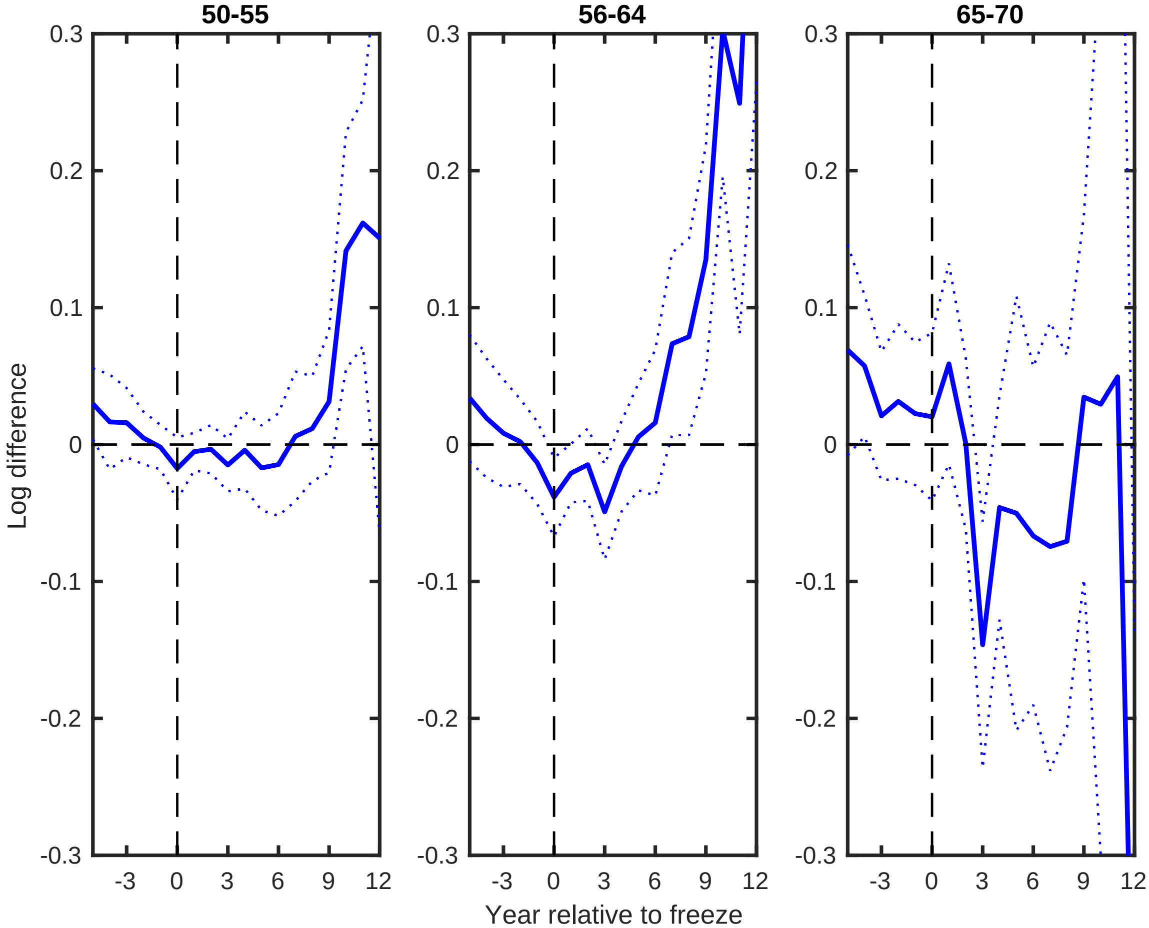
Notes: This figure shows the time path of the treatment effect of the freeze on employment rates for workers in the three different groups based on age at the time of the freeze. The p-values for F-tests that the coefficients for periods -5 to -1 are jointly zero are .41, .38, and .35 for each of the three age groups respectively. Dotted lines show 95 percent confidence intervals which are based on standard errors clustered at the firm-level.

Figure 1.5: Impact of freezes on retirement



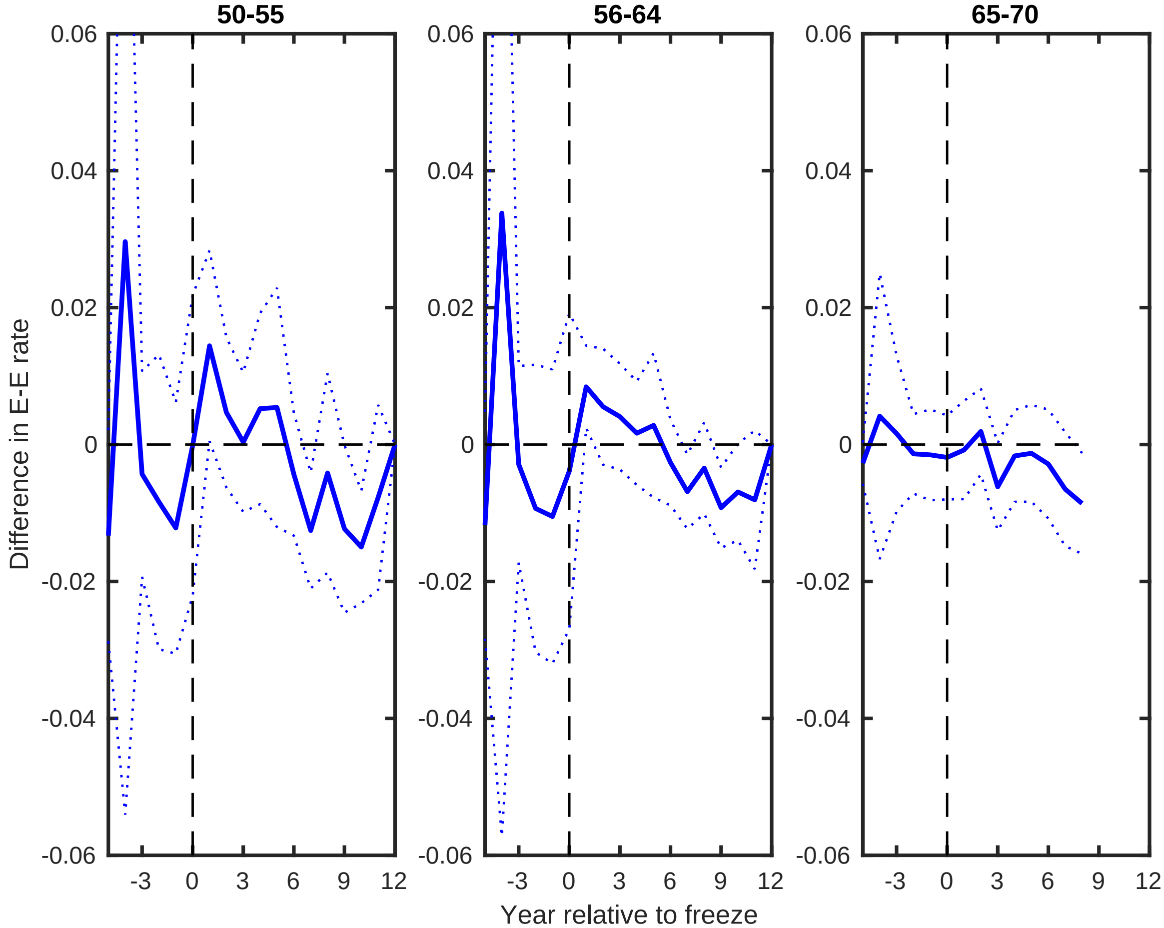
Notes: This figure shows the time path of the treatment effect of the freeze on retirement rates for workers in the three different groups based on age at the time of the freeze. The p-values for F-tests that the coefficients for periods -5 to -1 are jointly zero are .64, .77, and .35 for each of the three age groups respectively. Dotted lines show 95 percent confidence intervals which are based on standard errors clustered at the firm-level.

Figure 1.6: Impact of freezes on log annual earnings



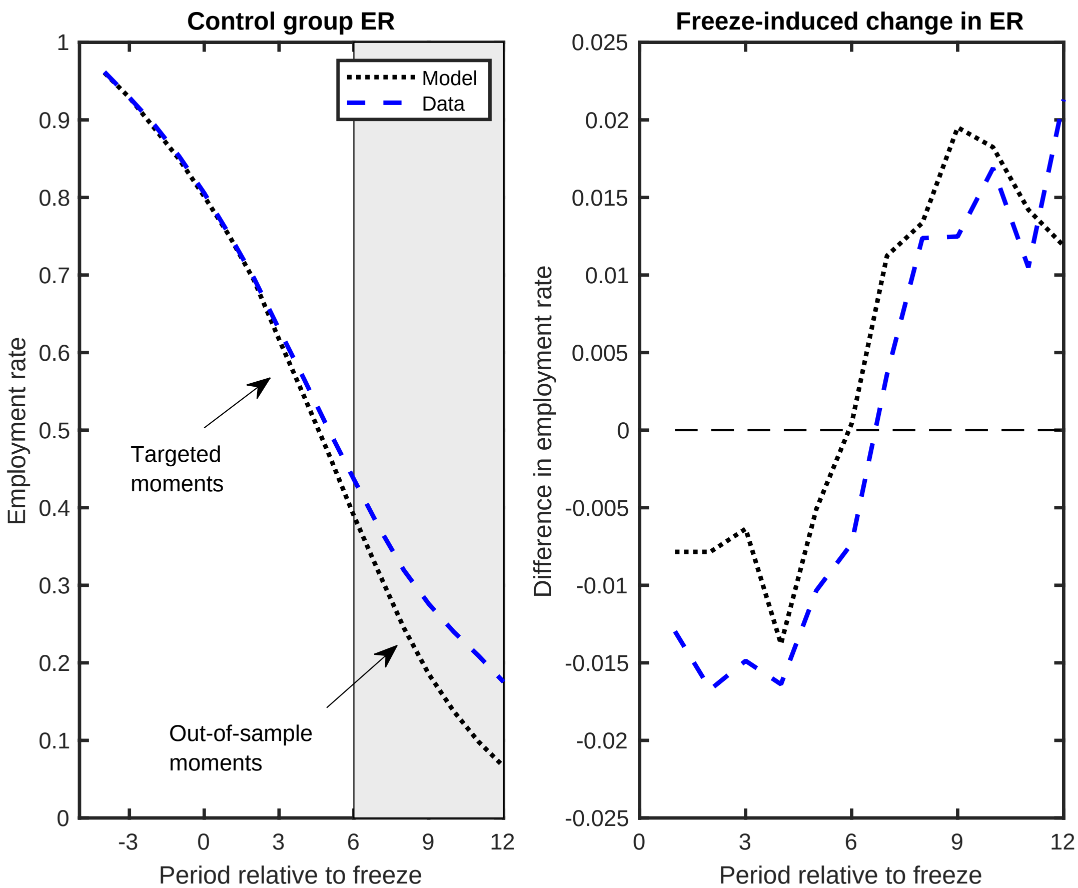
Notes: This figure shows the time path of the treatment effect of the freeze on log annual earnings for workers in the three different groups based on age at the time of the freeze. The p-values for F-tests that the coefficients for periods -5 to -1 are jointly zero are .17, .30, and .31 for each of the three age groups respectively. Dotted lines show 95 percent confidence intervals which are based on standard errors clustered at the firm-level.

Figure 1.7: Impact of freezes on employer attachment



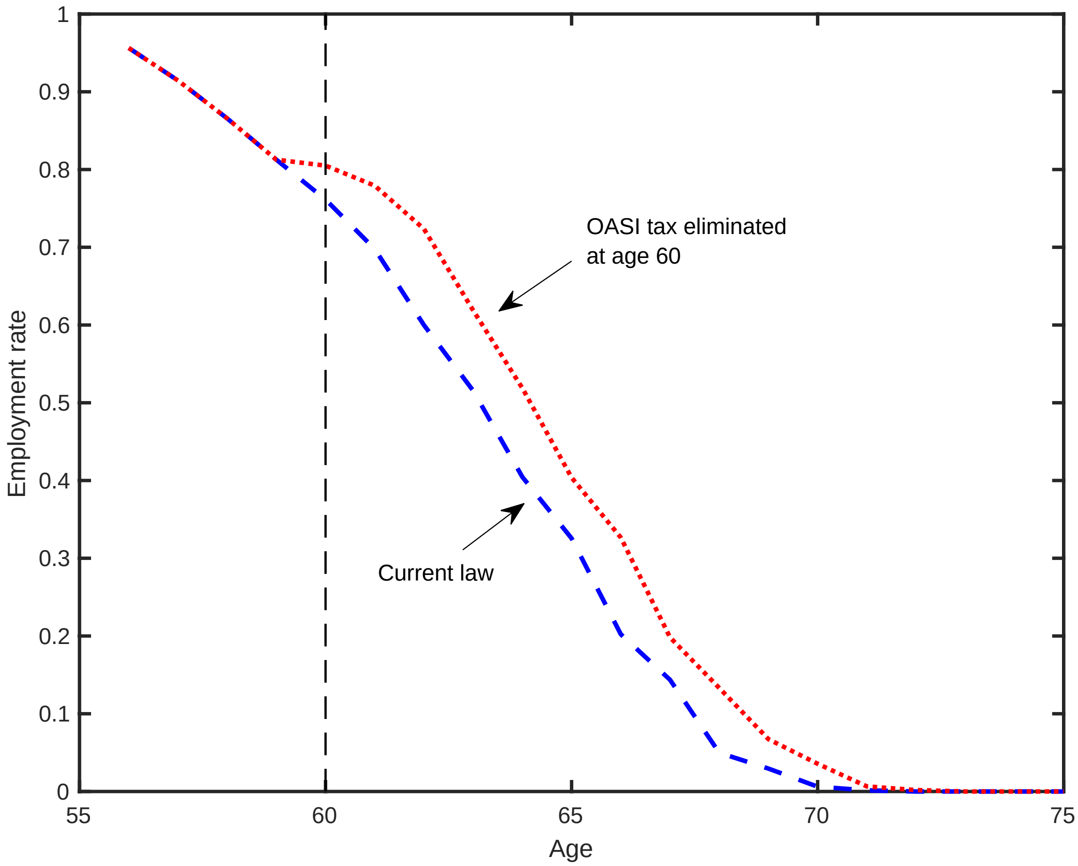
Notes: This figure shows the time path of the treatment effect of the freeze on log annual earnings for workers in the three different groups based on age at the time of the freeze. The p-values for F-tests that the coefficients for periods -5 to -1 are jointly zero are .01, .06, and .58 for each of the three age groups respectively. Dotted lines show 95 percent confidence intervals which are based on standard errors clustered at the firm-level.

Figure 1.8: Data moments compared to model simulated moments



Notes: The left panel shows the employment rate trend for 56-64 year-old workers, all of whom are employed 5 periods prior to the freeze. The right panel shows the difference between the treated and control group employment rates. Dotted lines show moments from simulated data. Dashed lines show moments from observed data. The shaded region includes moments that are not targeted in the estimation.

Figure 1.9: Effect of OASI tax sunset at age 60 on employment rates



Notes: This figure shows the evolution of the employment rate under two scenarios. The dashed blue line shows the employment rate trend under current law. The dotted red line shows the counterfactual employment rate trend under a regime where the OASI payroll tax is unexpectedly eliminated at age 60. The trends are based on model simulations for a sample of workers who are employed at age 55.

Table 1.1: Pre-period summary statistics split by age group

Variable	50-55 Comp. mean	50-55 Diff.	50-55 p-value	56-64 Comp. mean	56-64 Diff.	56-64 p-value	65-70 Comp. mean	65-70 Diff.	65-70 p-value
Worker characteristics									
Age	52.5	0.001	0.93	59.6	-0.031	0.32	67.0	0.016	0.56
Male	0.475	0.025	0.46	0.487	0.023	0.55	0.494	0.014	0.76
High school	0.232	-0.001	0.93	0.232	-0.001	0.93	0.294	0.001	0.98
Some college	0.324	-0.006	0.55	0.312	-0.007	0.42	0.290	-0.005	0.58
College or more	0.386	0.008	0.81	0.392	0.010	0.77	0.311	0.008	0.76
White	0.791	0.022	0.20	0.826	0.020	0.25	0.838	0.027	0.25
Black	0.096	-0.011	0.15	0.079	-0.010	0.12	0.069	-0.010	0.21
Hispanic	0.068	-0.003	0.81	0.053	0.000	0.97	0.049	-0.005	0.61
Other race	0.045	-0.008	0.27	0.043	-0.009	0.23	0.044	-0.013	0.31
Earnings ($)	65,140	-1,477	0.75	65,280	-2,337	0.58	47,770	-3,562	0.13
Tenure at $l-5$	7.8	-0.943	0.15	8.2	-0.977	0.11	8.0	-0.623	0.17
Retired	0.022	0.000	0.99	0.056	0.002	0.57	0.189	0.006	0.59
In labor force	0.964	-0.001	0.83	0.927	-0.003	0.47	0.779	-0.006	0.66
Switched $l-5$ employer	0.047	0.012	0.22	0.042	0.012	0.24	0.030	0.003	0.45
Pension and firm characteristics									
Log DB pension wealth/active participant	10.07	0.072	0.86	10.15	0.039	0.92	10.43	0.043	0.82
Log DB pension accrual/active participant	7.43	0.094	0.80	7.49	0.054	0.88	7.60	0.052	0.77
Pension plan claim age	62.8	-0.1	0.70	62.9	-0.1	0.72	64.4	0.0	0.88
Log firm size	8.55	-0.017	0.98	8.40	0.007	0.99	7.31	-0.162	0.74
Fraction workforce ≤ 45	0.580	0.010	0.44	0.568	0.007	0.61	0.542	-0.003	0.80
Fraction workforce [46,50]	0.146	-0.002	0.62	0.141	0.000	0.91	0.135	0.001	0.79
Fraction workforce [51,55]	0.124	-0.002	0.61	0.127	-0.001	0.78	0.124	0.001	0.89
Fraction workforce [56,60]	0.087	-0.002	0.63	0.095	-0.002	0.67	0.098	0.001	0.74
Fraction workforce [61,65]	0.044	-0.002	0.40	0.047	-0.002	0.52	0.066	0.000	0.99
Fraction workforce [66,70]	0.012	-0.001	0.38	0.013	-0.001	0.54	0.022	0.000	0.87
Fraction workforce ≥ 71	0.007	-0.001	0.33	0.008	-0.001	0.56	0.013	0.001	0.70
Comparison group workers		383000			373000			77000	
Treated group workers		60000			66500			11000	
Comparison group firms		7700			8600			4600	
Treated group firms		1500			1700			900	

Notes: Unless otherwise noted, statistics reported in the table average over the five year period preceeding any freeze activity. Pension wealth per active participant is computed as the present value of the liability owed to active participants divided by the number of active participants. Tenure is understated because the LEHD does not capture the complete history of an employer-employee relationship when states enter the dataset after a given employee-employer relationship is established. P-values for the difference between treatment and control groups are obtained by regressing the statistic of interest on a indicator variable for treatment status and clustering standard errors at the firm-level.

Table 1.2: Structural model parameter estimates

Parameter		Estimate	Standard error
IES	σ	0.958	0.004
Labor disutility persistence	ρ	0.869	0.003
Labor disutility standard deviation	σ_v	0.108	0.0005
Labor disutility age slope	ϕ	0.046	0.0001
Labor disutility constant	γ	-2.59	0.006
χ^2 statistic, 17 d.f.		236.8	

Notes: This table shows preference parameter estimates and standard errors for the structural model. See Appendix A.5 for details.

Table 1.3: Employment elasticity estimates

		Age		
		58	60	62
Intertemporal elasticity	$\hat{\eta}$	0.180	0.154	0.302
		(0.044)	(0.040)	(0.057)
Option value elasticity	$\hat{\eta}^{PV}$	1.17	0.92	1.77
		(0.279)	(0.238)	(0.322)

Notes: This table shows two different definitions of the employment elasticity for older workers. Standard errors are computed using the delta method. See Section 1.7.4 for definitions.

Table 1.4: Counterfactual effects of OASI tax sunset at age 60

	Current law				Reform				Difference
Variable	p25	p50	p75	Mean	p25	p50	p75	Mean	Mean
Retirement age	62	64	66	64.6	64	65	67	65.7	1.1
DC wealth at retirement ($)	130,040	184,790	277,880	225,575	133,870	189,060	286,970	233,518	7,943
PV income tax remitted after 60 ($)	62,230	81,329	94,662	79,380	83,153	103,343	121,465	102,540	23,160
PV payroll tax remitted after 60 ($)	22,569	43,692	63,251	48,290	14,844	18,236	24,605	19,992	-28,298
					p25	p50	p75	Mean	
Equivalent variation ($, PV at age 60)					50,620	67,941	92,874	75,134	

Notes: Estimates are derived from model simulations under current law and under the reform. Simulated workers in the sample are all employed at age 55. Estimates are conditioned on not having retired before age 60. Monetary estimates are reported in 2010 dollars and present values are computed as of age 60. See text for details.

CHAPTER II

Bad Times, Bad Jobs? How Recessions Affect Early Career Trajectories

2.1 Introduction

The state of the business cycle at labor market entry has substantial and persistent effects on the earnings and career trajectories of young workers.[1] Because of the strong element of chance associated with labor market entry during a recession, the popular press has labeled cohorts with the misfortune of being exposed to years of earnings losses as "unlucky."[2]

Initial wage losses for unlucky cohorts stem from the fact that availability of high-wage jobs is strongly procyclical (see, e.g., Okun, 1973 and McLaughlin and Bils, 2001); however, the translation of short-term fluctuations in wages into long-term scarring effects is a product of several different factors. First, search frictions can hinder the movement of workers between employers thereby extending the duration of recession-induced losses. Second, when these frictions rise with tenure, they can generate long-term human capital mismatch if recession entrants do not find work in jobs, occupations, and industries for which they have already specialized (see, e.g., Oreopoulos et al., 2012). Third, employers may be slow to learn about the true quality of recession entrants relative to expansion entrants because of greater initial mismatch. Finally, the fact that much of the labor market is characterized by long-term wage-setting rather than a spot market slows down the convergence between recession entrants and expansion entrants (see, e.g, Beaudry and DiNardo, 1991).

In this paper, we make two advances relative to the literature. First, we provide a precise estimate of the importance of employer-specific and non-employer-specific factors

[1]See, e.g., Kahn (2010) and Altonji et al. (2016) for evidence from the United States and Oreopoulos et al. (2012) for evidence from Canada.

[2]See, e.g., during a recession: Catherine Rampell, "Many With New College Degree Find the Job Market Humbling," *New York Times*, May 8, 2011. In contrast, during an expansion: Ben Casselman, "This Year's College Grads Are The Luckiest In A Decade," *FiveThirtyEight*, May 6, 2016.

in explaining the long-term impact of recessions on the wages of labor market entrants. While prior studies have found that employer and occupational characteristics play a role in explaining recession-induced penalties (see, e.g., Oyer, 2006, Oyer, 2008, Oreopoulos et al., 2012, and Rinz, 2019), the precise, quantitative role employers play in generating scars for recessionary entrants remains unknown. Our approach relies on the two-way fixed effect wage decomposition developed in Abowd et al. (1999) (AKM), which we use to partition recession-induced wage losses into components that are employer-specific and non-employer-specific. This exercise provides a more concrete comparison of between- versus within-employer explanations of the long run consequences of entering the labor market during a recession relative to existing literature.

Second, and perhaps more importantly, we provide the first evidence regarding how non-pecuniary compensation changes for unlucky cohorts relative to lucky cohorts. We estimate the value of working for each employer in our dataset by implementing a revealed preference-based estimator of job utility developed in Sorkin (2018). Within this framework, non-pay amenities are measured as variation in employer-specific pay holding employer value fixed, whereas rents accruing to workers are measured as variation in employer-specific pay that is explained by employer value.[3] From a welfare perspective, this exercise facilitates a more holistic accounting of recession-induced scarring because utility losses depend not only on losses in pay but also on changes to non-pay amenities (see, e.g., Rosen, 1986).

Our analysis uses linked employer-employee administrative data from Germany and studies the trajectories of new graduates of vocational training programs. Over two-thirds of the German workforce holds a vocational training degree, making our results relevant for a large proportion of the labor market—including a variety of skill types, occupations, and industries. Using these data, we establish three central findings. First, we show that the broad story of recession-induced scarring exists even in an economy with strong active labor market programs for employment and re-training. The typical recession in Germany lowers wages for new entrants by 4.9 percent cumulated over a 10 year horizon.[4] Second, we show that 1.9 percentage points (40 percent) of the total loss is explained by workers matching to lower paying firms. The remaining 3 percentage points (60 percent) of the total loss comes from non-employer-specific factors, including human capital mismatch, slow market-wide employer learning, and infrequent wage renegotiation. Our results therefore indicate that

[3] Under the revealed preference approach, job utility combines all unobserved amenities including such factors as job security, hours flexibility, commuting convenience, etc.

[4] The magnitude of recession-induced losses for unlucky cohorts in Germany is approximately equal to that found in Canadian data. Oreopoulos et al. (2012) find that a typical recession results in a 5 percent loss of earnings cumulated over 10 years. Estimates for recession-induced losses on unlucky cohorts in the United States are substantially larger. For each percentage point increase in the unemployment rate, Kahn (2010) finds a 6-7 percent loss in wages that decays to 2.5 percent after 15 years.

the majority of recession-induced losses in pay come from factors that are not specific to employers but rather a product of the labor market in general.

Finally, we show that 1.5 percentage points of the 1.9 percentage point employer-specific-pay penalty is compensated for by non-pay amenities whereas only 0.4 percentage points reflect losses from rent sharing. Thus, fully three-fourths of the employer-specific pay penalty is explained by higher non-pecuniary compensation. After netting out relative gains in non-pay amenities, the cost of recessions for young German labor market entrants drops by 30 percent from 4.9 percent of wages cumulated over 10 years to 3.4 percent. These estimates indicate that the welfare costs of labor market entry during a recession are overstated by a non-trivial margin if evaluated using only pecuniary losses.[5]

Using rich data on workers and establishments, we then assess the mechanisms that drive disparate career trajectories in the face of differing entry conditions. We show that low-pay, high-amenity employers are more likely to make job offers during recessions. Young workers in unlucky cohorts are beholden to the offer distribution they face upon entry, and their set of options are tilted towards towards employers that offer amenities in lieu of pay. This finding is consistent with prior research that links compensating differentials to unemployment risk (see, e.g., Abowd and Ashenfelter, 1981). It is also consistent with the observation that sectors which feature cyclically stable labor demand, such as education or healthcare, are often associated with higher job satisfaction than sectors which feature cyclically sensitive labor demand, such as construction or finance.

In addition, we show that recessionary entrants are much less likely to work at the same employer, in the same occupation, or in the same industry for which they trained. Our findings therefore suggest that much of the wage penalty faced by unlucky cohorts owes to losses in employer-, occupation-, and industry-specific human capital, analogous to the experience of workers who face involuntary job loss.[6] By illustrating the importance of changes in pay and amenities over the business cycle, our results also provide new evidence for the view that wage gains associated with industry switching during expansions are partly driven by lower non-pay amenities.[7]

[5]It is important to add that our conclusion on welfare implications relates only to employer-specific utility and may not capture other welfare relevant consequences of young workers' exposure to adverse aggregate conditions. For instance, Maclean (2013) and Schwandt and von Wachter (2016) find evidence of worse health and increased mortality among unlucky cohorts in the United States.

[6]See, e.g., Neal (1995) emphasizing the importance of industry-specific human capital in displacement scarring. See von Wachter and Bender (2006) emphasizing the importance of firm-specific human capital especially in the context of German vocational trainees.

[7]McLaughlin and Bils (2001) conjecture that such a phenomenon is possible but do not verify it empirically.

The rest of the paper is organized as follows. Section 2.2 provides a simple theoretical framework that formalizes the role of labor market risk in generating compensating differentials and explains how cyclical shocks alter the relative composition of wages and amenities available in the labor market. Section 2.3 describes linked employer-employee data from the Institute for Employment Research (IAB) of the German Federal Employment Agency, and how we utilized this data to construct our variables of interest. Section 2.4 describes our empirical strategy and presents our main results along with robustness checks. Section 2.6 concludes.

2.2 Theoretical framework

This section outlines a simple model of compensating wage differentials generated by variation in unemployment risk across firms. The model formalizes one important channel through which workers who enter the labor market during recessions could obtain higher non-pay amenities relative to those who enter during expansions.

2.2.1 Setup

Consider an economy with homogenous risk-averse workers and two types of firms: type R firms and type S firms. The production function for each firm type $j \in \{R, S\}$ in period t is

$$y_{jt} = z_{jt}F(h,k), \;\; F_h > 0, F_k > 0, F_{hh} < 0$$

where z_{jt} is an firm-specific productivity draw and F is a production function whose argument h represents hours employed in production and whose argument k represents capital. $F_{hh} < 0$ simply operationalizes a diminishing marginal product of labor. From an employment perspective, type R firms are deemed to be risky whereas type S firms are deemed to be safe in the sense that:

$$\begin{bmatrix} z_{Rt} \\ z_{St} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \bar{z} \\ \bar{z} \end{bmatrix}, \begin{bmatrix} \sigma_R^2 & \rho_{RS} \\ \rho_{RS} & \sigma_S^2 \end{bmatrix}\right) \tag{2.2.1}$$

$$\rho_{RS} > 0 \tag{2.2.2}$$

$$\sigma_R^2 > \sigma_S^2 \tag{2.2.3}$$

In equation (2.2.1), the mean productivity level $\bar{z}$ represents the steady state. Although the shocks are positively correlated, type R firms exhibit greater variance in productivity than type S firms. A cyclical shock is defined to occur when both firm types obtain productivity draws that are jointly above or jointly below the steady state level. In recessions, $z_{jt} < \bar{z}$ and in expansions, $z_{jt} > \bar{z}$ for each j.

For any realization of firm-specific productivity shocks and for fixed capital stock $\bar{k}$, equilibrium labor demand, h, can be generically written using the firm's first order condition for profit maximization as

$$F_h(h^*_{jt}, \bar{k}) = \frac{w}{z_{jt}}, \tag{2.2.4}$$

where w is the equilibrium wage rate and h^*_{jt} is the optimal hours demand for the firm. $F_{hh} < 0$ guarantees that $\frac{\partial h^*_{jt}}{\partial z_{jt}} \geq 0$.

2.2.2 Wage determination

We assume that the labor market does not function as a spot market. Instead, as in Azariadis (1975) and Abowd and Ashenfelter (1981), workers and firms agree to long-term implicit contracts where wages are fixed but hours are variable. The key element of these long-term implicit contracts is that they insure risk averse workers against future labor market shocks that are propagated through firms' productivity draws.[8] The wage associated with each contract can be ascertained as follows. First, define the indirect utility for a worker who receives a wage w at firm j as $V(w, h_{jt})$, with utility an increasing, concave function of earnings, $U(w \cdot h_{jt})$. At a common wage rate, w, the expected indirect utility associated with a type R contract will be lower than the expected utility associated with a type S contract. This conclusion follows from the fact that z_{Rt} is drawn from a distribution that second order stochastically dominates z_{St} and that higher variability in productivity translates directly into higher variability of hours demand through (2.2.4). Consequently, risk aversion on the part of workers implies that

$$E[V(w, h_{Rt})] < E[V(w, h_{St})].$$

Figure 2.1 illustrates this gap in expected indirect utility over a range of potential hours realizations around the steady state level $\bar{h}$. Because R and S firms compete to obtain labor, workers demand a wage premium to compensate for the additional risk associated with employment at type R firms. Define the risk premium as the compensating wage differential $\Delta \equiv w_R - w_S > 0$ where w_R is the wage rate for a type R firm and w_S is the

[8] See Rosen (1985) and references therein on other models of risk premia in long-term implicit labor market contracts.

wage rate for a type S firm. The equilibrium wage rates are those which ensure that the expected indirect utility associated with employment at each firm type is equalized:

$$E\left[V(w_S + \Delta, h_{Rt})\right] = E\left[V(w_S, h_{St})\right].$$

2.2.3 Effect of cyclical shocks

Cyclical shocks re-order the expected relative productivity of each firm type. In expansions the following condition holds due to (2.2.1) and (2.2.3)[9]:

$$E[z_{Rt}|z_{Rt} > \bar{z}] > E[z_{St}|z_{St} > \bar{z}].$$

That is, risky firms are expected to be more productive than safe firms. Conversely, in recessions[10]:

$$E[z_{Rt}|z_{Rt} < \bar{z}] < E[z_{St}|z_{St} < \bar{z}].$$

That is, safe firms are expected to be more productive than risky firms. Combining the productivity shifts over the cycle with the labor demand equation (2.2.4) generates the following implications. First, expansions generate an increase in hours demanded by risky firms relative to safe firms. Thus,

$$E[h_{Rt}|z_{Rt} > \bar{z}] > E[h_{St}|z_{St} > \bar{z}]. \tag{2.2.5}$$

Second, recessions generate an increase in hours demanded by safe firms relative to risky firms. Thus,

$$E[h_{Rt}|z_{Rt} < \bar{z}] < E[h_{St}|z_{St} < \bar{z}]. \tag{2.2.6}$$

Finally, combining (2.2.5) and (2.2.6) shows us how the state of the economy affects hours demanded by risky *relative to* safe firms[11]:

$$\underbrace{|E[h_{Rt}|z_{Rt} < \bar{z}] - E[h_{Rt}|z_{Rt} > \bar{z}]|}_{\text{Drop in hours demanded by } R \text{ during recession}} > \underbrace{|E[h_{St}|z_{St} < \bar{z}] - E[h_{St}|z_{St} > \bar{z}]|}_{\text{Drop in hours demanded by } S \text{ during recession}}.$$

[9]Specifically, $E[z_{jt}|z_{jt} > \bar{z}] = \bar{z} + \sigma_j\sqrt{\frac{2}{\pi}}$ because z_{jt} is distributed normally for each j.

[10]Specifically, $E[z_{jt}|z_{jt} > \bar{z}] = \bar{z} - \sigma_j\sqrt{\frac{2}{\pi}}$.

[11]The same intuition holds in extensions with multiple firm types and firms with different average productivities (and therefore different levels of hours demand during steady state).

This simple framework illustrates two key implications for the observed wages of labor market entrants. First, the potential for cyclical shocks generates compensating wage differentials that are designed to indemnify workers against risk. Second, the state of the economy materially influences the share of hours demanded by risky firms relative to safe firms. This generates a composition effect in which recessions see relatively more hours demanded by safe firms compared to risky firms. Thus, we should expect labor market entrants to be relatively more likely to receive offers from low-wage/low-risk firms during a recession. We work toward an empirical test of this hypothesis in the following sections.

2.3 Data and construction of key variables

Our analyses are based off the Sample of Integrated Labor Market Biographies (SIAB) (vom Berge et al., 2013). These data comprise a longitudinal two percent random sample of all individuals in Germany that ever worked, claimed unemployment insurance benefits, or sought job seeking assistance. In total, the sample describes the labor market histories of just over 1.6 million workers starting in 1975 and ending in 2010.[12]

The SIAB is a linked establishment-worker dataset that is organized in terms of spells. For employed individuals, each spell enumerates a match between a given worker and a given establishment. For non-employed individuals, a spell can enumerate a period of unemployment benefit receipt or a period of participation in an active labor market program. Spell lengths are measured at daily precision. In addition to demographics, data on individuals include average daily wages, educational qualifications, occupation, and state of residence. Establishment information is available at annual frequencies and includes industry codes, location, size, and median wage rates.

Because vocational training earnings are subject to Social Security contributions, spells of young workers who are apprentices in Germany's vocational training system are fully enumerated in the SIAB. With precise information on the occupation, wage rate, and start and end dates of training, the SIAB provides us with an unusually detailed set of information on workers both before and after they graduate from vocational training. In addition to exploiting the linked worker-establishment nature of the data, we rely on the timing of labor market entry made possible by observing workers before and after their training is complete.

Before delving further into how we estimate the impact of business cycle shocks on the career trajectories of young German workers, we first discuss how we use the SIAB

[12]The data exclude employment in the civil service as well as self-employment. Marginal part time employment or so called "mini-jobs" are tracked in the SIAB starting in 1999. Mini-jobs are low-wage jobs with a monthly income threshold of 450 euro. Participation in active labor market programs is tracked starting from 2000.

to estimate variables that are critical inputs in our analyses. These variables include establishment-specific measures of wage premia, utility, compensating differentials, rents, and occupation-specific unemployment rates at the state level.

2.3.1 Wage decomposition

Following AKM, Card et al. (2013) (CHK) use the Integrated Employment Biographies (IEB)—which is the universe of data from which the SIAB sample is drawn—to estimate person and establishment fixed effect components of daily average wages. For worker i employed at establishment j in year t, the decomposition can be summarized as

$$\log(\text{wage}_{it}) = \alpha_i + \psi_j \mathbf{1}\{i \text{ works at } j \text{ in } t\} + \boldsymbol{x}'_{it}\beta + r_{it} \tag{2.3.1}$$

Person fixed effects, α_i, incorporate individual specific skills that are rewarded equally across employers. The establishment fixed effect, ψ_j, is a proportional premium that is paid by establishment j to all its employees. $\boldsymbol{x}_{it}$ is a vector of unrestricted year dummies as well as quadratic and cubic terms in age fully interacted with educational attainment. These controls account for aggregate and life-cycle determinants of wages. Consistency of the parameter estimates requires that the error term, r_{it}, is uncorrelated with α_i, ψ_j and the $\boldsymbol{x}_{it}$. CHK provide a detailed discussion about the validity of the identifying assumptions.

The CHK person and establishment fixed effects are identified only within the connected set of establishments—i.e., the set of establishments that either hire from or lose workers to other establishments in the set. CHK find that the largest connected set encompasses 95 percent of establishments in the IEB. Because the SIAB is a two percent random sample of worker histories drawn from IEB, the set of connected employers is relatively small. Fortunately, the original CHK person and establishment fixed effects are provided as supplements to the SIAB dataset, precluding the need for analysts to re estimate (2.3.1) within a sparsely connected dataset.

2.3.2 Estimating establishment values

Much of the literature that builds on the AKM decomposition treats the establishment fixed effect as a measure of economic rents shared by workers. However, a long tradition in economics has posited that employer-specific components of pay (ψ_j) can vary not only because of factors such as rent sharing or efficiency wages, but also because of amenities that are priced in the labor market as compensating differentials.[13]

[13]See, e.g, Rosen (1974) and Rosen (1986) for theory, and Lucas (1977), Freeman (1978), and Brown (1980) for empirical evidence.

Building on this tradition, Sorkin (2018) proposes a novel methodology to estimate employer-specific utility, exploiting the voluntary movements of workers between employers in order to infer the relative utility of each employer, which is also referred to as employer value. Implementing this revealed preference argument requires three key assumptions. First, all workers have the same ex-ante preferences over jobs. Second, all jobs within an employer are deemed to be identical from the standpoint of non-pay characteristics. Finally, all workers—both employed and non-employed—search randomly from the same offer distribution.

Taking these assumptions to linked employer-employee data, Sorkin develops an estimator that aggregates the choices of workers into unique establishment values, the utility an employee derives from working at a particular establishment. Intuitively, the estimator rewards employers for making more hires from other high-quality employers and penalizes them for voluntary departures. Akin to the connectedness requirement in AKM and CHK, values are only calculable within the *strongly connected* set of employers. Strong connectivity is defined as a set of employers who both gain *and* lose workers to other employers in the set.

Because the establishment values are estimated using a revealed preference argument, a crucial step in this procedure is to separate voluntary from involuntary movements of workers. This distinction is required both for employer-to-employer movements (EE moves) and movements from employment to non-employment (EN moves).[14] Sorkin's methodology relies on the notion that workers who separate from establishments that are shrinking are more likely to be involuntarily displaced, whereas workers separating from growing establishments are more likely to be voluntary departures. Thus, comparing the rate of worker exit from shrinking and growing establishments provides a benchmark for the probability that a given move from a shrinking establishment is involuntary.[15] We mimic this procedure in our establishment value estimation.

Unlike Sorkin (2018), whose analyses are based on the universe of U.S. workers covered by unemployment insurance in the Longitudinal Household Employer Dynamics (LEHD) data, we rely on the set of establishments that employ workers in our two percent longitudinal sample. Differences across these data settings necessitate additional modifications to Sorkin's methodology. First, while he limits strong connectivity to firms linked only by EE flows,

[14]Note that unemployment and labor force non-participation are taken to be the same for the purposes of this estimation procedure.

[15]The rate of worker separation from growing establishments represents "expected" turnover that is fueled by worker quits. Then, the rate of worker separation from shrinking establishments that is above and beyond this "expected" rate disciplines the probability that a given worker move from a shrinking establishment is involuntary. The methodology thus assumes that all moves from growing establishments are voluntary. The likelihood of voluntary exit is separately estimated for EE and EN moves.

we expand the set to include employers linked both by EE flows as well as transitions of workers through non-employment. Because of the ubiquity of movements in and out of non-employment, this modification expands the scope of strong connectivity and makes the computation of values feasible even within a two percent random sample like the SIAB.[16] This modification is fully consistent with the estimating equations in Sorkin's model which allow non-employment to obtain its own relative value.[17] Second, when coding EE and EN transitions, Sorkin reduces the quarterly LEHD data where workers potentially have multiple employers in a given year, into a data set where each worker is associated with a single employer—also known as the annual dominant employer—in a given year. In the SIAB, job spell lengths are measured at daily precision, allowing us to take advantage of a more complete set of job-to-job transitions. When a worker has two or more overlapping job spells in a given year (i.e. works for multiple employers at the same time), we select the spell associated with the highest total earnings. We treat periods between jobs that are longer than 90 days as non-employment spells regardless of whether an individual received unemployment benefits or sought job seeking assistance. Third and finally, unlike Sorkin, we do not impose any earnings or age restrictions on workers in our sample in order to maximize the number of establishments that we can include in the strongly connected set.

2.3.3 Decomposing establishment-specific fixed effects into rents and compensating differentials

Within the framework of the utility posting job-search model he proposes, Sorkin assumes that the value of being employed at employer j (V_j) can be written as an additively separable function of the employer-specific component of pay and an employer-specific non-pay amenity:

$$V_j = \omega(\psi_j + a_j). \tag{2.3.2}$$

where ω is utility per log euro, ψ_j is the employer-specific component of pay, and a_j is the employer-specific non-pay amenity. Using establishment value estimates and CHK establishment fixed effect estimates, we re-arrange equation (2.3.2) to estimate

$$\psi_j = \pi V_j + \epsilon_j \tag{2.3.3}$$

[16] We rely on the MATLAB package `MatlabBGL` provided by David Gleich to find the largest strongly connected set in the SIAB.

[17] Put differently, workers who make EN and NE transitions help to identify both the estimate of non-employment value as well as estimates of employer values.

and then obtain the residual terms $\hat{\epsilon}_j = \psi_j - \hat{\pi} V_j$. Because the residuals are orthogonal to V_j by construction, they capture components of a_j that generate variation in pay holding utility constant. As such, they correspond to the non-pecuniary amenities defined as compensating differentials in Rosen (1986). The fitted value, $\hat{\pi} V_j$, is an estimate of rents accruing to workers because it captures variation in pay that is correlated with utility.[18]

Figure 2.2 shows the relationship between studentized values and establishment fixed effects in the SIAB replicating a Figure 5 from Sorkin (2018), but with German data. The blue circles plot average establishment value and average fixed effects within each ventile of the establishment value distribution. The red lines show one standard deviation bands of the establishment fixed effect distribution within each value quintile. The upward slope of the line of best fit shows that workers typically value employers that pay more, thereby indicating the importance of rents. Nevertheless, there is a wide spread of establishment-specific pay variation conditional on a given level of establishment utility, thereby indicating the presence of compensating differentials.[19]

Figure 2.3 plots average values and establishment fixed effects across firms within broad industry categories. The line depicting the overall relationship between V_j and ψ_j across firms is the same as that in Figure 2.2, but the scatter plot across industries gives us further information about which sectors are associated with high pay and/or high amenities. As expected, sectors associated with higher pay and low amenities (above the line) include Mining, Construction, and Finance. Meanwhile, establishments in the Health & Social Work sector impart roughly the same value to their employees as Manufacturing establishments despite paying substantially lower, on average. This indicates that Health & Social work establishments offer more non-pay amenities to their workers, consistent with our prior beliefs about relative work satisfaction in these sectors. In total, we view Figure 2.3 as important qualitative validation that the values we have estimated have economic meaning.

[18] The CHK establishment fixed effects are estimated separately by gender, giving us two observations per establishment. Due to sample size limitations, we pool together worker flows of both genders when estimating establishment values. Thus, we additionally include a dummy variable for gender on the right-hand side of equation (2.3.3) to remove average differences in establishment-specific pay between men and women.

[19] Our estimates of establishment value are based off a 2 percent sample of workers and therefore embody measurement error. As a consequence, the slope of the line of best fit shown in Figure 2.2 is attenuated. In Appendix B.2, we use a split-sample instrumental variable (IV) approach to evaluate the quantitative impact of measurement error in our analyses. We find that OLS-based estimates of equation (2.3.3) are indeed attenuated relative to the IV estimates. However, correcting for this bias has no economically substantive impact on our key findings. Given that $\hat{\epsilon}_j$ (compensating differentials) and πV_j (rents) are ultimately used as outcome variables, this is to be expected as long as the measurement error is classical.

2.3.4 Occupation-specific unemployment rates

In order to capture variation in business cycle conditions relevant to the young workers we study, we take advantage of the SIAB's extensive data and large sample size to estimate two customized unemployment rates. Our preferred measure is the state- and occupation-specific unemployment rate. This measure provides an effective representation of labor demand, especially for workers entering the labor market in a given state with training in a given occupation. In addition, we calculate the occupation-specific national unemployment rate. This unemployment rate provides a broader measure of labor demand that allows us to test whether our results are driven by endogenous migratory responses.

Each unemployment rate is calculated using a similar methodology. We first assign every worker in our dataset a status of employed, unemployed, or out of the labor force on the 15th of each month using the SIAB's labor market status variable.[20] We assign employed individuals the occupation of their current job and unemployed workers their last known occupation. For employed workers who are currently in multiple jobs with multiple occupations, we assign them the occupation of the job that paid higher daily wages. Using the state of residence variable included in the SIAB, we then aggregate these observations into monthly, occupation-specific unemployment rates. Finally, we aggregate monthly unemployment rates into yearly unemployment rates by averaging across months and weighting monthly rates by the underlying number of observations associated with each month.

The primary value of using the SIAB to construct unemployment rates is to better characterize the state of the business cycle specific to young workers as they enter the labor market. Given that the data come directly from administrative sources, they also have the advantage of capturing the the unemployment rate as well as, or better than, survey-based unemployment rates. Nonetheless, to demonstrate that unemployment rates calculated directly from the SIAB match publicly-available unemployment rate measures for Germany, we present a comparison between our estimates of the national unemployment rate from the SIAB and two measures provided by the Organization for Economic Co-Operation and Development (OECD) in Figure 2.4. The first OECD unemployment measure is survey-based while the second is the registered unemployment rate, calculated from administrative data by counting the number of individuals who register with the government as unemployed in order to receive unemployment benefits, auxiliary benefits like community assistance and health assistance, or to signal the need for assistance with job search. The administrative data-based OECD measure more closely aligns with the SIAB-based estimates in levels

[20]The SIAB provides detailed information about worker status that we aggregate to three simple categories.

because both use registration for unemployment benefits rather than survey self-reports to ascertain unemployment.[21]

The trends of these three measures track each other closely, which is the relevant check for our panel data-based analysis. Furthermore, at the sub-national level, a within-state regression of survey-based OECD unemployment rates on unemployment rates estimated from the SIAB produces a coefficient of 0.978 (t =16.39).[22] In sum, while the OECD does not report occupation-specific unemployment rates for direct comparison, we are confident that our measure identifies relevant changes to business cycle conditions. Meanwhile, constructing the unemployment rate with administrative data allows us to define an individual's labor market based on both occupation and state, which more accurately represents the conditions they face when entering the labor market.

2.3.5 Sample construction

An important characteristic of the SIAB is that it distinguishes between employment spells associated with vocational training from those that are not. Since this distinction is crucial for the implementation of our identification strategy, our most important sample restriction is to only consider workers whose first spell in the SIAB is in vocational training. We make four additional restrictions to hone in on the career paths of young trainees during the first ten years of their labor market experience. First, we limit our sample to those who complete their training before turning 30. Second, in order to keep comparisons consistent over time, we restrict our sample to April 1999 and later, after which the SIAB began enumerating mini-job work spells.[23][24] Third, we drop observations in occupation-state-years from which unemployment rates were computed with 20 or fewer observations. Finally, in order to avoid shifting composition based on outcome variables, we restrict the analysis

[21] The difference in levels between the two OECD sources arises for two reasons. First, non-employed survey respondents may indicate that they are job-seekers even though they are not registered as unemployed (if, for example, they are not eligible for unemployment benefits and thus see no advantage in registering). Second, survey respondents who have registered to receive unemployment benefits may indicate that they are not seeking a job when responding to the survey (if, for example, they register primarily to get auxiliary benefits such as health or community assistance). The latter group outweighs the former in our study period and likely also accounts for the fact that unemployment rates estimated from the SIAB are higher than the survey-based measure. In addition, unemployment rates obtained from the SIAB are higher in levels because these data exclude civil servants and self-employed workers who likely exhibit lower rates of joblessness than the rest of the labor force.

[22] The specification is $U_{st}^{\text{OECD survey-based}} = \alpha + \rho U_{st}^{\text{SIAB}} + \theta_s + \varepsilon_{st}$ where s is a state. Registered unemployment rates are not available at the state level from the OECD.

[23] Data drawn before and after this period are difficult to compare because of the change in enumeration.

[24] CHK estimate Equation (2.3.1) using different time windows in the IEB. We rely on estimates from the 2002-2009 time window and extrapolate these backward to 1999 and forward to 2010 to cover our sample window.

dataset to those individuals who are employed at establishments where establishment fixed effects and values are available, when they are employed. As noted in Sections 2.3.1 and 2.3.2, this restriction amounts to analyzing workers who are employed by establishments in the strongly connected set.

In order to align the timing of the analysis, we assign each worker a dominant employer for each year. The dominant employer is defined as the employer that pays a given worker the most in earnings for a calendar year. We obtain earnings for a given job spell by multiplying daily wages by spell length, then sum within employer to determine the total earnings from each employer for a given worker within a year. The employer with the highest earnings is given "dominant" status, and it is this employer's value, daily wage, and other associated characteristics that are used as outcomes below.[25]

2.4 Identification

Our identifying assumptions exploit key features of Germany's apprenticeship training system. Before proceeding to the empirical analysis, we first provide some institutional background on how the system works and why it lends itself to our estimation strategy.

Apprenticeship training in Germany is also known as dual vocational training because it combines workplace and classroom training in a roughly 60-40 split. The typical young worker begins her vocational training after secondary schooling by starting an employer-sponsored apprenticeship in one of approximately 350 officially recognized occupations. Employers, unions, and government agencies jointly regulate the course content and program length associated with training in each of the occupations to meet quality standards. Trainee wages are set by collective bargaining agreements which vary both by state and by occupation (Kuppe et al., 2013). Apprenticeships are the most common form of higher education in Germany, with over two-thirds of the workforce holding a vocational training degree.

Two aspects of the German apprenticeship system are particularly important for our analyses. First, occupational segmenting of the German labor market is a natural consequence of a system that is designed to promote occupation-specific skills among young workers. This feature makes between-occupation heterogeneity a more important dimension of youth labor market sorting in Germany than the United States, for example. Second, the duration of training programs are regulated, with the typical course taking about 3 years and culminating in a qualifying examination. The pre-set training duration makes it less

[25]In cases where there is more than one spell with the dominant employer, the daily wage is a weighted average across spells where the weight is the number of days.

likely that young workers can selectively enter the labor market when cyclical conditions are favorable. Even so, we use the mode within detailed training occupation to measure expected, rather than observed, training duration. This allows for a reduced form strategy in which we assign individuals a date of labor market entry based on expected training duration rather than their true labor market entry date, which could be subject to limited manipulation. Year of entry is thus defined by the training start date plus the modal training duration time in a given occupation.

Given this institutional setting, our identifying assumption is that occupation-specific unemployment rates prevailing at the time of expected labor market entry are unrelated to the timing of labor market entry. Given that we assign the timing of labor market entry based on the modal training duration within an occupation, this assumption rests on the notion that individuals do not change their training occupation or the timing of their training *start* based on labor market conditions that manifest (usually 3) years later. This broad assumption encompasses two important points: first, workers with particular unobserved characteristics cannot manipulate their initial labor market conditions through selective entry; and second, that the introduction of a particularly poor cohort in terms of unobserved variables is not responsible for adverse aggregate conditions.

Under this framework, we estimate the effect of initial aggregate conditions on labor market outcomes by exploiting variation in the occupation-state-cohort (*osc*) specific unemployment rate U_{osc} using the following specification:

$$y_{it} = \beta_e U_{osc} + \Gamma \mathbf{X}_i + \theta_{s(i)} + \theta_{o(i)} + \theta_{c(i)} + \theta_{e(i)} + \theta_t + \varepsilon_{it} \tag{2.4.1}$$

In equation (2.4.1), i represents an individual, $\theta_{s(i)}$ is a vector of state of training fixed effects, $\theta_{o(i)}$ is a vector of training occupation fixed effects, $\theta_{c(i)}$ is a vector of year-of-expected-entry (training start + modal training duration in the detailed occupation) fixed effects, $\theta_{e(i)}$ is a vector of potential experience (year minus year of expected entry) fixed effects, θ_t is a vector of year fixed effects, and $\mathbf{X}_i$ is a vector that contains a dummy for the individual's gender, a dummy for whether the individual is a German citizen, and a vector of fixed effects for the individual's age at start of training.[26] We use β_e to trace out the average effect of initial labor market conditions U_{osc} on outcomes y_{it} for the first ten years of young workers' careers, where career start is defined by predicted training end. We cluster standard errors at the occupation-state-cohort level.

Table 2.1 provides initial evidence for the validity of our identifying assumptions by showing that workers are similar in terms of age, training wages, nationality, gender,

[26] As in Oreopoulos et al. (2012), we identify $\theta_{c(i)}$, $\theta_{e(i)}$, and θ_t by dropping an extra year fixed effect.

and successful completion of apprenticeships whether they enter the labor market when the unemployment rate is below or above the median of a given occupation-state cell. To the extent that they are economically meaningful, any differences in these observed characteristics are eliminated because we include them in the conditioning set $\mathbf{X}_i$.

To provide further evidence for our identification strategy, we formally test whether young workers are able to speed up or delay entry based on aggregate labor market conditions by re-estimating Equation (2.4.1) using training duration as the outcome:

$$[\text{Training Duration}]_i = \beta U_{osc} + \Gamma \mathbf{X}_i + \theta_{s(i)} + \theta_{o(i)} + \theta_{c(i)} + \theta_{e(i)} + \varepsilon_{it} \tag{2.4.2}$$

We do this under two scenarios: in the first, $c(i)$ is an individual's actual year of entry, and in the second, $c(i)$ is expected year of entry based on modal training duration within occupation—our preferred measure. Table 2.3 shows the results from this validation exercise. Two key points emerge. First, there appears to be some ability for workers to manipulate training duration based on the business cycle, but on average, this ability is very small. When entry is defined by the last day of an individual's training, a one standard deviation increase in U_{osc} generates 12 day increase in training duration, on average. In contrast, when entry is defined by training start date plus the modal training completion time in an occupation, there is no relationship between U_{osc} and training duration. We take this as evidence that our reduced form strategy corrects a small endogeneity bias relative to assigning individuals their true date of training completion.

2.5 Effect of cyclical shocks on pecuniary and non-pecuniary compensation

2.5.1 Primary results

The four panels of Figure 2.5 show the effect of a 1 percentage point increase in U_{osc} at expected labor market entry on daily wages, establishment fixed effects, rents, and compensating differentials over 10 years of potential experience. The top left panel illustrates that young workers face an initial wage loss of about 0.6 percent, a gap that steadily narrows over the next 10 years. The wage losses incurred in these first 10 years are economically meaningful: our estimates imply that a one standard deviation increase in U_{osc} at the time of labor market entry induces a 4.9 percent present discounted value (PDV) loss in daily wages cumulated over the next decade.[27] Assuming away differences in earnings arising from

[27]To conduct this calculation, we use the mean daily wage at each potential experience year, $\bar{w}_e$, the standard deviation of U_{osc}, $\sigma_U = 7.34$, a discount rate $r = 0.05$, and the β_e coefficients in the following

the number of days worked, our PDV daily wage loss estimate from Germany is similar to the 5 percent earnings loss accrued over 10 years for the average Canadian recession graduate estimated in Oreopoulos et al. (2012).[28]

Establishment fixed effects, shown in the top right panel, drop by about 0.2 percent on impact, and slowly rise. Taken on its own, the establishment fixed effect result hides important dynamics that would be invisible were they not decomposed into rent and compensating differential components. The effect of adverse entry conditions on these sub-components of wages are shown in the lower two panels. The initial gap in rents is small and closes steadily over time, with no statistically significant that it still exists after 10 years. In contrast, non-pay amenities (or negative compensating differentials) rise by about 0.15 percent on impact and then slowly converge by year 10. These results imply that unlucky workers start their careers in lower-paying jobs that feature higher non-pay amenities compared to their lucky counterparts. Over time, these workers are successful in transitioning to higher paying establishments and closing the gap in rents. However, they continue to work at establishments that provide a relatively higher share of non-pay amenities until year 10.

2.5.2 Career paths of "lucky" and "unlucky" cohorts

Figure 2.6 illustrates the career paths of young workers in absolute rather than relative terms. Establishment value (V) is shown on the horizontal axis and establishment fixed effects (ψ) are shown on the vertical axis. The lines show the first 10 years of career trajectories in (V, ψ) space for two different initial conditions, obtained by plotting the predicted values of V and ψ from Equation (2.4.1) for each level of potential experience using counterfactual values of U_{osc}. The "Low U" group faces initial unemployment rates at the 10th percentile of within occupation-state unemployment rates, while the "High U" group faces initial unemployment rates at the 90th percentile. Lastly, the gray dashed lined plots a fitted line estimated from Equation (2.3.3) showing the average relationship between establishment fixed effects and establishment values for employers in the SIAB.

Both "Low U" and "High U" groups move to the northwest, gaining in value and in employer specific pay on essentially the same line—an approximation of the early-career job ladder in (V, ψ) space. Workers who face low initial unemployment rates begin their careers

---

expression: $100 \times \left(1 - \frac{\sum_{e=0}^{10}\left[\bar{w}_e(1+\sigma_U\beta_e)/(1+r)^e\right]}{\sum_{e=0}^{10}\left[\bar{w}_e/(1+r)^e\right]}\right)$.

[28] Oreopoulos et al. (2012) construct their estimate assuming that recessions induce a 5 percentage point change in the regional unemployment rate. During our study period (1999-2010), the regional unemployment rate in Germany exhibited a standard deviation of 4.9 percentage points.

at higher paying, higher value firms. However, a part of this pay gap comes from lower non-pay amenities (this is evidenced by the fact that the “Low U” career path is further above the dashed line than the “High U” career path in any given year). Both types of workers retain the higher-pay/lower-amenity trade off as they gain experience, but “High U” workers start from a lower level. In general, the career path has a higher slope than the average relationship between establishment fixed effects and establishment values indicating that individuals trade non-pay amenities for pecuniary compensation as they move up the job ladder.

2.5.3 Decomposition of recession-induced losses

We next decompose recession-induced losses into four major categories in Table 2.4.[29] Appendix B.1 provides details on how we calculate the various estimates presented in the table. As mentioned above, the total pecuniary loss for cohorts who enter the labor market during a recession is estimated as a 4.9 percent reduction in the present value of wages, cumulated over a decade.

The subsequent rows of the table illustrate various novel decompositions that we are able to calculate by applying unique data and empirical techniques. The first row shows that 1.9 percentage points, or 40 percent, of the total loss is attributable purely to the fact that unlucky cohorts match with lower paying employers. These employer-specific losses are further split into rents and amenities in the second and third rows of the table. The 1.9 percentage point reduction in employer-specific pay arises from a 0.4 percentage point reduction in rents and a 1.5 percentage point gain in amenities. Thus, fully three-quarters of employer-specific pay reductions are compensated for by relative gains in amenities. The final row of the table shows that 3 percentage points, or 60 percent, of the total loss is not explained by employer-specific factors. This 60 percent incorporates losses due to factors such as human capital mismatch, changes in outside offers, slow market-wide learning, and infrequent wage re-negotiation.

Taken together, the results shown in Figure 2.5, Figure 2.6, and Table 2.4 present new perspective about the effects of cyclical conditions on the early career outcomes of young workers. The welfare cost of recession entry, when viewed only in pecuniary terms, is a 4.9 percent reduction in the present value of wages. However, when we account for the fact that 1.5 percentage points of that loss is compensated for by relative gains in non-pay amenities, the overall impact of typical recession is effectively a 3.4 percent reduction in the present value of wages. Consequently, using pecuniary measures alone overstates the overall loss by

[29]These estimates reflect a one standard deviation increase in unemployment rates at entry which is intended to simulate a recession.

about 30 percent. In other words, a purely pecuniary comparison between the two cohorts presents an incomplete picture about the welfare losses imposed by initial labor market conditions.

2.5.4 Mechanism 1: cyclical variation in hiring by establishment type

Having illustrated the major implications of cyclical shocks on young workers' career trajectories, we now turn to investigate the mechanisms that drive the results we discussed above. The first of these mechanisms is based on the model we presented in Section 2.2 which explicitly incorporates a compensating differential for employment stability (lack of unemployment hazard) into the competitive equilibrium wage rate. In this framework, some establishments are less cyclical in their hiring and firing decisions, and can therefore offer lower equilibrium wages. The overall composition of hiring establishments would tilt toward these stable establishments during a contraction, decreasing the average establishment fixed effect measured for contractionary entrants and increasing the average measured compensating differential. This divergence across firms could occur for a variety of reasons, including both cross- and within-sector differences in the cyclicality of productivity.

Here, we investigate whether firms with higher measured non-pecuniary amenities are less cyclical in their hiring by relating establishment growth to the business cycle, stratified by an establishment's measured time-invariant amenity value, a_j. The regression model generally takes the form

$$g_{jst} = \alpha + \gamma_U(U_{st}) + \gamma_a(a_j^{CD}) + \gamma_{int}(U_{st} \times a_j^{CD}) + \alpha_s + \alpha_t + \varepsilon_{jst} \tag{2.5.1}$$

where g_{jst} represents establishment growth between year t and $t+1$, U_{st} is the state-year unemployment rate, and a_j^{CD} is a studentized version of $-\hat{\epsilon}_j$ from Equation (2.3.3) — our measure of non-pecuniary amenities that generate variation in pay holding utility constant. γ_{int} is our key coefficient of interest, measuring the extent to which establishments with higher non-pecuniary amenities are more ($\gamma_{int} > 0$) or less ($\gamma_{int} < 0$) cyclical in their hiring.

The results from estimating Equation (2.5.1) can be seen in Table 2.5. They offer evidence that establishments that provide higher non-pay amenities are less cyclical in their net employment growth. This suggests a direct link between the amenities we measure and an establishment's hiring decisions: those establishments that are more likely to hire during a recession naturally offer more job security, and this job security is itself an amenity. Workers are naturally more likely to receive job offers from more stable establishments (and industries) when they enter the labor market during an economic contraction, which naturally increases their non-pecuniary compensation and decreases their wage rate.

2.5.5 Mechanism 2: displacement-induced human capital mismatch

We further use the rich structure of the SIAB to uncover patterns of worker mobility that underpin the results from Section 2.5. The four panels of Figure 2.7 summarize these mechanisms by showing the impact of a one percentage point increase in U_{osc} at labor market entry on the probability of remaining in one's training industry and the probability of remaining in one's training occupation. As with the previous results, we show these impacts over a 10 year horizon.

Figure 2.7 indicates that adverse initial conditions have the effect of displacing newly trained workers from their training industries and training occupations. Because training employers are important contributors to wage growth in the German context, the displacement we see in these figures suggests that there are strong parallels between labor market entry during adverse cyclical conditions and the unemployment scar that follows involuntary job loss (e.g. Jacobson et al., 1993, Davis and von Wachter, 2011, and Krolikowski, 2017). Furthermore, mismatch in industry- and occupation- specific human capital engendered by this displacement likely explains some of the wage penalties, which also has analogs in the literature that proposes mechanisms for earnings losses following job loss (e.g. Jarosch, 2015 and Krolikowski, 2017).

However, focusing purely on pecuniary losses as much of the prior literature has, would have led one to conclude that the cumulative effect of these mobility scars is overwhelmingly negative. In contrast, we find that affected cohorts recover completely in terms of rents and values even though they are more likely to work outside of the industries and occupations in which they trained. These displacements also shed light on the nature of cyclical upgrading in the career paths traversed by lucky and unlucky cohorts. For example, McLaughlin and Bils (2001) note that the wage gains obtained by workers who enter high-wage industries during expansions could, in fact, be attributable to a compensating differential channel.[30] While they do not seek to empirically verify this possibility, the patterns we present here provide evidence both of its existence and of its magnitude. Taken together, the evidence on outcomes and mechanisms suggests that the overall costs of industry and occupational displacement may not be as damaging as one might have been led to conclude in the absence of richer measures of compensation.

2.5.6 Robustness checks

In this section we investigate the robustness of our identification strategy to a variety of potential threats. We consider the effect of potential simultaneity that arises from estimating

[30] McLaughlin and Bils (2001) focus on a Roy-model and a queuing model to explain cyclical upgrading.

values as well as regression coefficients with the same underlying sample of worker mobility. Next, we examine the extent to which connectedness restrictions required for the estimation of establishment fixed effects and values generate selection bias in our sample. We find that our results remain robust to these concerns.

2.5.6.1 Simultaneity of value and regression model estimation

A potential concern about our empirical strategy comes from the fact that identification of the β_e coefficients in Equation (2.4.1) exploits some of the same worker moves that identify establishment values as described in Section 2.3.2. If rents and compensating differentials, dependent variables in Equation (2.4.1), are estimated using the same underlying source of variation as the β_e coefficients, then our estimates could exhibit simultaneity bias. In particular, the establishments that employ the workers in our sample could have higher values precisely because the workers in our sample chose to work there.

In order to purge our econometric analysis of such a bias, we re-estimate the establishment values by dropping all individuals who were trainees as of April 1999 or later from the set of worker moves. The results of this exercise are presented in Figure 2.8. The upper panel shows results based on values estimated from the restricted sample, whereas the lower panel duplicates the results from Figure 2.5 based on values estimated from the full sample. It appears that there is some scope for upward bias in estimated coefficients from Equation (2.4.1) on value-related outcomes. However, this bias does not alter the basic patterns of our results: workers who face worse initial aggregate conditions recover in terms of rents but not in terms of compensating differentials eight years later.

2.5.6.2 Sample selection bias

As discussed in Section 2.3.1 and 2.3.2, our analysis is restricted to the set of establishments for which establishment fixed effects and establishment values are estimable. The first restriction requires connectedness and the second requires strong connectedness. While these restrictions allow us to estimate Equation (2.4.1) using a variety of outcomes for a fixed sample, we want to ensure that our results are not influenced by endogenous sample selection. To do so, we take advantage of the fact that daily wages are available for all individual-establishment observations in our data, regardless of whether the establishments are in either of these connected sets.

Figure 2.9 contrasts our Figure 2.5 results for log wages from the restricted sample in the left panel to an unrestricted sample of all young workers in our data in the right panel.[31]

[31] These estimates are obtained using actual year of entry and U_{osc} (our preferred specification from Section

Comparing the two panels, we see that adverse initial labor market conditions have a very similar effect on log wages across the restricted and unrestricted samples, starting from a 0.6 percent dip for each percentage point increase in U_{osc} and recovering to around 0.4 percent by the eighth year of an individual's post-training career. Thus, while our analysis is confined to the 115,673 individuals represented by the left panel, this test provides good evidence that their labor market experience reflects the broader early career paths of young German workers.

2.6 Discussion and conclusion

This paper provides a new perspective on the costs imposed by cyclical shocks on young workers. Using rich employer-employee linked data from Germany we replicate the major finding of existing research showing that adverse cyclical conditions at labor market entry generate persistent wage losses on affected workers. Implementing a revealed preference based method of measuring compensating differentials across a large sample of establishments, we find that low wages for unlucky cohorts are explained by reduced rents and by offsetting compensating differentials, with the latter accounting for the majority. We offer suggestive evidence that employers hiring workers during downturns provide more long-term job security than those who hire workers during expansions, which explains part of the offsetting compensating differential that unlucky cohorts obtain. Our findings indicate that focusing on earnings losses alone overstates the long-term welfare losses of labor market entry during recessionary periods.

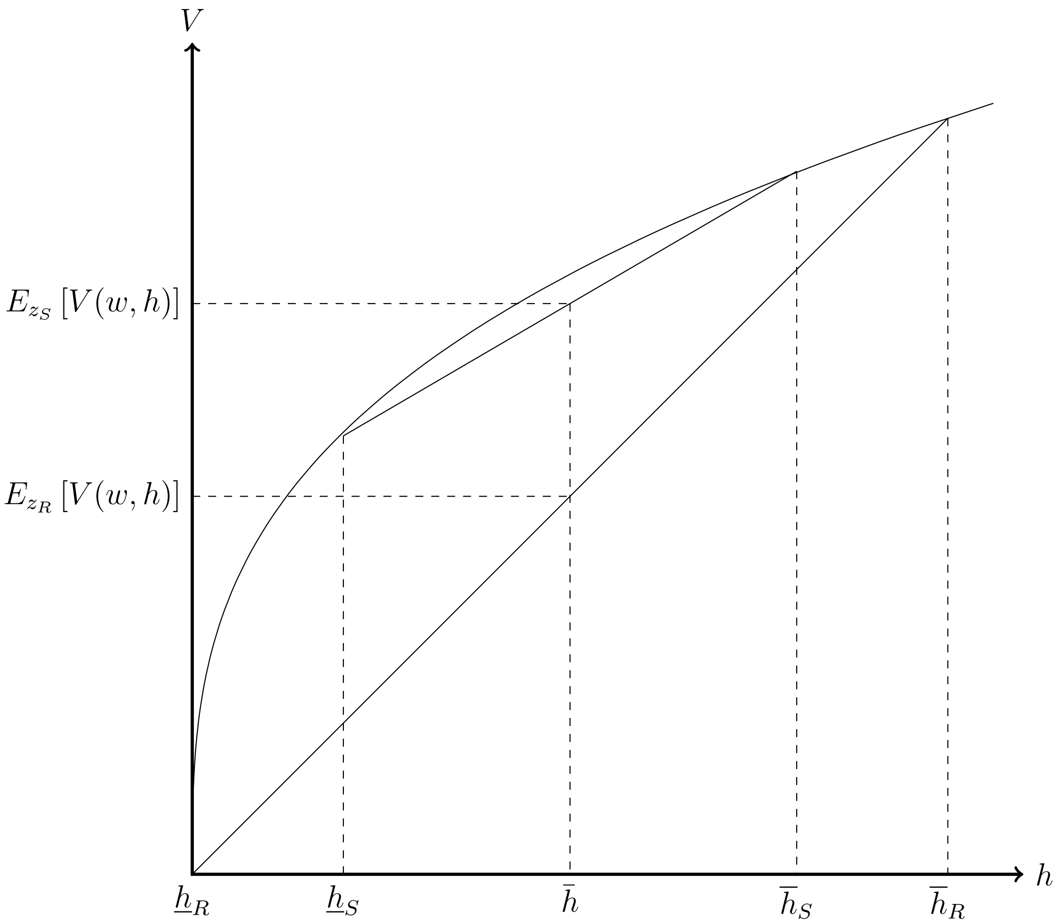
We find that cyclical shocks in Germany displace workers away from the occupations and industries in which they gained specialized apprenticeship training. As such, the wage losses that unlucky workers face is explained by the mismatch in human capital similar to that experienced by workers who suffer involuntary job loss. Nevertheless, our results indicate that these unlucky workers continue to climb the utility ladder of job quality by accruing returns in the amenity rather than rent component of establishment-specific pay. These findings shed new light on the role that compensating differentials play in cyclical upgrading and downgrading.

The sample of newly trained workers that we study includes a wide variety of skill types, broadening the relevance of our results beyond college graduates. Nevertheless, our conclusions are specific to the German labor market which differs substantially in terms of employment protections, unemployment benefits, healthcare provision, and re-training

2.5), but the comparative results are qualitatively similar within any of the robustness checks discussed in this section.

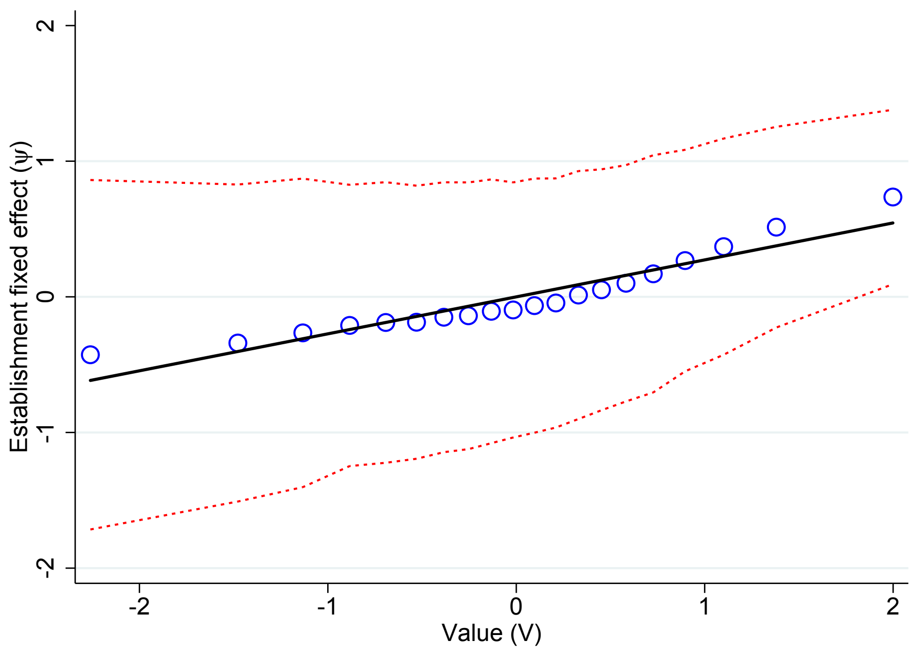
programs relative to the United States and Canada which are the two other countries in which recession-induced wage losses have been studied. Furthermore, the utility consequences that we focus on are isolated to job specific components and do not speak to factors such as health or consumption.

Figure 2.1: Compensating wage variation due to hours volatility



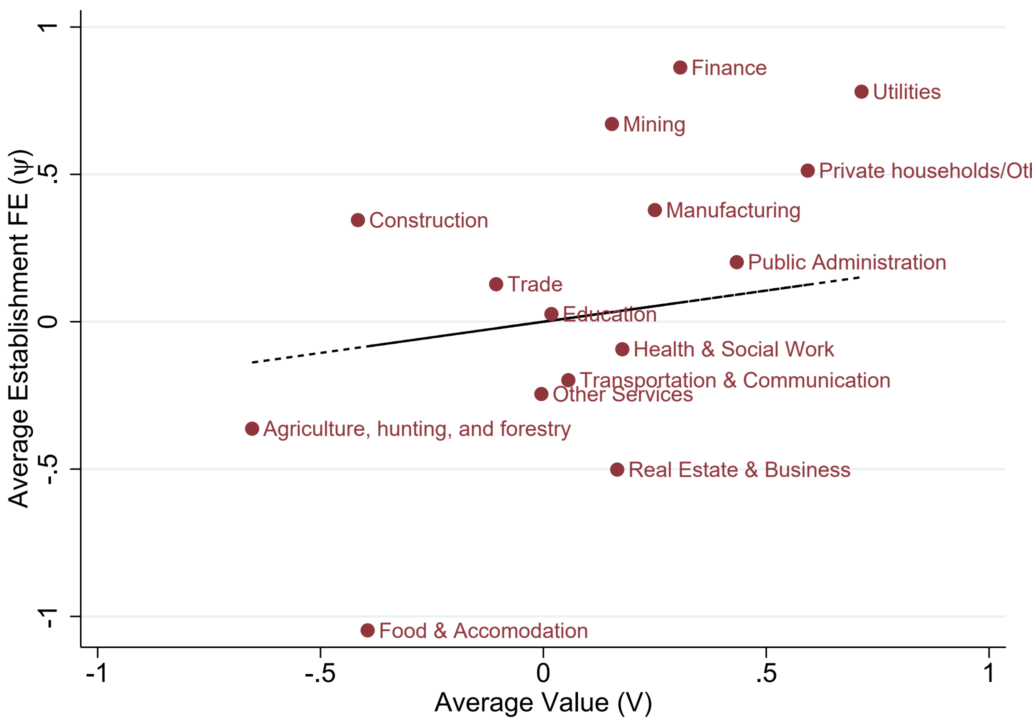
Notes: Thresholds illustrated on the horizontal axis represent the bounds of potential hours demand at risky (R) and safe (S) firms. $\bar{h}$ is the steady state level of hours demand associated with productivity $\bar{z}$. Underbars represent average values of h conditional on $z < \bar{z}$. Overbars represent average values of h conditional on $z > \bar{z}$.

Figure 2.2: Relationship between establishment values and compensation



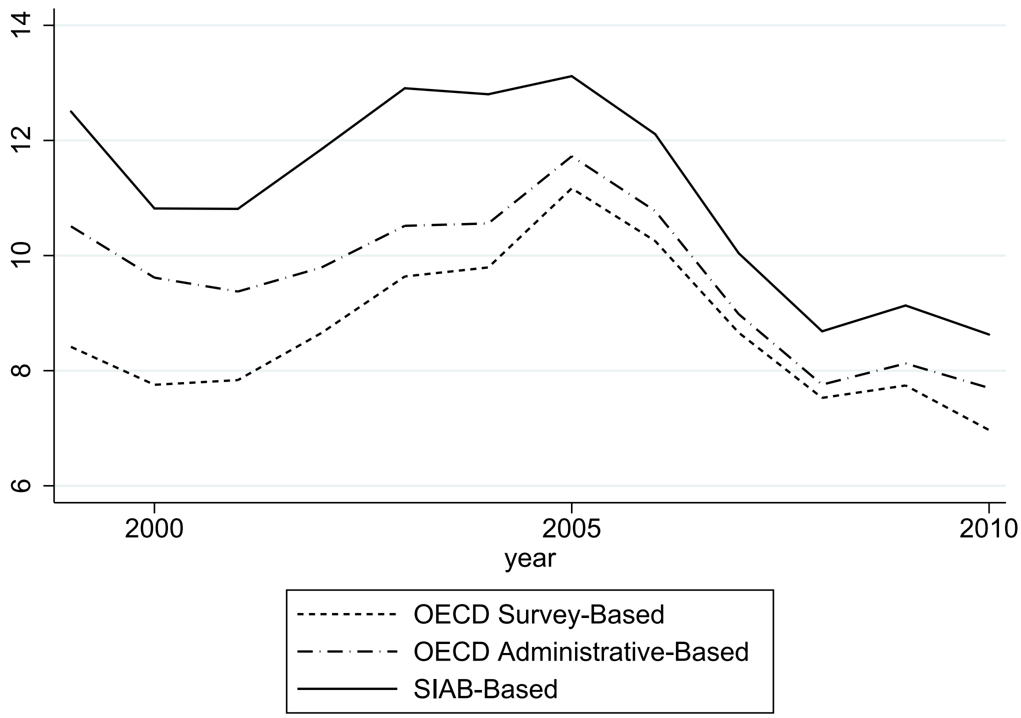
Notes: The value and establishment fixed effect estimates shown in this graph are studentized. The blue circles show the average establishment fixed effect and average establishment value within each ventile of the establishment value distribution. The red lines show one standard deviation bands of the establishment fixed effect distribution.

Figure 2.3: Relationship between establishment values and compensation across industries



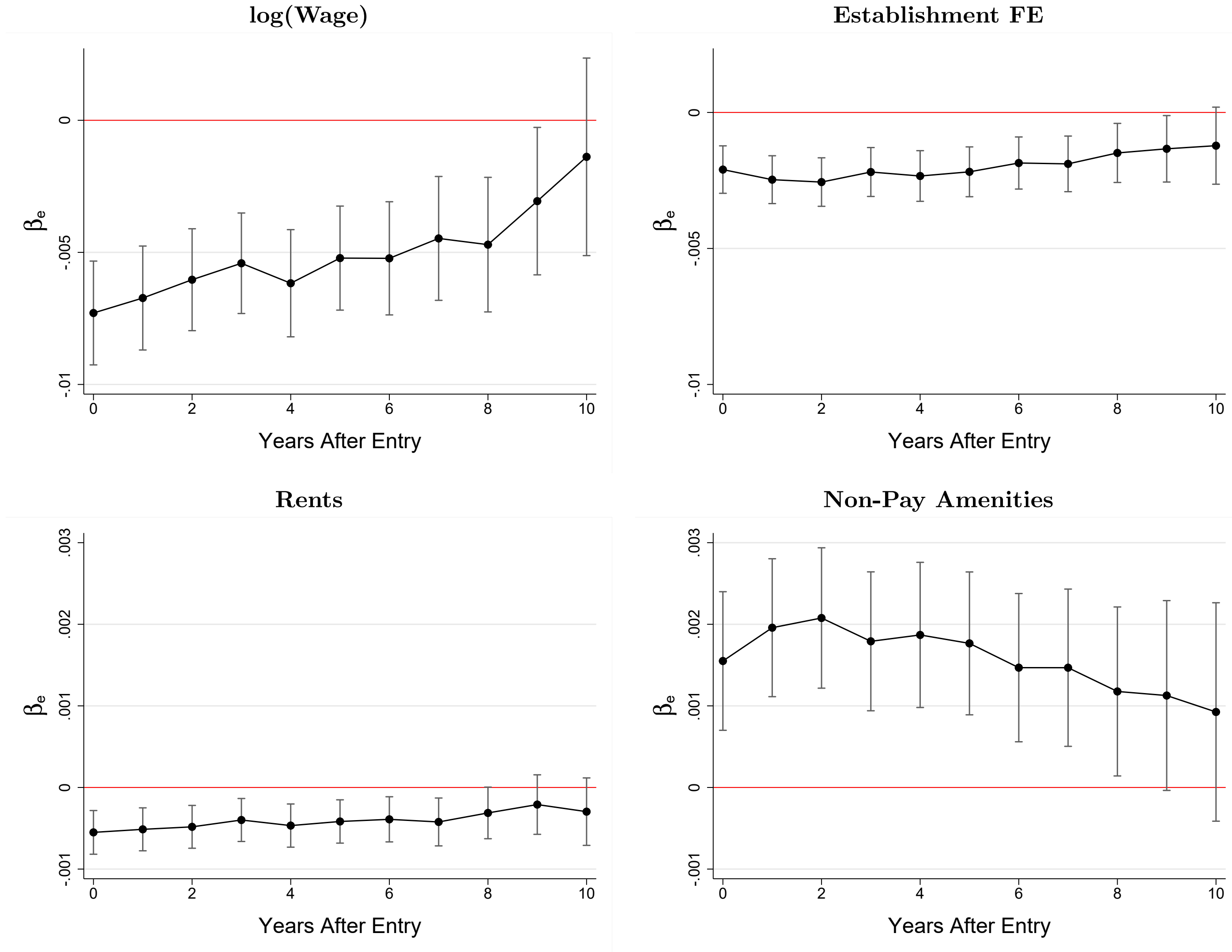
Notes: The value and establishment fixed effect estimates shown in this graph are studentized.

Figure 2.4: OECD versus SIAB national unemployment rates



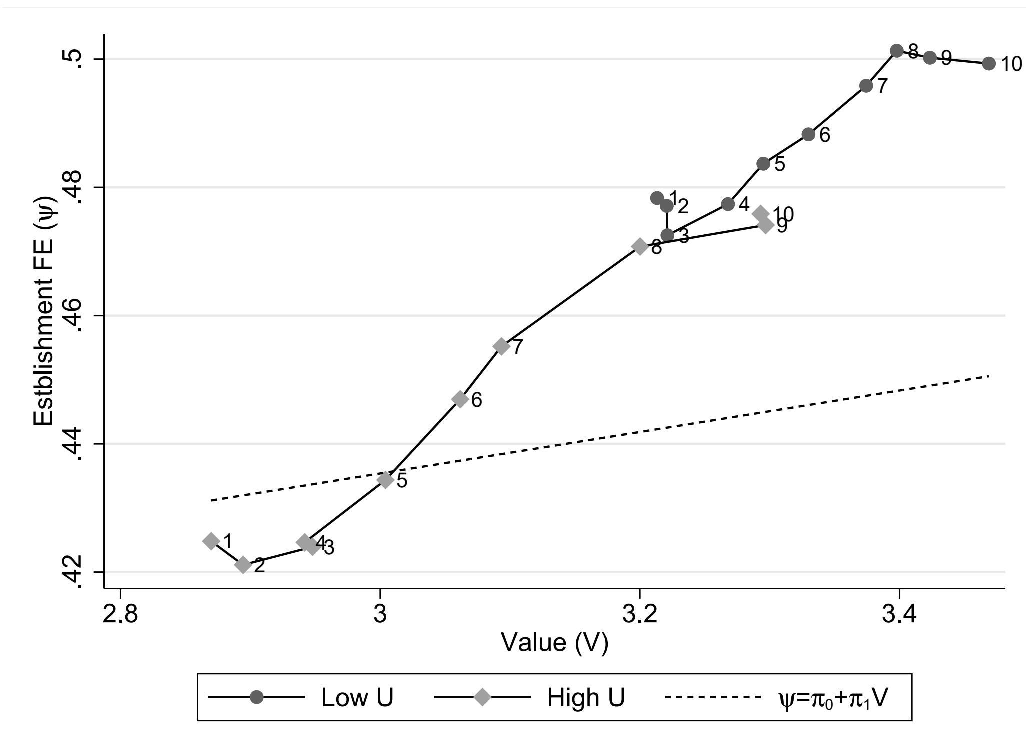
Notes: OECD unemployment rate obtained from OECD Stat. See Section 2.3.4 for details of of how unemployment rates are estimated using the SIAB.

Figure 2.5: The effect of entry conditions (U_{osc}) on early career outcomes



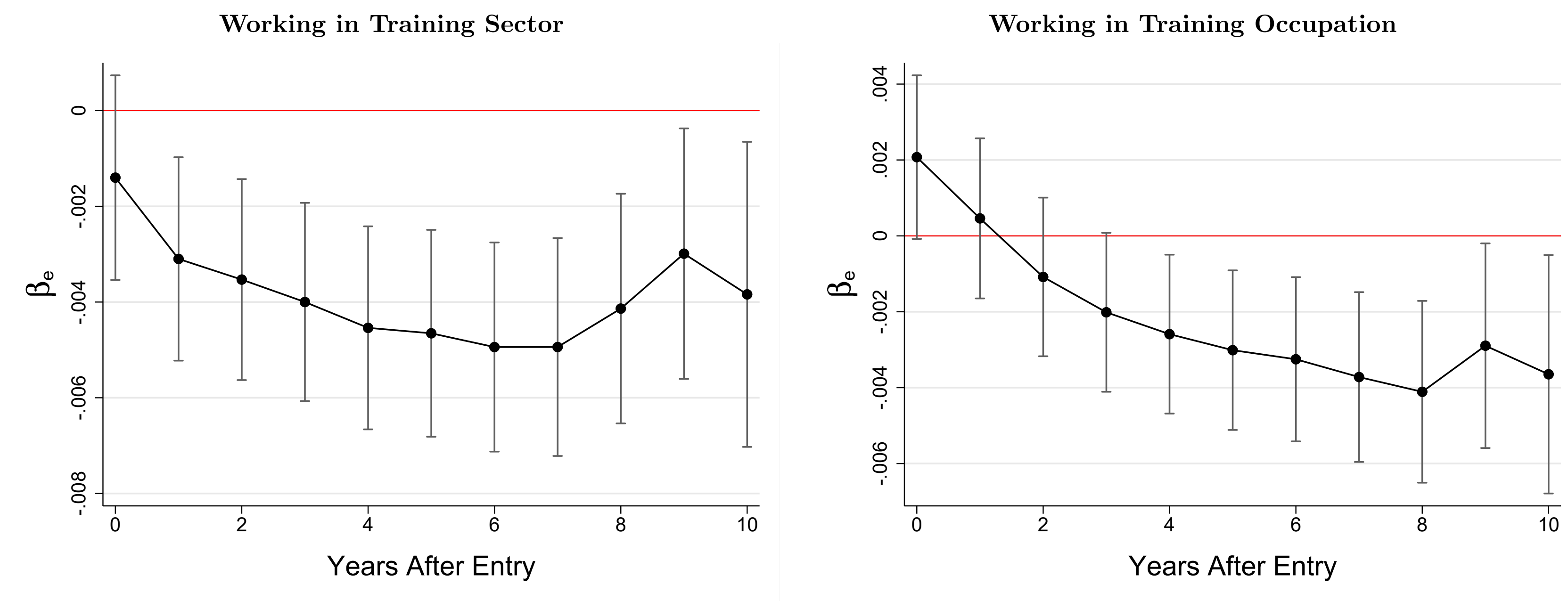
Notes: See Equation (2.4.1). All estimated coefficients are in log wage units. Sample size for each specification is 115,673. 95% confidence intervals represented by bars, with standard errors clustered at the state-cohort-occupation level. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Figure 2.6: Implied career paths of young workers



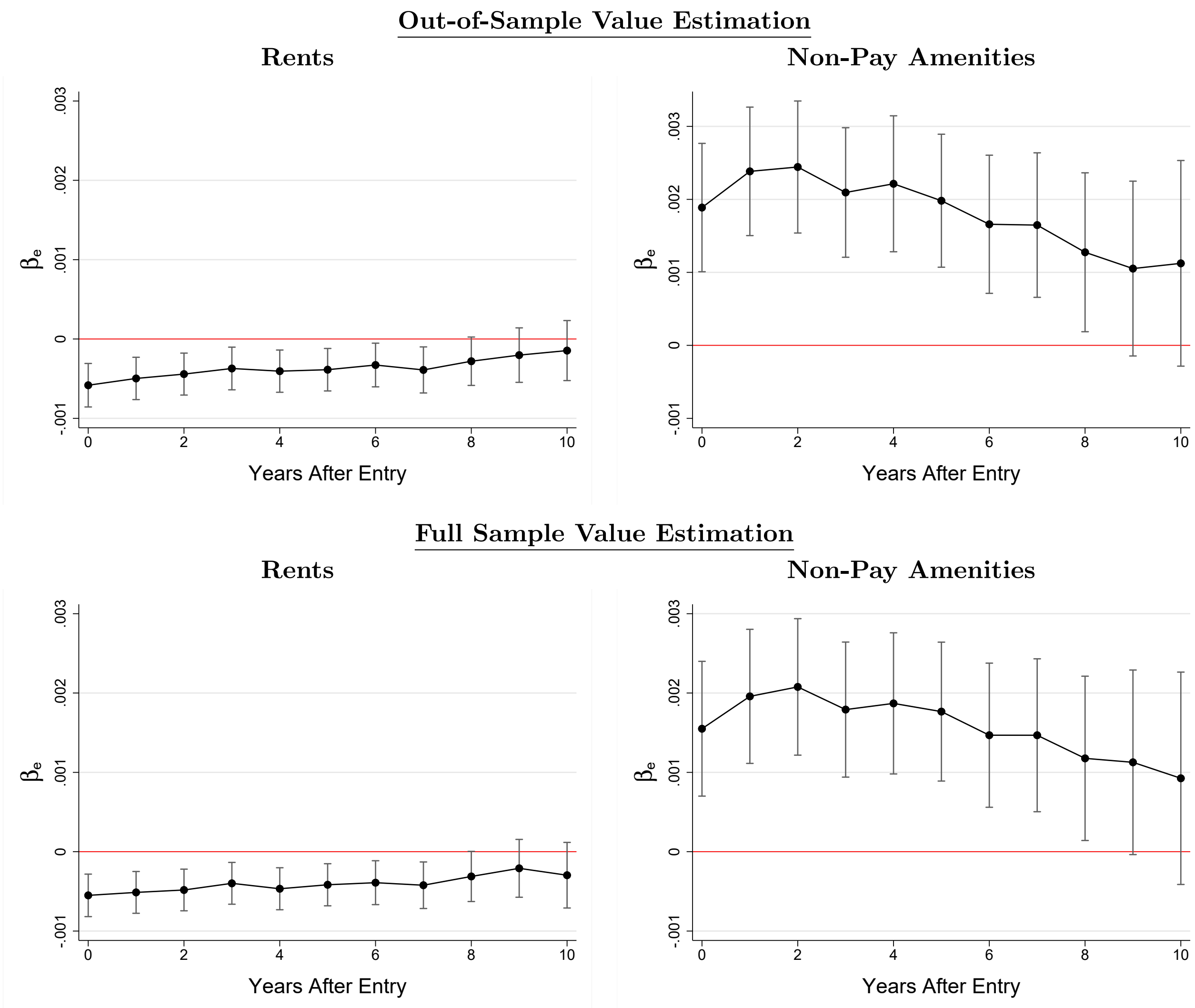
Notes: See Section 2.5 for details.

Figure 2.7: The effect of entry conditions (U_{osc}) on early career mobility



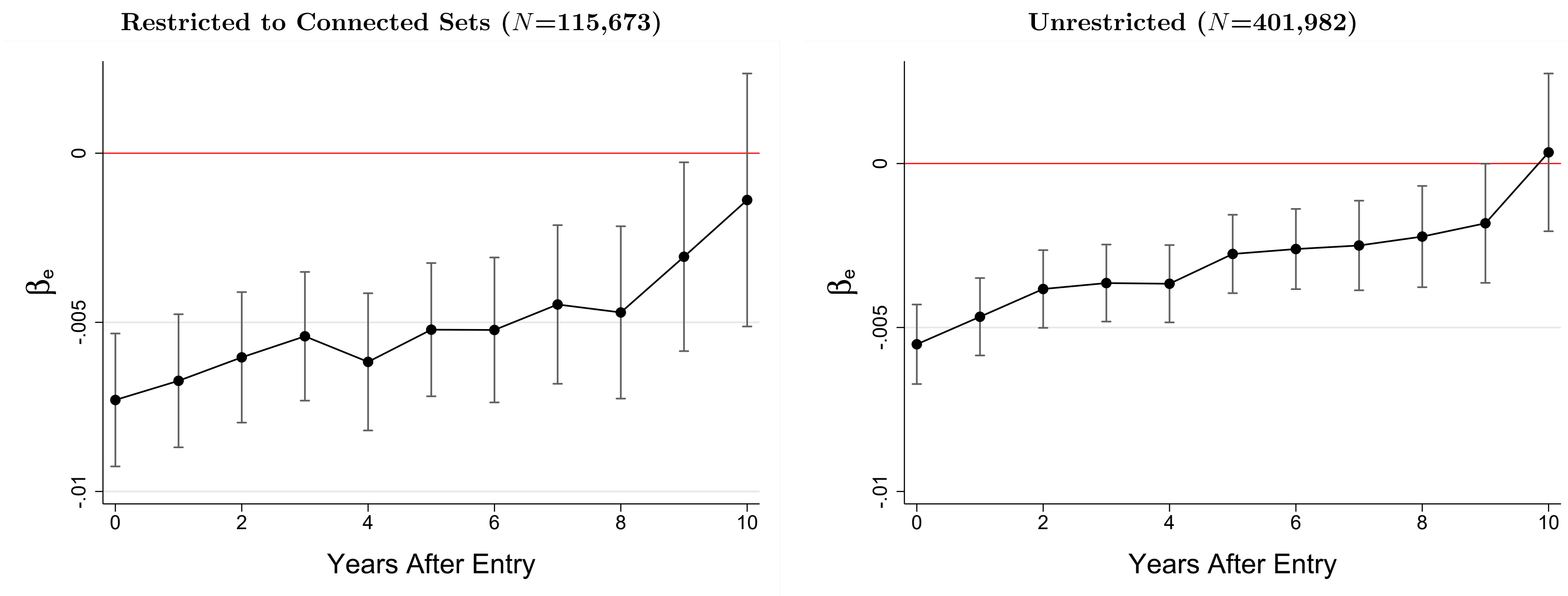
Notes: See Equation (2.4.1). All estimated coefficients are in log wage units. Sample size for each specification is 115,673. 95% confidence intervals represented by bars, with standard errors clustered at the state-cohort-occupation level. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Figure 2.8: Simultaneity bias robustness



Notes: Sample size for "Out-of-Sample Value Estimation" figures is 119,988. Sample size for "Full Sample Value Estimation" figures is 125,363. 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Figure 2.9: The effect of entry conditions (U_{osc}) on log wages—restricted versus unrestricted sample



Notes: 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German national indicator variable, and a female indicator variable.

Table 2.1: Mean individual characteristics by aggregate entry conditions

	$U_{osc} < U_{os}^{MED}$	$U_{osc} \geq U_{os}^{MED}$
Age at start of training	17.4	17.3
Median wage at training firm	87.9	86.2
Share German	0.94	0.95
Share female	0.41	0.43
Obtained vocational training certification (=1)	0.79	0.78

Notes: U_{os}^{MED} refers to the median occupation-state-specific unemployment rate during the study sample, 1999-2010. Median wage at training firm refers to median wage at individual's training firm in the year of training completion.

Table 2.2: Identifying variation—unemployment rates faced by young workers at time of entry

				Percentile				
Independent Variable	Description	Mean	St. Dev.	10	25	50	75	90
U_{osc}	State-Occupation Unemployment Rate	10.88	7.34	4.05	5.56	8.85	14.20	21.05
U_{oc}	National Occupation Unemployment Rate	9.94	3.80	6.18	7.67	9.57	11.24	12.86

Notes: Occupation assigned to individuals based on the occupation of their training apprenticeship. Unemployment rates calculated as described in Section 2.3.4.

Table 2.3: Identification strategy—validation tests

	Training Duration (yrs)	Training Duration (yrs)
U_{osc}	0.0044**	-0.0015
	(0.0021)	(0.0030)
Year of Entry Definition	Last day of training	Predicted last day of training based on occ. mode

Notes: See Equation (2.4.2). * $p<0.1$ ** $p<0.05$ *** $p<0.01$.

Table 2.4: Decomposition of recession-induced losses

	Δ PDV wage (10 years)
Losses in employer-specific pay	1.9%
Due to losses in rents	0.4%
Due to relative gains in non-pay amenities	1.5%
Other losses	3.0%
Total wage loss	4.9%

Table 2.5: Establishment growth over the business cycle

	(1) Establishment Growth	(2) Establishment Growth	(3) Log Establishment Size
U_{st}: State-Year Unemployment Rate	-0.0050***	-0.0051**	-0.0079***
	(0.0012)	(0.0024)	(0.0018)
a_j^{CD}: Non-Pay Amenities	0.0019	—	-0.0035
	(0.0054)	—	(0.0069)
$U_{st} \times a_j^{CD}$	0.0017***	0.0088**	0.0015**
	(0.0005)	(0.0011)	(0.0007)
Establishment FE	No	Yes	No
Lagged Log Establishment Size	No	No	Yes

Notes: All models include state and year fixed effects. a_j^{CD} is studentized.
* p<0.1 ** p<0.05 *** p<0.01.

CHAPTER III

Finding Needles in Haystacks: Multiple Imputation Record Linkage Using Machine Learning

3.1 Introduction

Increasingly, researchers are interested in linking survey and administrative data for measurement and analysis. In most record linkage applications, the units being linked originate from the same frame. For instance, individuals in a given dataset are linked to same individuals in a different dataset, or businesses in one dataset are linked to the same businesses in another dataset. In this paper, we consider the problem of linking across frames. We match individual respondents in household survey data to administrative data on the universe of employers. How does one use a household report of a business to link to the correct employer? It would be possible to build in these linkages from the start, especially where a sampling frame is created from administrative data. In that case, linkage is part of the design. This paper addresses the problem of linking individuals and employers where the linkage is not pre-designed into a survey. This situation typically arises in surveys of households, which are built from sampling frames of household addresses, often without the purpose of linkage as part of the design. Even in an idealized world where the survey and administrative frames were developed in tandem, additional linkages to other administrative data, that are not part of the design, may be desirable.

In the absence of unique identifiers, we treat record linkage as a missing data problem where true match status is unknown and must be imputed. To do this, we need to accomplish two related tasks. First, we need a way to predict whether any given pair of records drawn from the two datasets constitutes a true match. Second, we need a way to characterize uncertainty in the prediction of true matches and propagate that uncertainty into inferences drawn from the linked data set by subsequent analyses.

The task of predicting true match status is difficult because the size distribution of firms is very skewed. Consider the striking empirical fact that 0.3 percent of all firms employ 53 percent of all workers in the United States.[1] There are however millions of firms, so the flip side of this fact is that most firms are very small. With so many small firms, matching individuals to employers is inherently noisy because a large number of small employers among a set of potential candidates are feasible matches for any given respondent. This is our needle in the haystack problem. For example, imagine a set of candidate matches included a large insurance company, an independent credit union at the same location using the name of the insurance company, and a cafeteria operated by a third-party vendor also at the same location and using the name of the insurance company. The names and address of the candidate matches are all similar. When presented with this information, a human reviewer may use auxiliary information such as the respondents industry, occupation, or reported firm size to guess the correct match. We automate and speed-up what a human reviewer would do by using a supervised machine learning (ML) approach to predict the matching firm. ML is particularly valuable for record linkage because it makes flexible use of a very large number of predictors, including auxiliary information, to mimic the heuristics used by a human reviewer. Furthermore, relying on a rich set of predictors lends support to the assumption that imputation errors are ignorable thereby improving inferences in subsequent analyses of the imputed data. Finally, our ML estimator is tuned to deliver high out-of-sample accuracy which improves the quality of predictions beyond the sample used for estimation.

To propagate uncertainty about true match status, we rely on multiple imputation (MI). For each record in household survey data, our procedure samples multiple candidates from administrative employer-level data by using ML-based match probability estimates as weights. In cases where the match probability estimates are highly concentrated, household survey records are linked to just one employer. Conversely, for cases where the match probability estimates are highly diffuse, household survey records are linked to many different employers. In the completed (matched) dataset, variability between implicates for a given household survey record captures uncertainty associated with the linkage for that household. Subsequent analyses of the completed dataset can then combine the multiple implicates for valid statistical inference as in Rubin (1987).

The plan of this article is as follows. Section 3.2 describes record linkage methodologies in deterministic and probabilistic contexts. Section 3.3 provides details on the files that we attempt to link and explains the three major steps of our record linkage procedure. Section 3.4 assesses the fit of our imputation model and evaluates the accuracy of the linkage. Section 3.5 compares selected employer characteristics derived from HRS self-reports to

[1] 2017 Statistics of U.S. Businesses (SUSB), U.S. Census Bureau.

BR imputations obtained using a variety of imputation methods. Section 3.6 illustrates an application of the matched data to shed new light on the incidence of nonclassical measurement error and nonresponse bias in household survey respondent's reports of employer and establishment size. Section 3.7 concludes.

3.2 Foundations of Record Linkage

This paper builds on an important literature that developed widely-used techniques for record linkage. The simplest approaches are non-probabilistic. In these deterministic file matching applications, researchers accomplish record linkage by isolating a set of variables that are common to a given record in both files. This procedure constitutes both the first and the last step in the linkage. It is the first step because it enumerates the set of possible matches. It is the last step because only those records that have exactly one match conditional on variable agreement are retained. In some instances, a sufficiently rich set of accurately measured variables can allow a large fraction of the original file to be unequivocally matched (see, e.g., Warren et al. (2002), Hammill et al. (2009), Lawson et al. (2013), and Setoguchi et al. (2014)). In other cases, the matched file consists of a small and potentially non-random subset of the original file that limits the usefulness of the matched dataset for analysis. This concern is highlighted in the context of linking historical data, for example, in Bailey et al. (2017).

The Fellegi and Sunter (1969) (FS) method is an early and widely-used probabilistic linking approach that picks the best match from the set of multiple potential matches. In this method, researchers estimate the probability that a particular characteristic (such as gender or first and last name) agrees in the two files, given that the records should link (match) and given that they should not link (non-match). To estimate match probabilities the FS method relies on the strong, and sometimes untenable, assumption that the agreement status of each characteristic is independent conditional on true match status. Next, the data are used to determine log odds cutoffs above which potential matches are coded as true matches and below which they are treated as non-matches. Candidate pairs that fall between the cutoffs are evaluated manually, a procedure which has been criticized, for example, in Belin and Rubin (1995) because the error properties of manual review are unknown, may be subject to inconsistent standards across reviewers, and may fail to yield a substantial number of unequivocal matches.[2]

ML methods for record linkage constitute a growing alternative to the FS approach.

[2]While manual review of the entire set of blocked records has been adopted in some applications (e.g., Ferrie (1996)), it is prohibitively expensive in many settings and remains subject to the same criticisms as the manual review step of the FS method.

These methods estimate highly flexible non-parametric functions using observed covariates and classify record pairs into matches and non-matches. For example, Cochinwala et al. (2001) and Elfeky et al. (2002) use decision trees for classification while Christen (2008a) and Christen (2008b) rely on support vector machines. ML approaches have been implemented with training data (supervised) and without it (unsupervised), with the former typically yielding more accurate linkage (see, e.g., Christen (2008b)). The key advantage of the ML-based record linkage approach is its high degree of accuracy. These implementations of ML create a deterministic classifier. Hence, like FS, existing ML-based record linkage applications select the best match among a set of candidate matches. That is, conditional on the matching algorithm, matches are treated as deterministic.

The Bayesian approach to record linkage characterizes uncertainty associated with parameters in the linkage process (Fortini et al. (2001) and Larsen (2004)). In this method, researchers specify prior distributions of parameters that govern the mixture of matches and non-matches that generate the comparison vector of agreement status for variables observed in both files. Draws from the posterior predictive distribution of the parameters are then used to produce estimates of pair-specific match probability. One-to-one matching is enforced using the mode of the posterior predictive distribution or by minimizing a loss function. Tancredi and Liseo (2011) refine this procedure by relying on actual observed discrete matching variables rather than a comparison vector of agreement status for those variables. Steorts et al. (2016) provide a method of linking multiple files, each with potentially duplicated records, within the Bayesian framework. Gutman et al. (2013) and Gutman et al. (2014) further develop the Bayesian approach by applying it to situations where variables used in the linkage model are available in both files as well as variables available in only one file. Moreover, they jointly model the linkage step as well as relationships between variables in the linked dataset (the analysis step). Then, by repeatedly sampling from the posterior distribution of the linkage step parameters they generate multiple implicates of linked datasets that are used in the analysis step and combined using the formulas in Rubin (1987). This procedure has the advantage of propagating uncertainty in the linkage step parameters into the analysis step.

Work that is highly germane to the household-employer record linkage problem we consider in this paper began as a part of the Longitudinal Employer Household Dynamics (LEHD) program in two projects that were initiated in the early years of that effort. The first of these projects linked employers to job histories in the 1990-1996 Surveys of Income and Program Participation (SIPP).[3] Abowd and Stinson (2013) evaluates this linkage and

[3]This work also developed improved linkages within the 1990-1993 SIPP job histories, and integrated data from the Census Business Register into the SIPP (Stinson (2003)).

uses it to compare self-reports and administrative reports of earnings. The LEHD program also links establishments in the Quarterly Census of Employment and Wages, called the Employer Characteristics File in LEHD, to individual workers via the state Unemployment Insurance account number, called the SEIN in LEHD. This linkage starts with deterministic methods using the SEIN. When these methods do not find a one-to-one match, a Bayesian posterior predictive distribution is used to generate ten implicates linking establishments to the candidate worker employment history (Abowd et al. (2009)).[4] These ten implicates are used to associate workplace characteristics to each worker history.[5] The ten implicate threads are processed according to the Rubin (1987) combining formulas to produce the Quarterly Workforce Indicators (QWI). McKinney et al. (2017) provides a complete assessment of the total variability in the QWIs due to the MI and other edit procedures.

The methodology we develop relies on the accuracy of the ML approach to record linkage while using MI to characterize uncertainty in the linkage and to propagate that uncertainty into subsequent analyses. To our knowledge, this combination of methods has not been previously employed in record linkage applications.[6] Our ML approach allows to us to leverage a very large number of predictors to estimate match probabilities including both discrete and continuous observed variables from either file as well as agreement status variables constructed using both files. Furthermore, the flexibility inherent in this method accommodates rich complementarity between predictors and allows us to dispense with the assumption that predictor variables are independent conditional on true match status as has been posited in many prior applications. In addition, tuning our prediction models to achieve high out-of-sample accuracy facilitates scalability and precision linkage in a way that is difficult to achieve using Bayesian or FS methods. Finally, unlike prior ML-based record linkage methods that use binary classification to select the single best match, the ML algorithm that we use provides a match probability estimate for each record pair. By repeatedly sampling candidate matches from the estimated match probability distribution when constructing MI linkages, our procedure characterizes uncertainty regarding latent match status.

[4]See Goldstein et al. (2012) for a similar approach applied to medical records.

[5]Other incomplete data in the LEHD infrastructure, such as incomplete data for education, are completed using similar Bayesian methods.

[6]ML methods have been used to improve MI in applications that do not involve record linkage. See, e.g., Reiter (2005) for the creation of partially synthetic public use microdata and Burgette and Reiter (2010) for missing variable imputation.

3.3 Our procedure

In this section, we describe our MI-based record linkage procedure for matching household-level survey data to employer-level administrative data. We rely on three distinct steps to address the needle in the haystack problem associated with record linkage in this setting. We begin by enumerating the set of candidate employers that constitute feasible matches for each survey report about a particular job using a technique known as blocking. Second, we assign match probabilities to each of the feasible candidates by building a training dataset and using it to estimate a flexible ML model that is tuned for out-of-sample accuracy via cross-validation. The model takes a given record pair as an input and returns the estimated match probability. The third step of our procedure uses MI to construct links for each survey report by resampling employer candidates using the estimated match probabilities as weights.

Before delving further into the details of our methodology, we briefly describe the datasets that we use in our application. The household survey that we use is the Health and Retirement Study (HRS) which samples more than 22,000 Americans over the age of 50 every two years. It is a large-scale longitudinal project that studies the labor force participation and health transitions that individuals undergo toward the end of their work lives and in the years that follow. The HRS elicits information about employer identity from respondents to construct measures of pension wealth. These data are obtained at the baseline (i.e., when new respondents are enrolled in the study, generally every six years when a new cohort is added to the study) and in each subsequent wave if the respondent reports having changed jobs. Although the names, addresses, and phone numbers captured in these reports were originally intended to aid the HRS in contacting employers about pension benefits, we use this information for record linkage. We match employer reports in the HRS to the Census Business Register (BR) which is the Census Bureau's list of essentially all employers in the United States. The BR contains information on employer names and address, company affiliation, size, payroll, industry classification and other employer-level characteristics and can be linked to other Census Bureau survey and administrative data on employers. We refer to the dataset created by matching the HRS to the BR and associated Census Bureau employer-level data as the CenHRS.[7]

[7]The original design of the CenHRS was predicated on having Federal employer identification numbers (EINs). EINs would provide tight, but not perfect, linkage to an employee's firm. The reliance on business name and address matching was necessitated by challenges in receiving permissions to use EINs for linkage. The BR includes EINs, and most HRS respondents have given permission to the Social Security Administration (SSA) to provide EINs for their employers to HRS for purposes of enhancing the HRS data infrastructure. We expect that we will evaluate the approach implemented here with a comparison to the matches achieved using EINs when access is obtained. This paper aims to link to both firm and establishment

Notably, employer names and addresses collected by the HRS are not provided to researchers in order to protect respondent confidentiality. Similarly, information about employers stored in the BR is restricted by Federal statute. While we have obtained permission to use personally identifying information to create the CenHRS dataset, the underlying data needed for record linkage will remain restricted. Consequently, imputing matches and then analyzing the matched dataset are necessarily disconnected steps in our application.

Step 1: Blocking

Let jobs in the HRS be indexed by $i = 1, ..., N_{\text{HRS}}$. A job in the HRS is defined as a spell of employment with a unique employer. Let employers in the BR be indexed by $j = 1, ..., N_{\text{BR}}$. If we start with the prior that every record in the BR is a potential match for each job in the HRS, we would need to search over a set of $N_{\text{BR}} \times N_{\text{HRS}}$ pairs. Because this set is of the order $10^6 \times 10^4$, the computational cost of such a prior is prohibitive.

To reduce the dimensionality of the search problem, we establish a blocking strategy. Blocking is a way of grouping together record pairs that share specific characteristics wherein pairs that agree on at least one characteristic are regarded has having a positive probability of being matches, while pairs that fail to agree on any characteristics are deemed as non-matches (see, e.g., Christen (2012)). We block on 3-digit zip code, 10-digit phone number, telephone area code, and city-state. Any BR candidate match that fails to share at least one of the blocking values with an HRS job is assumed to have 0 probability of being a true match. Employing this strategy substantially reduces the number of potential BR candidate matches paired with each HRS job.

Step 2: Supervised machine learning for match probability estimation

Training data

We create a training data set by selecting a sample of $N^T \approx 1000$ unlabeled HRS-BR pairs for human review. We heavily oversample pairs with a high likelihood of being true matches using data and the methodology described in Appendix C.1.

Define $\boldsymbol{x}_i^H$ as a vector of self-reported employer characteristics for HRS respondent i that includes, for example, name, address, phone number, industry, size, etc. Define $\boldsymbol{x}_j^B$ as a vector of characteristics for employer j drawn from administrative data in the BR. Let

(i.e. the specific location at which they work), so even were EINs available, the method developed here would be necessary to link to establishments. We also want to have the capacity to link households that provide employer names and addresses, but may not consent to linkage with SSA administrative data.

$k = 1, \ldots, N^T$ index HRS-BR pairs in the unlabeled sample. These data are examined by reviewers who observe certain pair characteristics, $\boldsymbol{x}_k$, which are a subset of $(\boldsymbol{x}_i^H, \boldsymbol{x}_j^B)$; Table 3.1 lists the elements of $\boldsymbol{x}_k$. Notably, each HRS-BR pair is evaluated by two reviewers. Define $y_{k,r} = 1$ if reviewer r scores pair k as a match and $y_{k,r} = 0$ otherwise. To the extent that they disagree, the two reviewer assessments — i.e. the $y_{k,r}$ — reflect uncertainty about latent match status.[8]

For each HRS-BR pair in the unlabeled sample, reviewers consider employer and establishment match status separately. An employer match means that the employer's identity (e.g., Dunder Mifflin Paper Company) in the HRS corresponds to the employer's identity in the BR. In contrast, an establishment match implies that, in addition to an employer match, the workplace identified by the HRS respondent exactly corresponds to the physical location in the BR (e.g., Dunder Mifflin Paper Company, 1460 Main Street, Scranton, PA). This distinction is important because workplace characteristics can differ substantially even at different locations of a single employer. For example, different establishments of a given employer may experience differential expansion or contraction, produce different types of goods or services, or employ workers of different skill types or ages. Consequently, we construct two different training datasets. In the employer match dataset, $y_{k,r}$ refers to employer match status; in the establishment match dataset, $y_{k,r}$ refers to establishment match status.

Once reviewers complete their assessments, each of the training datasets can be represented by the following matrix:

$$\boldsymbol{T} = \begin{bmatrix} y_{1,1} & \boldsymbol{x}_1 \\ y_{1,2} & \boldsymbol{x}_1 \\ \vdots & \vdots \\ y_{N^T,1} & \boldsymbol{x}_{N^T} \\ y_{N^T,2} & \boldsymbol{x}_{N^T} \end{bmatrix}. \tag{3.3.1}$$

Because there are two reviewer outcomes associated with each HRS-BR pair, $l = 1, \ldots, 2N^T$ indexes the rows of $\boldsymbol{T}$.

Estimating match probabilities

We estimate match probabilities using a set of variables that supplements the information observed by reviewers ($\boldsymbol{x}_l$). The supplemented set of predictors is given by the vector

[8] A total of eight reviewers conducted these reviews inside the Federal Statistical Research Data Center (FSRDC) computing environment.

$\tilde{\boldsymbol{x}}_{l(ij)} = f(\boldsymbol{x}_l, \boldsymbol{x}_i^H, \boldsymbol{x}_j^B)$ where the function $f(\cdot)$ supplements and transforms observed data. Table 3.2 shows the elements of $\tilde{\boldsymbol{x}}_{l(ij)}$. The first two variables are cubic splines of Jaro-Winkler (JW) scores for employer name and establishment address which flexibly capture reviewers' assessments of the similarity in the HRS and BR names and addresses.[9] The next two variables capture the importance of specific employers in the local (i.e. within blocking variable) and national labor market on match probability. These variables account for institutional factors such as specific knowledge about dominant employers that reviewers may have relied upon in ascertaining match status. To flexibly capture all complementarities across name and address similarity and the role of specific employers, we fully interact all four cubic splines together, expanding the set of predictors substantially.

The lower panel of Table 3.2 shows the set of binary predictors. These variables capture agreement between the HRS-BR candidate match on a number of dimensions. Some predictors, such as 10-digit phone agreement can be highly influential in predicting match probability, but it is rare for candidate pairs to share such granular characteristics. On the other hand, sharing SIC industry codes or 4-digit zip codes is more likely but less predictive of a match. The final two variables — employer provision of health insurance and retirement plans — incorporate information obtained purely from HRS respondents. We include these predictors because they are typically associated with large employers and serve as proxies for employer size when such information is missing in the HRS. Altogether there are 1413 variables in $\tilde{\boldsymbol{x}}_{l(ij)}$.

Having defined the set of predictors, we then estimate

$$y_{l(ij)} = \Lambda(\tilde{\boldsymbol{x}}'_{l(ij)}\boldsymbol{\beta} + \epsilon_{l(ij)}), \tag{3.3.2}$$

where $\Lambda(\cdot)$ represents the logistic transformation. Because of the high dimension of $\tilde{\boldsymbol{x}}_{l(ij)}$, we rely on the Elastic Net (EN) shrinkage estimator for model selection and estimation of $\boldsymbol{\beta}$ (Zou and Hastie (2005)). Tuning parameters for the EN estimator are estimated using 10-fold cross validation; see Appendix C.2 for additional details. Plugging $\hat{\boldsymbol{\beta}}$ into Equation (3.3.2) provides an estimate of $P(y_{l(ij)} = 1|\tilde{\boldsymbol{x}}_{l(ij)})$ which is the probability that a human reviewer would regard a given pair as a match.

Notably, because each pair is scored by two reviewers, observationally equivalent cases (i.e. pairs with the same $\tilde{\boldsymbol{x}}_{l(ij)}$) can receive different evaluations about true match status (i.e. different $y_{l(ij)}$). We posit that these differences in reviewer evaluations owe to random factors that are independent of the $\tilde{\boldsymbol{x}}_{l(ij)}$ which constitutes ignorability as in Rubin (1976).

[9] JW scores, which range from 0 to 1, combine edit distance and q-gram-based comparison techniques to measure string similarity.

Formally, our linkage procedure is unbiased if

$$P(y_{l(ij)} = 1|\tilde{\boldsymbol{x}}_{l(ij)}) = P(y_{l(ij)} = 1|\tilde{\boldsymbol{x}}_{l(ij)}, \boldsymbol{z}_{l(ij)}), \tag{3.3.3}$$

where $\boldsymbol{z}_{l(ij)}$ represents a vector of all other observed and unobserved variables that influence match status for a given pair. The high dimension of $\tilde{\boldsymbol{x}}_{l(ij)}$ with non-linearities and rich interactions makes assumption (3.3.3) more tenable and facilitates valid inferences for a richer set of subsequent analyses with the linked data.[10]

Step 3: Multiple-imputation of matches

In the final step of our procedure, we follow MI-based approaches to record linkage which we implement as follows.

1. We estimate model (3.3.2) using the training data and denote the parameter vector by $\hat{\boldsymbol{\beta}}$.

2. We then plug $\hat{\boldsymbol{\beta}}$ into Equation (3.3.2) to obtain the match probability estimate for each HRS-BR pair. Because of the asymmetric nature of the size distribution of firms, each HRS job is mapped to a large number of BR candidates that have a trivially low estimated match probability. Consequently, sampling an implicate with probability proportional to the estimated match probability will often result in a low-quality match being selected — typically a smaller firm or establishment. To overcome mismatch errors of this nature, we impose a data-driven match probability cutoff that further restricts the set of potential candidates surviving the blocking step. The cutoffs are estimated with the training data so as to optimally trade-off the out-of-sample true positive rate (sensitivity) and the out-of-sample false positive rate (1-specificity). We provide additional details about this procedure in Section 3.4.2.1.

3. After eliminating BR candidates with estimated match probability below the cutoffs, we normalize the probabilities to sum to one over the surviving candidates. Then, for each HRS job, we sample $M = 10$ BR candidates with replacement using the normalized match probabilities as weights.

This procedure yields M multiply imputed employer and establishment links for each HRS job thereby constituting M completed datasets. For any statistic generated using imputed data, we can combine estimates obtained from the M completed data sets using the

[10] See, e.g, issues of congeniality in imputation models as noted by Meng (1994), Rubin (1996), and Murray (2018).

formulas in Rubin (1987) to compute the variance owing to sampling noise (within-implicate variability) and the variance due to linkage uncertainty (between-implicate variability).[11] For some scalar parameter of interest θ, let $\hat{\theta}_m$ represent estimates derived from the $m = 1, \ldots, M$ completed data sets. Let $\hat{\sigma_m^2}$ represent the variances associated with each of the M parameter estimates. The multiply imputed estimate of θ is

$$\hat{\theta} = M^{-1} \sum_{m=1}^{M} \hat{\theta}_m \tag{3.3.4}$$

The within-implicate variance is

$$\hat{\sigma_W^2} = M^{-1} \sum_{m=1}^{M} \hat{\sigma_m^2} \tag{3.3.5}$$

The between-implicate variance is

$$\hat{\sigma_B^2} = (M-1)^{-1} \sum_{m=1}^{M} \left(\hat{\theta}_m - \hat{\theta}\right)^2 \tag{3.3.6}$$

The total variance associated with $\hat{\theta}$ is

$$\hat{\sigma^2} = \hat{\sigma_W^2} + (1 + M^{-1})\hat{\sigma_B^2} \tag{3.3.7}$$

3.4 Assessing model fit and linkage accuracy

In this section we implement our record linkage methodology by matching the 2010 wave of the HRS to the BR. We begin by showing partial effects of the EN-based employer and establishment matching models and compare their accuracy with simpler logit models. We then explain how we estimate data-driven cutoffs for implicate selection and examine the extent to which their use improves the quality of the imputations. Finally, we show a variety of statistics to validate the predictive accuracy of the matching models.

[11] In complementary work in a regression context, potential matches can be aggregated using match probability estimates as weights as in Lahiri and Larsen (2005).

3.4.1 ML matching model estimates

3.4.1.1 Partial effects of matching model

Figure 3.1 shows the partial effects of the employer and the establishment matching models. The top row of the figure shows the partial derivative of the employer prediction model for four selected covariates holding all other variables at their means. The bottom row shows analogous effects for the establishment prediction model. For both models, similarity between HRS and BR names and addresses deliver the largest effect on the likelihood that a pair is a true match. This effect only manifests at very high levels of similarity and does so in a highly non-linear fashion. The partial effects in the employer model are uniformly higher than those in the establishment model, reflecting that employer matches are easier to confirm than establishment matches.

The partial effects of the two models underscore the value of employing the EN estimator and relying on cubic splines to model the covariate space. The models we estimate capture sharp inflection points in the curvature of the match likelihood, reflecting non-linearities in reviewer decisions that would be infeasible to replicate using a simpler parametric approach.

3.4.1.2 Predictive accuracy evaluated using cross-validated ROC curves

Although we do not use our models as binary classifiers, we illustrate their predictive performance by showing receiver operating characteristics (ROC) curves in Figure 3.2. For probability thresholds ranging from 0 to 1, the ROC curve plots the true positive rate (sensitivity) on the vertical axis against the false positive rate (1-specificity) on the horizontal axis. A model that was only as good as chance in classifying matches would have an ROC curve that ran along the 45-degree line, while a perfect classifier would have an ROC curve that hugged the left and top edges of the graph. The area under the curve (the c-statistic) would be 0.5 for the good-as-chance classifier, while it is 1.0 for a perfect classifier. As such, the c-statistic captures the predictive ability of the model in a single number which aids in evaluating in the quality of the model relative to the two extremes of 0.5 and 1.0.

The left panel of Figure 3.2 compares employer match prediction performance using ROC metrics. The blue curve is based on the EN estimator while the red curve is based on a traditional logistic regression model estimated using JW scores of name and address. Applying tuning parameters from the minimum deviance EN model, we estimate model coefficients using 9/10ths of the training data and compute sensitivity and specificity estimates by projecting the model on the remaining 1/10th of the sample.[12] Iterating through each hold-out tenth yields the out-of-sample ROC estimate. The c-statistics

[12]Deviance is defined as -2 times the maximized log likelihood. Lower deviance implies better model fit.

from the two models are 0.98 and 0.94 respectively. The right panel shows the same contrast for the establishment prediction model; the c-statistics for these two models are 0.94 and 0.88 respectively. Employer matches are easier to ascertain than establishment matches. Employer matches depend mainly on address while establishment matches depends importantly on address in addition to name, so establishments are unconditionally less likely to be found relative to employers.

Table 3.3 compares the relative precision of each model at pre-selected sensitivity levels of 0.85, 0.90, 0.95 and 0.99. For the employer match model, EN attains false positive rates (1-specificity) which are 2-to-5 times lower than the corresponding values attained by traditional logistic regression. At the same sensitivity levels for the establishment match model, EN attains false positive rates which are 1.5-to-3 times lower than the corresponding values attained by traditional logistic regression. Taken together, the results shown in Figure 3.2 and Table 3.3 indicate that the EN prediction models deliver very high predictive ability out-of-sample, outperforming the simpler logit models in both cases.

3.4.2 Implementing probabilistic matches

3.4.2.1 Cutoff estimation

Although the blocking step dramatically shrinks the number of potential BR candidates, the number of candidate matches is still very large for many HRS jobs. This empirical regularity is a consequence of seeking to match household-level survey data to an establishment-level frame. Most individuals are employed at a relatively small number of large employers whereas the vast majority of employers are, in fact, very small. As a consequence, the set of potential candidates for each HRS job is populated with many small employers each of which receive a trivial match probability, a problem that is made more acute when EINs cannot be employed for blocking. To illustrate the impact of block size on match uncertainty, consider an example where an HRS job is paired with 1000 BR candidate matches with one large employer candidate being the correct match and 999 small employer candidates being non-matches. Suppose the large employer candidate obtains a match probability of 0.5 while the remaining 999 small employer candidates receive match probability $= \frac{0.5}{999}$. In this instance, random small employers will populate half of the implicates even though they are two orders of magnitude less likely to be the right match relative to the large employer. To address this concern, we propose a data-driven procedure that mitigates the impact of block-size induced noise in the linkage process.

The details of the procedure are as follows. First, we estimate out-of-sample ROC curves using the Bayesian bootstrap weighted training data. We then pick the probability threshold

p^* that minimizes

$$D(p) = \left((1 - \text{sensitivity})^2 + (1 - \text{specificity})^2\right)^{1/2}. \tag{3.4.1}$$

The cutoff probability p^* minimizes the distance between the ROC curve and the upper left corner of the graph. Put differently, p^* is the feasible cutoff closest to the infeasible point where the sensitivity and specificity are both 1 (see, e.g., Coffin and Sukhatme (1997) and Youden (1950)).[13] We estimate these cutoffs separately for each quartile of the block size distribution in the training data. Table 3.5 shows the average number of pairs in each quartile of the block-size distribution along with estimates of the associated cutoff probabilities.

Having obtained block-size dependent cutoffs, we leave as unmatched any candidate pairs whose estimated match probability is below the cutoff. Any matching procedure, of course, should admit the possibility that there is no reasonable match. Our procedure handles this possibility systematically based on the estimated matching model, its uncertainty, and a well-specified objective function. Hence, determining that a case is a non-match is entirely integrated into the procedure. It does not rely on ancillary determination, for example, that a case is treated as a non-match if the match probability is below a externally specified threshold.

3.4.2.2 Selecting the implicates

We now have the ingredients to implement the match. The previous subsections describe estimating the matching model which assigns a probability to each record pair, and finding the optimal cutoffs where probabilities below the cutoff are set to zero when selecting the implicates. We re-normalize the probability of pairs above the cutoff to sum to one over the set of surviving candidates and select implicates from this set using the normalized probabilities as weights. We implement this procedure on the 2010 wave of the HRS, the remainder of the paper uses this dataset to evaluate the properties of the match procedure and to do an application.

[13] This loss function gives equal weight to sensitivity and specificity and, of course, one could consider other weights. Given that the estimated ROC curves for our matching models are symmetric and hug the left and top edges of the graph, varying the weights would not change the cutoffs substantially.

3.4.3 Evaluating the linkage

3.4.3.1 Match uncertainty reduction using cutoffs

Table 3.6 illustrates the degree of concentration for MI linked data.[14] Concentration among the implicates is defined as the proportion of unique matches among the 10 multiply imputed matches for each HRS job. The left panel shows employer-match concentration whereas the right panel shows establishment-match concentration; for both models, we show concentration rates without cutoffs and with cutoffs. The first row of the table shows the fraction of HRS jobs for which a single BR record populated all 10 implicates, which is the maximum level of concentration. Subsequent rows show the share of implicates associated with successively higher numbers of unique BR entities. Cases with 5 or more unique matches are binned together.

The table shows that the imposition of minimum match probability cutoffs increases the concentration of BR entities across the implicates; that is, less disparate BR entities are realized as potential links after imposing the cutoff. For both models, the fraction of HRS jobs mapped to a single BR candidate rises by a factor of approximately 50 after cutoffs are used to exclude low quality matches. Concentration increases in the range of 10-to-30 fold are seen for jobs mapped to two or three unique BR entities. Because employer matches are easier to ascertain, the level of concentration is higher overall as compared with establishment matches.

3.4.3.2 Internal validation via employer identity concordance across models

Our models predict employer and establishment match status independently. As an internal check on prediction accuracy, we use MI linked data to compare the concordance between employer identity from the employer match model with employer identity from the establishment match model.[15] A high degree of overlap in this dimension indicates that both models select the same set of employers. Greater agreement between the two models is therefore not only an internal consistency check, but also a measure of match certainty.

Table 3.7 shows the average fraction (out of 10) of employer identities that are common to both the employer and establishment predictions for each HRS job. These rates are reported for each quartile of the block-size distribution, and for the sample as a whole. Prior

[14] For 33 percent of HRS jobs, exactly 0 BR candidates survive the employer match threshold. For the establishment match model, 8 percent of HRS records have 0 BR candidates. These HRS jobs represent cases where there is not enough information to produce a plausible match candidate from within the blocked set. Nevertheless, to the extent that exclusion from the imposition of cutoffs is non-random, it can generate selection bias. We investigate this issue in Appendix C.3.

[15] Employer identity is ascertained on the basis of Census firm identifiers. All establishments associated with a particular employer have the same Census firm identifier.

to imposing probability cutoffs, smaller block sizes yield greater concordance between the models: the average rate in the bottom quartile is 14 percent whereas the average rate in the top quartile is almost halved to 6.6 percent. This decline occurs because uncertainty grows in the number of potential matches. Once cutoffs are imposed, these concordance rates increase by 4-to-8 fold. Furthermore, the monotonic decline in match uncertainty vanishes in the cutoff based sample. Taken together with the concentration improvements highlighted earlier, these statistics show that the application of a simple filtering technique can dramatically reduce between-implicate variability among multiply imputed matches.

The less than perfect concordance reflects intrinsic uncertainty in record linkage. Researchers might not be happy with this uncertainty, but making it explicit is clearly superior to choosing a deterministic procedure and proceeding as if it were exact.

3.4.3.3 Out-of-sample validation using pension plan information

We exploit the availability of pension plan information for respondents from the 2010 wave of the HRS as a secondary check of the fit of our employer prediction model that is entirely independent of the training data set. For respondents in the 2010 wave, the HRS pension project collected employer pension plan information from Internal Revenue Service Form 5500 (F5500) filings which include a Federal Employer Identification Numbers (EINs). Using our sample of MI linked data, we match each respondents F5500 EIN to the BR to determine the employer who sponsored their pension plan.[16] Then, for each HRS job, we ascertain whether the employer identified by the HRS pension project appears among the set of 10 implicates. Table 3.4 shows that the average of this concordance rate over approximately 1900 jobs with linked F5500 EINs is 0.42. When we impose data-driven cutoffs to filter away low quality matches, the concordance rate rises to 0.63.Notably, this validation method understates the accuracy of the matching model because, although many pension plan EINs represent the respondent's current employer, others could represent EINs for union-sponsored pensions that do not have corresponding entries in the BR.

3.5 Comparing self-reported and imputed employer characteristics

In Table 3.8 we show moments of the employer and establishment size distribution. We consider four different methods of measuring employer and establishment size which we

[16]Employer identity is ascertained on the basis of a variable known as the Census firm identifier. All establishments associated with a particular employer have the same Census firm identifier even if they have different EINs.

illustrate by reporting averages within selected percentiles of employer and establishment size distribution. The top panel of the table shows employer size statistics while the bottom panel shows establishment size statistics. The percentiles of the size distribution are re-computed separately for each row. The table hence illustrates what a researcher would infer about the size characteristics of HRS respondents using either the HRS self-report or the population of linked employers or establishments.

For each panel, the first row is the HRS self-reports and the next three rows use different imputation methods for linkage to the BR:

1. SI-highest selects the implicate with the highest predicted match probability; this procedure is similar to FS methods where subjective judgment is used to reconcile the presence of multiple potential matches. The benefit of SI-highest over purely subjective match selection is that match probability estimates provide a well-defined metric to select between alternative candidates.

2. MI draws 10 implicates with replacement using normalized match probability estimates as weights.

3. MI-optimized first imposes data-driven cutoff probabilities on the estimated match probability distribution, eliminates cases below the cutoff, and then re-samples implicates with replacement using the normalized probabilities as weights.

For the MI-based statistics we show, in addition to means and standard errors, the fraction of variance that owes to between-implicate uncertainty (the missingness ratio). The missingness ratio captures variability that is relevant for inference that is ignored using SI-based methods. The standard errors and missingness ratio for MI-based estimates are computed using Equations (3.3.5)-(3.3.7).

Comparing moments of the employer and establishment statistics using each of these procedures facilitates comparison with other linkage applications and highlights the value of our preferred method. Employer and establishment size as reported by HRS respondents is consistently larger than MI imputation from the BR.[17] In contrast, the SI-highest and MI-optimized based estimates of employer and establishment size are larger than MI estimates. Reflecting the approximately log-normal distribution of firm size, most matching blocks contain many small firms. In contrast, the dominant fraction of workers are employed

[17]Employer size in the HRS is first elicited as a continuous variable. If respondents do not report a number, they are given the option of reporting one of six bins: [1,4], [5,14], [15,24], [25,99], [100,499], and 500+. In Table 3.8, we convert binned reports of employer and establishment size to continuous values by using the midpoint of the interval. For respondents who report "500+", we impute a continuous value by randomly drawing an employer size from the set of continuous valued reports that are above 500.

at large firms. This dichotomy has the potential to generate non-trivial bias especially because our task is to match household survey data to an establishment level frame.The MI procedure over-represents small firms because it draws implicates from a set of candidates where the count of small firms far exceeds large firms. Because self-reports of establishment size are likely to be more reliable than employer size (individuals know how many workers are at their workplace more readily than how many workers a firm employs across workplaces), we use the lower panel of the table to establish the bias reduction gains of SI-highest and MI-optimized imputation strategies. Across much of the establishment size distribution, these two measures more accurately correspond to self-reports while conventional MI is consistently downward biased. Improvement in imputation accuracy obtains because SI-highest and MI-optimized select larger employers with greater probability and, therefore, produce estimates that are closer to HRS self-reports.[18]

While both SI-highest and MI-optimized mitigate bias, only MI-optimized incorporates linkage uncertainty into the standard error estimate. For the middle tenth of the employer and establishment size distribution, about 3.5 percent of the variance in average size owes to linkage uncertainty respectively. This variability is ignored in the SI-highest procedure thereby downward biasing the associated standard errors. Secondary to the improvement of MI-optimized over SI-highest, we see that the cutoffs based procedure generally reduces between variability relative to conventional MI, reinforcing earlier measures of concentration and concordance among the 10 implicates.

3.6 Application: The wage-firm size gradient

Using both household and firm level survey data as well as administrative employer-employee linked data, a number of studies have established that larger employers pay observationally equivalent workers higher wages (see, e.g., Brown and Medoff (1989), Oi and Idson (1999), and Bloom et al. (2018)). In this section we discuss an application of CenHRS by re-examining the relationship between wages and employer size. In particular, our approach reveals how systematic biases generated by measurement error and nonresponse would remain hidden without establishing linkages to administrative data.

[18]As we show in the next section, HRS self-reports of employer and establishment size are downward biased due to nonclassical measurement error and nonresponse bias.

3.6.1 Wage-firm size gradient in household-survey data

Consider the following statistical model for the relationship between firm size and worker wages in the cross section

$$w_{ij} = \gamma_0 + \gamma_1 s^*_{ij} + v_{ij}, \tag{3.6.1}$$

where w_{ij} is the log hourly wage of worker i employed at firm j, s^*_{ij} is an error-free measure of the log of worker $i's$ employer's size, and v_{ij} is an error term that captures other factors influencing worker wages. The HRS provides household survey-based measures of hourly wages, w_{ij}, as well as potentially error-ridden measures of employer size, s_{ij} , where

$$s_{ij} = s^*_{ij} + u_{ij}. \tag{3.6.2}$$

Under the classical measurement error model, discrepancies in survey reports are not systematically related with the underlying variable of interest implying that $Cov(u_{ij,}v_{ij}) = 0$. Given this framework, it is well known that the presence of added noise in the explanatory variable attenuates the ordinary least squares (OLS) estimate of γ_1. Alternatively, if discrepancies in survey reports are systematically related to the underlying variable of interest — i.e. if the measurement error is non-classical — then the OLS estimate of γ_1 may be either amplified or attenuated depending on the sign of $Cov(u_{ij,}v_{ij})$ and its magnitude relative to $V(s_{ij})$. In the following subsections, we use our MI-optimized measures of employer size to determine whether the measurement error is classical or non-classical and assess its effect on the wage-size gradient.

3.6.2 Using MI-optimized variables to assess the nature of measurement error

Define $\hat{s}^{*(m)}_{ij} := g(\tilde{\boldsymbol{x}}_{ij}; \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{p}}^*)$ as the m-th implicate of log firm size obtained using the MI-optimized procedure. Recall that $\tilde{\boldsymbol{x}}_{ij}$ is the high-dimensional vector of match predictors, $\hat{\boldsymbol{\beta}}$ is the parameter vector that indexes the matching model used for imputation, and $\hat{\boldsymbol{p}}^*$ is a vector of block-size specific match probability cutoffs. We can write the true value of log firm size under our imputation procedure as

$$s^*_{ij} = \hat{s}^{*(m)}_{ij} + \eta_{ij}. \tag{3.6.3}$$

Ignorability as posited in Equation (3.3.3) implies the following moment conditions:

$$Cov(\hat{s}_{ij}^{*(m)}, \eta_{ij}) = 0 \tag{3.6.4}$$

$$Cov(\hat{s}_{ij}^{*(m)}, v_{ij}) = 0 \tag{3.6.5}$$

i.e., the imputed variable is uncorrelated with imputation error, η_{ij}, as well as the error in the regression model, v_{ij}.

The probability limit of $\hat{\gamma}_1$ estimated using the MI-optimized measures of firm size by OLS is

$$\begin{aligned}\hat{\gamma}_1 \underset{p}{\rightarrow} & \frac{Cov(\hat{s}_{ij}^{*(m)}, \gamma_0 + \gamma_1 s_{ij}^{*} + v_{ij})}{V(\hat{s}_{ij}^{*(m)})} \\ &= \gamma_1 \frac{Cov(\hat{s}_{ij}^{*(m)}, s_{ij}^{*})}{V(\hat{s}_{ij}^{*(m)})},\end{aligned} \tag{3.6.6}$$

where the second expression follows from Equation (3.6.5). Finally, from Equations (3.6.3) and (3.6.4) it follows that $Cov(\hat{s}_{ij}^{*(m)}, s_{ij}^{*}) = V(\hat{s}_{ij}^{*(m)})$ which implies that the estimate of $\hat{\gamma}_1$ based on MI-optimized variables is consistent.

Having established the consistency of OLS estimates of γ_1 based on MI-optimized measures, we turn next to assess the nature of measurement error in HRS reports by comparing the wage-size gradient obtained using survey self-reports of workplace size along side MI-optimized measures of workplace size.

3.6.3 HRS reporting error about workplace size is correlated with wages

We begin by showing non-parametric evidence of the positive wage-size gradient in our sample. The left panel of Figure 3.3 shows average log hourly wages for each decile of the log employer-size distribution, while the right panel shows average log hourly wages for each decile of the log establishment-size distribution. In both panels, the gradient based on self-reported size is steeper than the gradient based on MI-optimized size. If employer and establishment size were subject to classical measurement error — as is often the case in self-reports of earnings — one would expect the survey based gradient to be attenuated relative to the administrative data based gradient. However, the converse is true.

Figure 3.4 explains amplification bias by showing how measurement error and selective nonresponse in HRS are systematically correlated with wages. The top left panel shows average log employer size for each decile of the log wage distribution and illustrates a stark pattern: workers in lower deciles of the wage distribution underreport the size of their employer. This error diminishes as wages increase but does not vanish even at the top of the wage distribution. As such, self-reporting error about employer size is positively correlated with wages. Thus, $Cov(u_{ij,}v_{ij}) > 0$, and we observe amplification bias in the survey-based wage-size gradient. The lower left panel of Figure 3.4 shows the same qualitative relationship between self-reported error in establishment size and wages. Relative to employer-size discrepancies, the magnitude of these errors are substantially smaller and the confidence interval estimates for the two curves overlap across the entire wage distribution. Biases such as these could occur if low-wage workers are less informed about the employment structure of a firm than are say, high-wage, managerial workers who have more institutional knowledge of the firm's operations. They could also emerge if low-wage workers at multi-establishment firms tend to report establishment size as a proxy for employer size more frequently than do high-wage workers.

To illustrate the role of nonresponse bias, the two right-hand panels of Figure 3.4 show similar contrasts but restrict the sample of administrative data imputations to coincide with the sample where HRS respondents provide self-reports. This restriction eliminates nonresponse bias as a reason for the difference between self-reports and administrative data imputation by focusing purely on reporting error. The similarity between the plots in the left half of the figure and those in the right half indicate that reporting error is, in fact, the main driver of amplification bias.

To examine these differences more carefully, Table 3.9 shows measurement error and nonresponse bias within each decile of the wage distribution.[19] Measurement error is consistently negative and declining in wages for employer and establishment size, whereas nonresponse errors are typically positive and are largest in the 3rd, 4th, and 5th deciles of the log wage distribution. With a few exceptions in the tails of the wage distribution, these data reveal that non-responders in the HRS predominantly work at larger firms and establishments than do responders. Furthermore, because selective nonresponse is concentrated in lower deciles of the log wage distribution, relying purely on self-reports would make it appear that lower-wage workers are employed at smaller firms and establishments than is actually the case. This bias further amplifies the survey-based wage-size gradient.

The patterns discussed here provide new evidence on how survey responses about employer characteristics are selectively misreported or not reported at all. With linkages to administrative information on employers in CenHRS, we are able to characterize measurement and nonresponse errors that are unobservable in other household survey datasets.

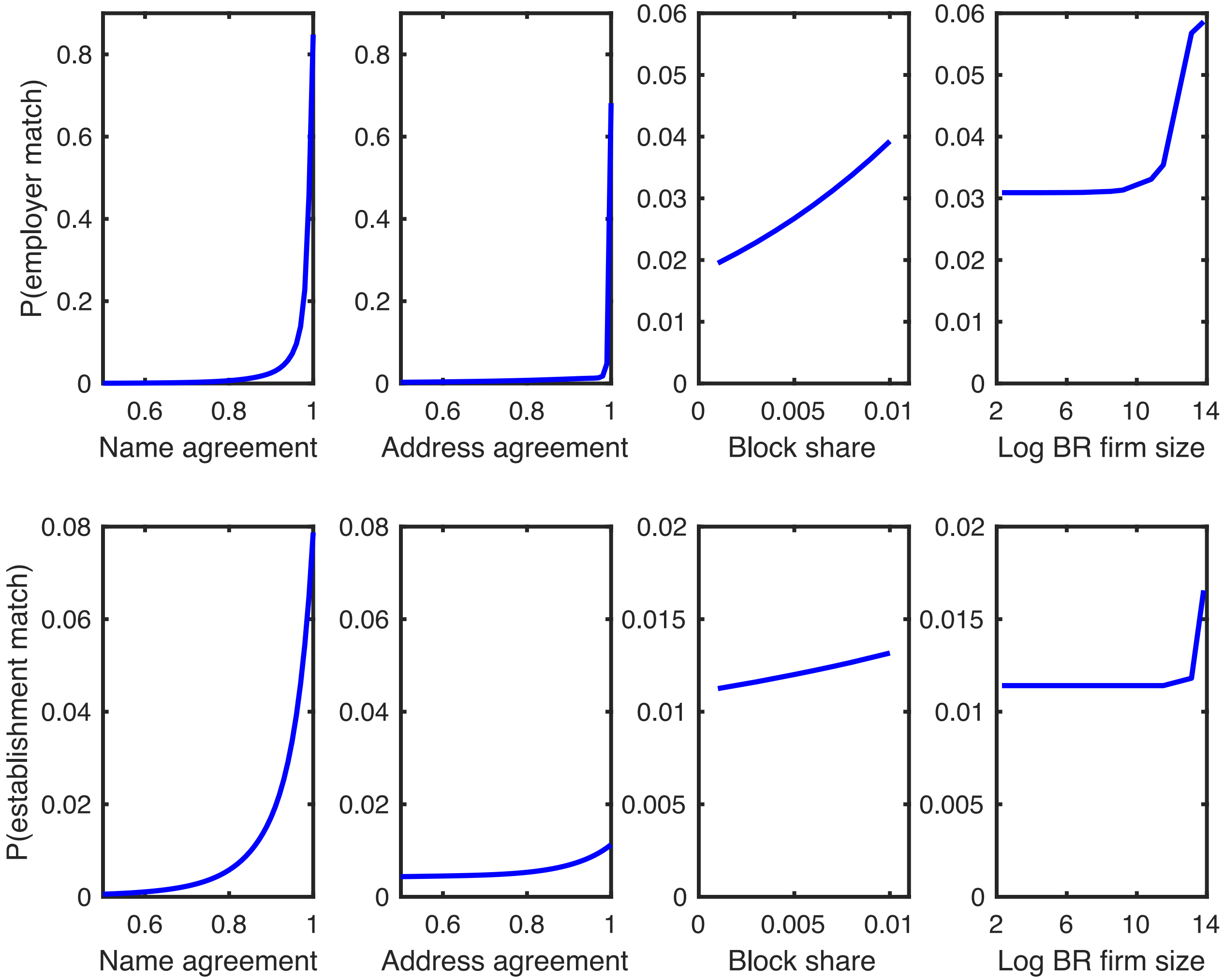
[19]Appendix C.4 formalizes how measurement error and nonresponse bias terms are computed using moments from HRS self-reports and MI-optimized variables.

3.7 Conclusion

This paper describes the construction of a new dataset, CenHRS, that is obtained by linking a household-level survey to an establishment-level frame in the absence of unique identifiers. The between-frame linkage task that we undertake is complicated by asymmetries in the distribution of employment across firms that makes matching inherently noisy. To address these issues, we resort to probabilistic linkage and utilize a supervised machine learning model to estimate the probability that specific employers and establishments in the BR are matches for individuals in the HRS. Our prediction model relies on a rich set of covariates and a high degree of flexibility to replicate important non-linearities inherent in the training data. Using probabilities estimated from the model, we employ MI to characterize uncertainty in the linkage. To further refine the posterior distribution of candidate matches we estimate probability cutoffs that provide the best sensitivity and specificity combination out-of-sample. Eliminating candidate matches that fail to meet these cutoffs dramatically reduces between-implicate variability while also reducing biases inherent in the between-frame linkage that we construct. We use these newly linked data to provide new evidence that reporting errors as well as selective nonresponse to survey questions on employer characteristics vary systematically with wages.

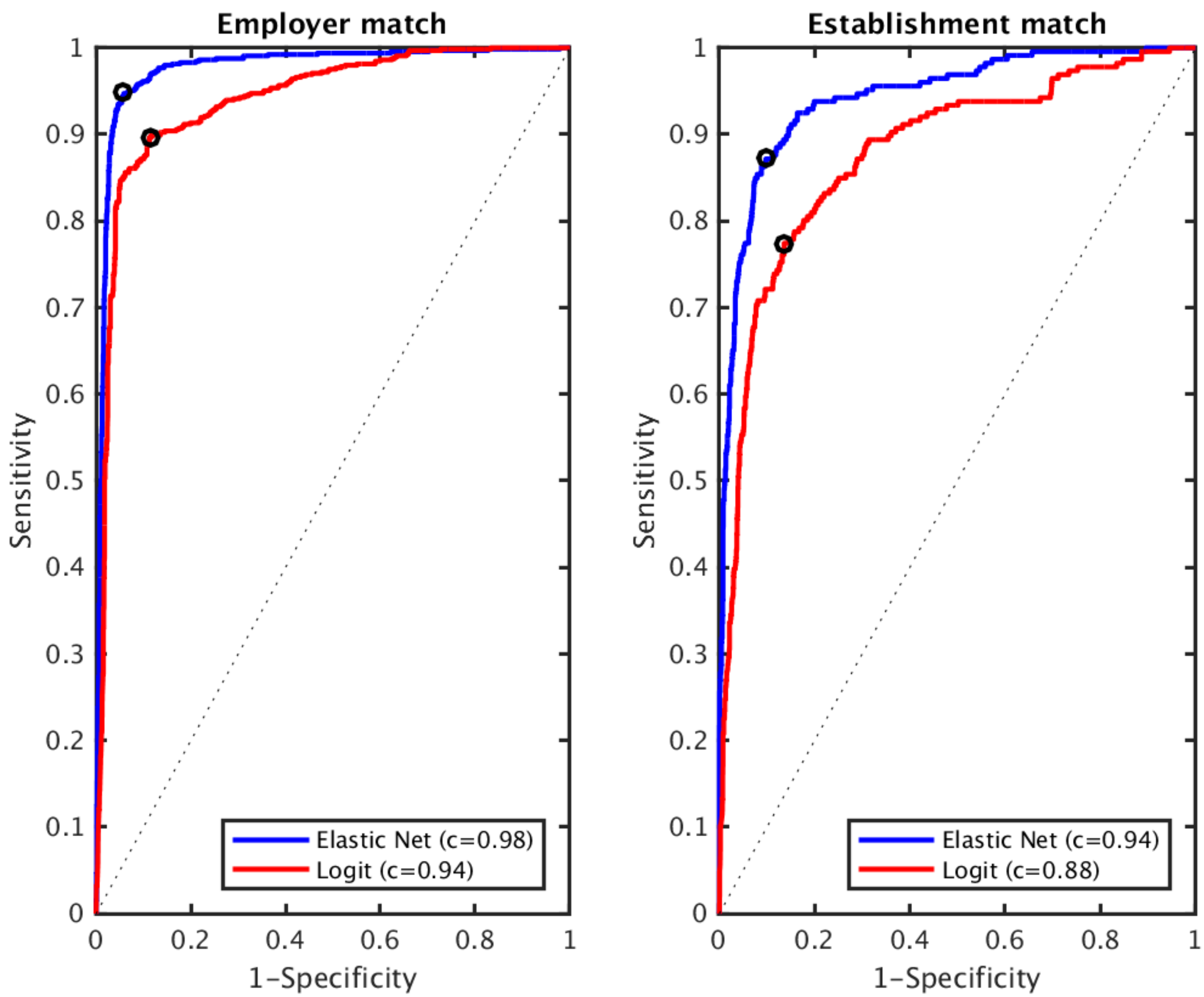
Beyond issues related to record linkage, CenHRS opens new avenues for research by extending pre-existing measures of activities, experiences, and outcomes for individuals from their family and home context to the work context. These new measures will provide data necessary for a more comprehensive understanding of the determinants of health and well-being over the lifespan. To validate and extend the linkages that we have developed in this paper we will exploit the availability of EINs in subsequent efforts, substantially improving the quality of these data for future research.

Figure 3.1: Partial effects of the matching models



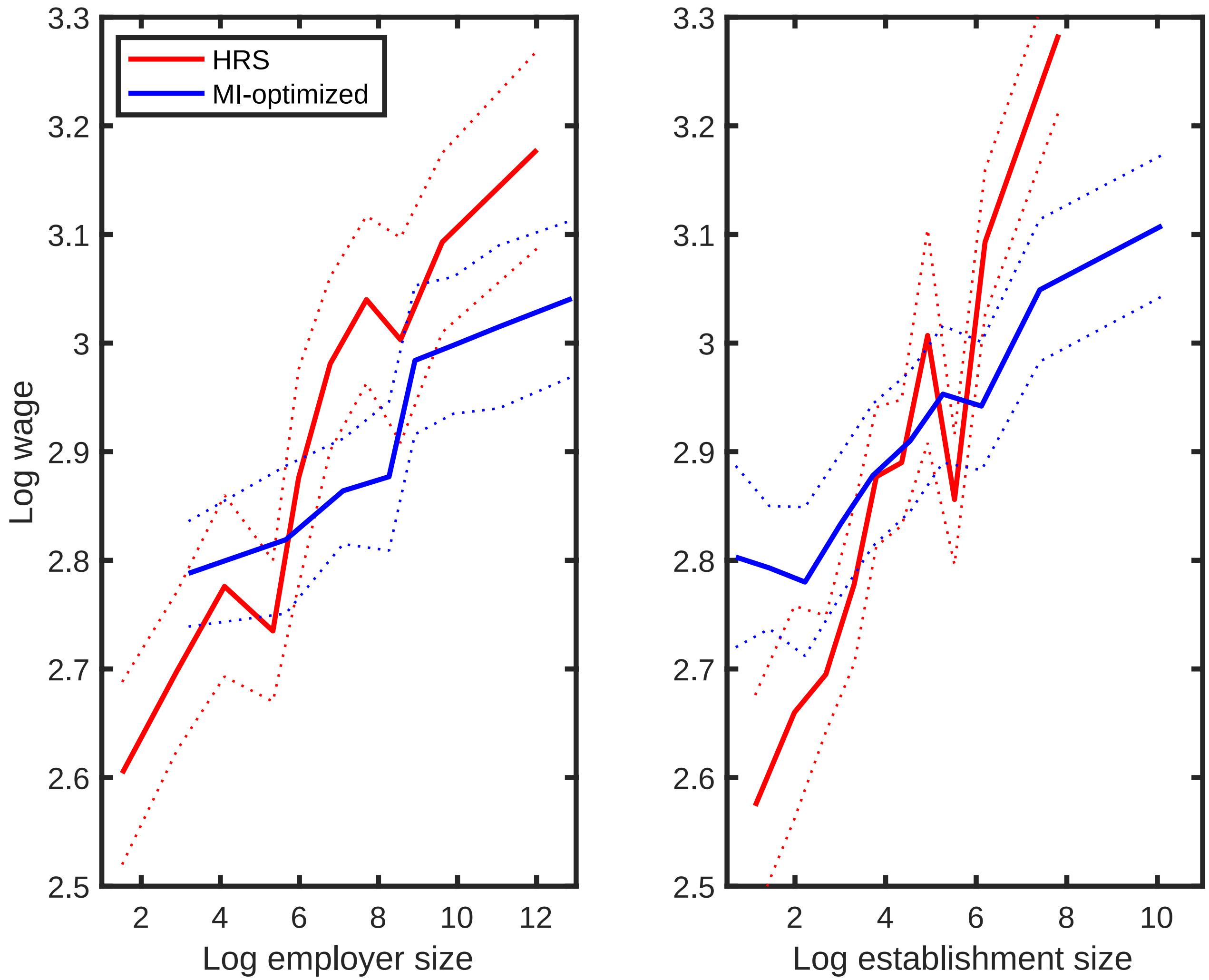
Notes: Each graph shows the partial derivative of the matching function for a given predictor holding all other predictors at their mean. Name and address agreement are based on Jaro-Winkler scores for similarity between HRS and BR names and addresses. These variables range from 0 to 1. Block share is the fraction of employment within the block (i.e. 3-digit zip, 10-digit phone number, telephone area code, or city-state that are common between a given HRS-BR pair) accounted for by a given establishment in the BR.

Figure 3.2: ROC curves of the matching models



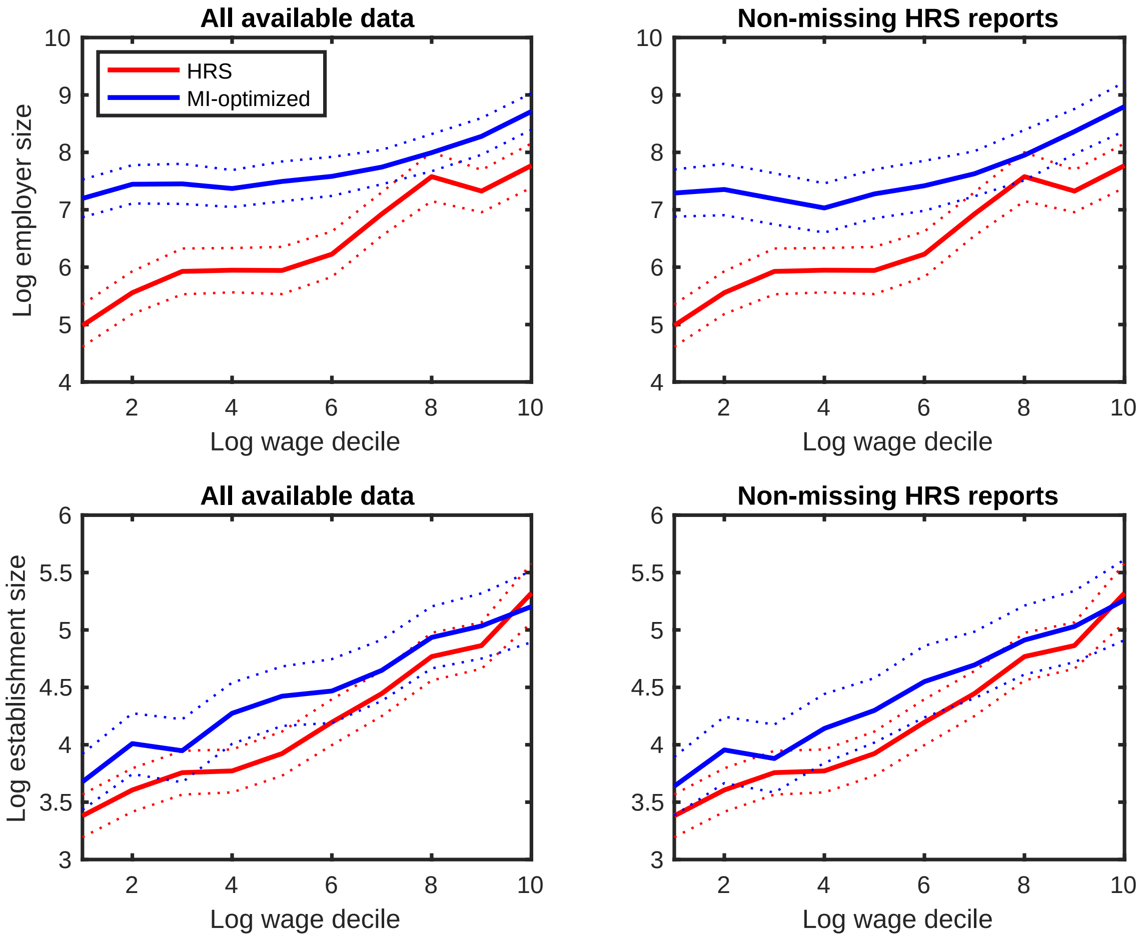
Notes: ROC estimates are computed using 10-fold cross-validation. Hollow black circles show the model's optimal cutoff probability which minimizes the distance to the top left corner of the graph (i.e. yield maximum sensitivity and specificity).

Figure 3.3: HRS versus CenHRS wage-size gradient



Notes: Dotted lines show 95 percent confidence intervals. Some HRS respondents report hourly wages directly. Others report compensation at daily, weekly, monthly, or annual levels. When compensation is reported at a different level than hourly, we convert it using the respondent's report of how many hours per week and weeks per year worked. Data in the 1-3 deciles and the 4-5 deciles of the BR imputed employer size distribution are binned together to prevent disclosure of information in small cells.

Figure 3.4: Measurement error and nonresponse bias across the wage distribution



Notes: Dotted lines show 95 percent confidence intervals. For multiply imputed statistics from the BR, the confidence interval accounts for within- and between-implicate variability. Some HRS respondents report hourly wages directly. Others report compensation at daily, weekly, monthly, or annual levels. When compensation is reported at a different level than hourly, we convert it using the respondent's report of how many hours per week and weeks per year worked. In the left-hand column, the MI-optimized plot uses all available data, regardless of whether respondents provide a report of their employer's or establishment's size. In the right-hand column, the MI-optimized plot is restricted to the sample of respondents that report employer and establishment size.

Table 3.1: Reviewer's information set

Category	Review variables (HRS and BR)
1	Employer name, establishment address, and phone number
2	Employer single or multi-unit status
3	Employer and establishment size
4	Employer industry code and code description

Table 3.2: Key predictors in the matching models

Predictor	Description
Cubic spline JW score name	Similarity in HRS and BR name
Cubic spline JW score address	Similarity in HRS and BR address
Cubic spline block share	Importance of establishment in local (within blocking variable) labor market
Cubic spline employer size (BR)	Importance of employer in national labor market (employer size from BR)
Full interaction of cubic splines	All complementarities between continuous variables
Employer size agreement	1 if employer size in HRS and BR agree, 0 if missing or disagree
Establishment size agreement	1 if establishment size in HRS and BR agree, 0 if missing or disagree
Multi-unit status in BR	1 if employer is multi-unit, 0 if single-unit
7-digit phone number agreement	1 if 7-digit phone number in HRS and BR agree, 0 if missing or disagree
10-digit phone number agreement	1 if 10-digit phone number in HRS and BR agree, 0 if missing or disagree
3-digit zip code agreement	1 if 3-digit zip code in HRS and BR agree, 0 if missing or disagree
4-digit zip code agreement	1 if 4-digit zip code in HRS and BR agree, 0 if missing or disagree
5-digit zip code agreement	1 if 5-digit zip code in HRS and BR agree, 0 if missing or disagree
SIC industry code agreement	1 if SIC code in HRS and BR agree, 0 if missing or disagree
Employer provides health insurance	1 if HRS respondent indicates employer provides health insurance, 0 if missing or no provision
Employer provides retirement plan	1 if HRS respondent indicates employer provides retirement plan, 0 if missing or no provision

Notes: The cubic splines for JW scores for name and address have 10 cut points each. The cubic splines for block share and employer size have 3 cut points each. There are a total of 1413 predictors prior to model selection.

Table 3.3: Prediction accuracy: Elastic Net versus Logistic regression

	1-Specificity			
	Employer match		Establishment match	
Sensitivity	Elastic Net	Logit	Elastic Net	Logit
0.85	0.026	0.057	0.080	0.266
0.90	0.034	0.136	0.145	0.361
0.95	0.068	0.354	0.309	0.697
0.99	0.311	0.634	0.605	0.887

Notes: Sensitivity and specificity estimates are computed using 10-fold cross-validation.

Table 3.4: Validation with pension plan information

	No cutoffs	Cutoffs
Employer ID agreement	0.421	0.625
N	1900	1250

Notes: Pension plan EINs are obtained through an independent linkage exercise where the HRS used employer names to search for IRS Form 5500 pension filings.

Table 3.5: Receiver Operating Characteristics curve-based cutoff estimates

Quartiles of block size	Avg. block size	Cutoffs	
		Employer match	Establishment match
1	13620	0.154	0.066
2	26390	0.343	0.154
3	44340	0.600	0.171
4	99270	0.534	0.123
Full sample	46220	0.236	0.157

Notes: ROC estimates are computed using 10-fold cross-validation. Cutoff estimates provide probability thresholds which minimizes the distance to the top left corner of the ROC graph (i.e. yield maximum sensitivity and specificity). The training data set has $\approx$ 2000 observations.

Table 3.6: Concentration of multiple implicates

	Employer match		Establishment match	
Unique matches	No cutoffs	Cutoffs	No cutoffs	Cutoffs
1	0.009	0.478	0.001	0.057
2	0.015	0.283	0.002	0.058
3	0.020	0.158	0.002	0.051
4	0.031	0.055	0.002	0.060
5-10	0.926	0.027	0.993	0.774
N	5700	3700	5700	5200

Notes: This table is based on the set of working HRS respondents in the 2010 wave. Totals may not sum to 1 because each cell is independently rounded. HRS jobs with 5 or more matches are binned together to prevent disclosure of information in small cells.

Table 3.7: Concordance between employer and establishment models

Quartiles of block size	No cutoffs	Cutoffs
1	0.140	0.547
2	0.103	0.694
3	0.066	0.718
4	0.066	0.522
Full sample	0.110	0.586

Notes: This table is based on the set of working HRS respondents in the 2010 wave. Block size is defined as the number of candidate BR pairs within a block defined on an HRS job (i.e. 3-digit zip, 10-digit phone number, telephone area code, or city-state).

Table 3.8: Employer and establishment size statistics

	A: Employer size percentile				
Source	[0,25)	[25,45)	[45,55)	[55,75)	[75,100]
HRS	17.6	177.8	710.5	2945	164800
	(.51)	(3.56)	(10.92)	(50.89)	(13040)
SI-highest	2.4	124.6	917.1	6668	190100
	(.08)	(3.47)	(15.63)	(129.10)	(8270)
MI	0.8	7.1	61.4	3537	126700
	(.02)	(.13)	(1.79)	(107.80)	(5643)
	[0.001]	[0.007]	[0.029]	[0.008]	[0.001]
MI-optimized	63.4	804.6	2673	8474	247300
	(2.27)	(15.97)	(31.74)	(156.20)	(12210)
	[0.002]	[0.008]	[0.033]	[0.003]	[0.000]
	B: Establishment size percentile				
Source	[0,25)	[25,45)	[45,55)	[55,75)	[75,100]
HRS	5.6	25.4	54.5	127.3	1848
	(.11)	(.29)	(.32)	(1.70)	(229)
SI-highest	0.8	5.0	13.5	58.5	14910
	(.02)	(.06)	(.14)	(1.02)	(2613)
MI	0.5	2.4	4.1	8.6	2158
	(.01)	(.02)	(.02)	(.09)	(983)
	[0.001]	[0.067]	[0.472]	[0.023]	[0.001]
MI-optimized	1.0	7.8	26.45	120.4	42020
	(.02)	(.12)	(.29)	(1.97)	(4880)
	[0.022]	[0.030]	[0.034]	[0.008]	[0.001]

Notes: This table is based on the set of working HRS respondents in the 2010 wave. SI-highest selects the BR candidate associated with the highest predicted match probability. MI selects 10 BR implicates for each HRS respondent using estimated match probabilities as weights. MI-optimized imposes data-driven cutoff probabilities, eliminates BR candidates below the cutoffs, and then selects 10 BR implicates using estimated match probabilities as weights. Standard errors are shown in parentheses. MI standard errors incorporate within and between variablility. Missingness ratios (ratio of between-implicate variance to total variance) are shown in square brackets.

Table 3.9: Measurement error and nonresponse bias across the wage distribution

	Log employer size		Log establishment size	
Log wage decile	Measurement error	Nonresponse bias	Measurement error	Nonresponse bias
1	-2.308	-0.094	-0.258	0.038
2	-1.798	0.091	-0.349	0.055
3	-1.264	0.261	-0.123	0.068
4	-1.084	0.339	-0.370	0.132
5	-1.334	0.217	-0.376	0.124
6	-1.192	0.164	-0.354	-0.082
7	-0.701	0.116	-0.249	-0.047
8	-0.374	0.044	-0.144	0.023
9	-1.038	-0.084	-0.166	0.005
10	-1.028	-0.085	0.060	-0.057

Notes: This table is based on the set of working HRS respondents in the 2010 wave. Measurement error is the difference between MI-optimized measures conditional on the sample where HRS self-reports are nonmissing and the averages based on HRS self-reports. Nonresponse bias is the difference between ML-MI-optimized measures for the whole sample and ML-MI-optimized measures conditional on nonmissing HRS self-reports. See Appendix C.4 for details.

APPENDICES

APPENDIX A

Appendix to Chapter 1

A.1 Data Appendix

Form 5500

F5500 is an annual plan specific filing collected jointly by the IRS, DoL, and the PBGC to ensure compliance with ERISA.[1] Each plan in the F5500 database is identified by a combination of an EIN and plan number (PN). The PN is assigned by the plan's sponsor and stays fixed over the life of the plan. For form years 2000-2015, the DoL has prepared an edited research sample of the data in which logical and arithmetic errors are corrected and multiple filings for the same plan are de-duplicated. From 2000-2009, the research data include all pension plans with more than 100 participants and a 5 percent sample of plans with less than 100 participants. I use records from the research data where possible and add back small plans (i.e. those with less than 100 participants) from the raw data if they are excluded from the research sample. I de-duplicate multiple filings for the same plan in the raw data files by retaining the most recent filing in a given year. I obtained pre-1999 data through a FOIA request to the DoL. The sample that I use covers plan years ending 1996-2014.

I focus primarily on DB plans, but also obtain data on DC plans offered by employers who sponsor DB plans. Plan characteristics are coded using a set of numbers and letters. In post-1999 F5500 data, DB plans have prefix 1, DC plans have prefix 2 and 3, and welfare

[1] F5500 data for form years after 1999 are available at `https://www.dol.gov/agencies/ebsa/about-ebsa/our-activities/public-disclosure/foia/form-5500-datasets`.

benefit plans — such as employer provided health insurance — have prefix 4.[2] Hard frozen DB plans are recorded using code 1I. Cash balance plans are recorded using code 1F. I eliminate supplemental plans which I identify by searching for any case versions of the string "supplemental" in the plan name field. I restrict my sample to single employer plans, thereby eliminating multi-employer DB plans.[3]

In addition to the main form, I include data taken from the actuarial information attachment. Prior to 2007 this attachment was labeled Schedule B. After 2009 it was labeled Schedule SB for single employer plans. For 2008, I impute actuarial data by interpolating between the 2007 and 2009 values because actuarial information is unavailable in electronic format. The actuarial attachment contains important plan level data including detailed breakouts of plan assets and liabilities, accruals earned during the plan year, the average retirement age/benefit claim age, mortality and separation rate assumptions, etc.

Linking to the Census Business Register and Census Longitudinal Business Database

The BR is a database of the universe of establishments in the United States.[4] It includes information on business location, organization, industry, and information on revenue, payroll, and employment that is collected from administrative tax records as well as survey information. The relationship between establishments belonging to multi-unit firms are determined using responses to the company organization survey, the economic census, and the annual survey of manufactures. Establishments that are part of the same multi-unit firm share the same Census assigned firm identification number even if they have unique EINs.

To link the F5500 files to the BR, I match EIN-plan-end-years in F5500 to EIN-years in the BR. Because the massive scope of the BR, I am able to match approximately 92 percent of DB plan-years in the F5500 files to specific establishments in the BR (see row 1 of Table A.1). Non-matches occur when a plan EIN does not map to any establishment with positive payroll in the BR which could happen, for example, when a plan is sponsored by a union or an employer-association.

The set of plan-years represented in the F5500-BR merge contains a mix of firms that sponsor just one DB plan and firms that sponsor multiple DB plans.[5] I limit my sample to

[2]The codes are different prior to 1999. Welfare benefit plans only need to be reported with a F5500 filing when such plans cover more than 100 active participants.

[3]Multi-employer plans are arrangements between a group of firms and/or unions to provide pension benefits to eligible employees within the group.

[4]Information about the BR is confidential and protected by Title 13 and Title 26, US Code. The following information is drawn from `https://www.census.gov/econ/overview/mu0600.html`.

[5]Firms that sponsor multiple DB plans typically do so to cover different types of workers. For example, a firm may sponsor different DB plans for salaried and hourly workers or unionized and non-unionized workers.

firms that have a single plan within the 1996-2014 window for which I have F5500 data. When firms have multiple plans, I retain only those employers who choose either to never freeze their plans, or freeze them all at the same time. The principle driver of this restriction is that I cannot observe individual pension plan coverage. Consequently, when firms sponsor multiple plans, there is no way of knowing — using the F5500, LBD, or LEHD data — which plan a worker may be covered by. By imposing this restriction, however, I can ascertain whether workers at a given firm have been affected by a freeze in a given year. This sample restriction allows me to retain 94 percent of firm-years but only about 40 percent of worker-years (see row 2 of Table A.1). The discordance between these two rates reflects the fact that only the very largest employers sponsor multiple DB plans.[6] Using Census firm identifiers, I match these data with the LBD which is a cleaned and research ready version of the BR. The LBD covers private sector establishments with non-zero payroll but excludes some industrial sectors (see p.4 of Jarmin and Miranda (2002) for details).

Having matched F5500 records to the BR and the LBD, I structure the data as follows. I treat each year from 2001-2014 as an experimental year, which is indexed by l.[7] This terminology reflects the research design wherein each year yields a fresh sample of firm-level pension freezes. For a given experiment year, the panel dataset of workers employed at freezing firms constitutes the treated group while the panel dataset of workers employed firms that do not freeze their plans constitutes the comparison group. I impose the restriction that firms file F5500 for their DB plans 5 calendar years prior to the experiment year, which I refer to as the pre-period. By requiring plan stability in the lead up to the experiment year, I implicitly follow a specific set of workers covered by a DB plan regardless of whether their employer merges, grows from single-unit to multi-unit, or vice versa. I match information on DC plans offered by the set of DB sponsoring employers to these data using the same EIN-based linking procedure described above. The key DC-related variable is the number of workers covered by DC plan(s). When a firm offers multiple DC plans, I pick the maximum number of active participants across plans and use that count to estimate the DC coverage rate at the firm.[8]

[6]When firms sponsor multiple plans and pass the sample screen, I sum plan-level variables such as assets, liabilities, and participant counts across all plans sponsored by the firm. I compute the weighted average of the retirement age reported on F5500 using the number of participants in each plan as weights.

[7]I start with 2001 because it is the first year in which pension freezes are reported in F5500. Cash balance conversions are reported in earlier years but the number of firms making CB conversions before 2001 is small.

[8]I use the maximum across plans rather than the sum across plans because workers can participate in multiple DC plans. The DC coverage rate is the ratio of the maximum number of DC participants from F5500 to the count of employees from the LBD.

Linking to the Longitudinal Employer Household Dynamics

The LEHD is a quarterly matched employer-employee dataset constructed from state-level unemployment insurance (UI) records. The UI system covers 96 percent of wage and salary employment nationally, although the data exclude independent contractors, the unincorporated self-employed, railroad workers covered by railroad unemployment insurance, and some other minor categories of workers who are not covered by state-level UI laws. State and local government employees are included in the data but elected officials, members of the judiciary, and some emergency employees are excluded. Federal government workers and workers employed in Alabama are excluded from the version of the data that I use in this paper.

An important feature of the LEHD is that states enter the dataset at different points in time. For example, Maryland enters in 1985:Q2 whereas Mississippi enters only in 2003:Q3. Because of staggered entry, the scope of the data grows continuously over time.[9] I use the 2014 snapshot version of the LEHD, which provides matched employer-employee histories from each state's entry quarter up through 2015:Q1. I eliminate the single quarter of 2015 from these data as it represents partial year information on earnings and is not representative of the annual data structure that I employ.

In the LEHD, employers are identified using a state UI account number known as the SEIN while workers are identified using a variable known as a protected identification key (PIK). I begin by matching firm-level data from the F5500-LBD linked sample to the T26 Employer Characteristics File (ECFT26) in the LEHD. The ECFT26 is a SEIN-quarter-year level file that contains the Census firm identifier associated with each SEIN. Using this common unique identifier, I can match national plan- and employer- level characteristics from the F5500-LBD linked sample to state level employers in the LEHD. I recover 89 percent of firm-experiment years which corresponds to 93 percent of employee-experiment years from the F5500-LBD linked sample (see row 3 of Table A.1). From this sample, I drop a small percentage of observations where certain pension plan variables are missing (see row 4 of Table A.1).[10]

Worker sample

When considering the implications of pension freezes on worker decisions, it is important to reiterate that I do not observe individual information on pension plan coverage. To study

[9] A number of populous states enter the data relatively early. Illinois enters in 1990, California and Pennsylvania in 1991, Florida in 1992, and New York and Texas enter in 1995.

[10] Critical pension plan information includes plan assets and liabilities, accruals earned in the filing year, the average benefit claim age for the plan.

worker responses in a way that limits the potential for misclassification error, I restrict the sample to firms where DB eligibility is near universal. I impose this restriction by retaining firms where the DB coverage rate is 80 percent or greater in the pre-period.[11] Within the sample of high-coverage rate firms, I use the LEHD Employment History File (EHF) to obtain matched employer-employee data. The EHF is a SEIN-PIK-year level file that provides the earnings history associated with each employer-employee combination. To these data, I add information on date of birth, race and ethnicity, and education from the Individual Characteristics File (ICF). I then select all workers employed at a DB sponsoring firm in $l-5$, who have at least two years of tenure as of $l-5$, and who will be between the age of 50 and 70 in the experiment year .[12]

Table A.1: Data linkage and sample restrictions

Match/restriction type	N	Match rate	Person weighted rate
F5500-BR (plan-years)	852000	0.922	0.947
Multi-plan restriction (firm-years)	699000	0.942	0.378
LBD-LEHD (firm-experiment years)	1419000	0.885	0.933
No missing pension data (firm-experiment years)	1256000	0.930	0.972

Notes: F5500 data are based on years ending 1996-2014. Pension data is treated as missing if plan liabilities, assets, accrual amounts, and claim ages are either missing or unreadable in electronic format and cannot be interpolated.

A.2 Imputing retirement in the LEHD

In the LEHD, a worker is employed if they have positive earnings in a given year. The definition of retirement is somewhat more involved. I classify an individual as having retired in year t if the last year in which she received non-zero earnings was $t-1$. In this definition, retirement only occurs when an individual completely withdraws from paid employment. Recall, however, that both definitions exclude work in the form of self-employment because the LEHD data are based on UI covered earnings. To examine the potential for misclassification of retirement in administrative data, I compare the retirement

[11]The firm-wide DB coverage rate is the ratio of active participants in the plan as reported in F5500 to the count of total employees in the LBD. The 80 percent average coverage rate requirement is based on years $[l-5, l-2]$ — i.e. between 5 and 2 years prior to the experiment year. Restricting the sample this way likely eliminates soft freezes in which the firm's plan is closed to new workers. A firm that imposes a soft freeze is likely to see its DB coverage rate decline as workers quit or retire but are not replaced with new eligible workers.

[12]The two year tenure restriction ensures that workers are fully vested in their pensions as of the experiment year when they may become subject to a freeze. This calculation is based on the 7 year maximum full vesting period allowed for DB plans by ERISA.

rate of employed, DB eligible, respondents from the 2004 wave of the HRS with a comparable sample of individuals in the LEHD drawn from the 2004 experiment year whose pensions have not been frozen.[13]

These comparisons between HRS and LEHD data are shown in the three panels of Figure A.1, which split the samples into three age categories as of 2004. The retirement rates align very well for the 56-64 year-old age group but diverge somewhat for the 50-55 and 65-70 year-old age groups. For 65-70 year old individuals, the HRS-based rates are lower than the LEHD-based rates which potentially reflects the fact that self-employment is not covered in the LEHD. For the 50-55 year old age group, the HRS-based rates are higher than the LEHD-based rates. This discordance exists even when HRS retirement rates are constructed from a question asking respondents if they have zero earnings.

A.3 Firm and plan characteristics around the freeze

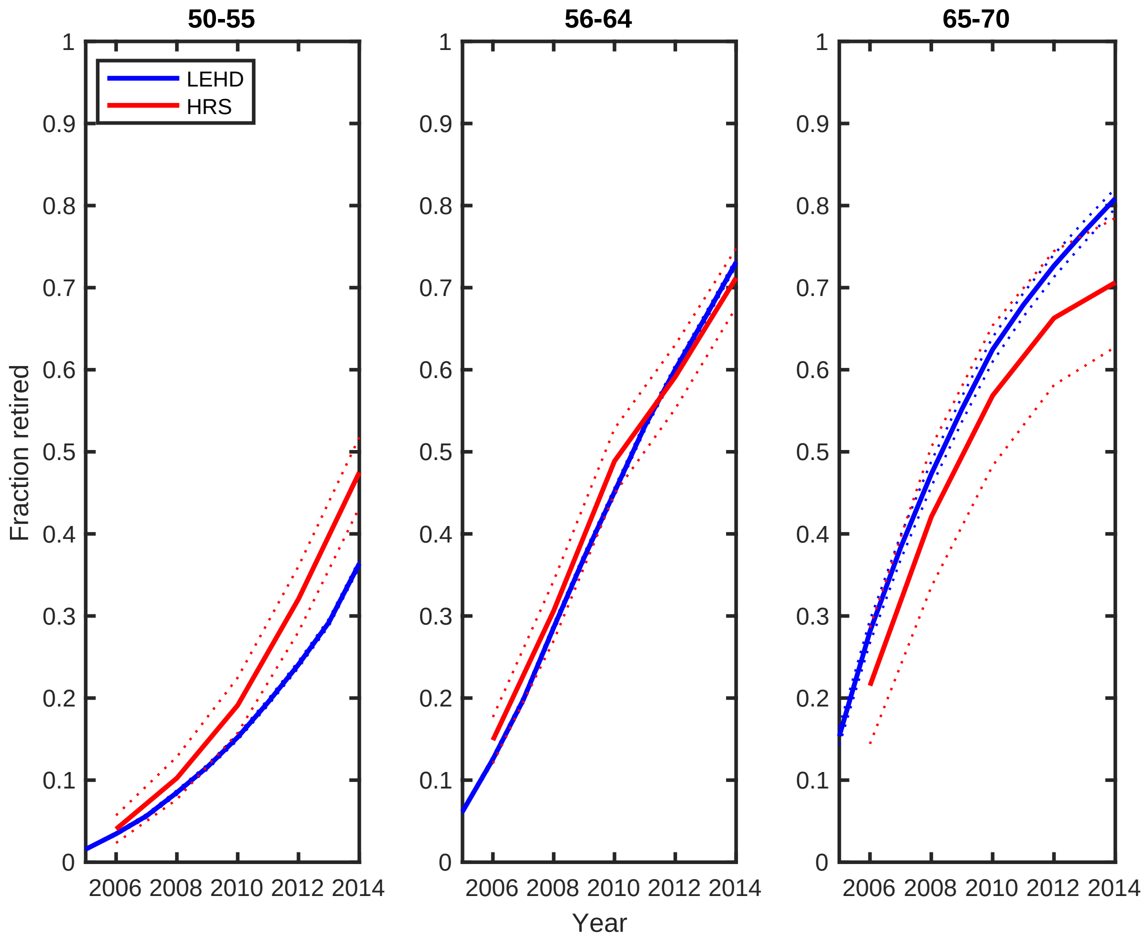
This appendix describes how firm and plan characteristics evolve around the freeze. I establish three facts. First, I show that worsening plan finances rather than worsening firm performance is the main predictor of DB pension freezes. Second, I show that the aftermath of a freeze leads to small but persistent reductions in firm size and average pay. Third, I show that DB freezes generate an immediate transition towards DC plan participation.

A.3.1 Pre-freeze environment

To describe the environment prevailing prior to the firms freeze decision, I show several characteristics of firms and their pension plans averaged over a five year pre-period in Table A.2. Comparing the left and right panels of the table shows the freezing firms are very similar in terms of size, average pay, employee age structure, and DB pension plan characteristics. DB and DC coverage rates within the two sets of firms are both approximately 70 and 35 percent respectively. The lack of any meaningful difference in employee age structure, pension liability, pension accrual rates, and claim ages indicates that the freezing firms are not disproportionately staffed by older workers at the threshold of retirement. Put differently, firms that freeze their plans are not on the brink of a large liability cliff. The likelihood of experiencing a mass layoff, which is recorded as a 30 percent reduction in employment and

[13] Retirement in the HRS is inferred from a respondent's labor force status report. To align with the LEHD-based definition of permanent departure from paid employment, I consider an HRS respondent as retired only if they continually report their labor force status as retired. In this definition, a respondent who reports being retired in 2006 but re-enters the labor force in 2010 will not be counted as a retiree in the dataset. The labor-force-status-based retirement statistics are very similar to those constructed from a question asking respondents if they have zero earnings from employment.

Figure A.1: HRS versus LEHD retirements



Notes: HRS data are based on respondents who are working as of 2004 and are eligible for employer sponsored DB pension plans. LEHD data are based on the sample of individuals employed at DB sponsoring firms as of 2004 where the firm-wide coverage rate is $\geq$ 80 percent. Sample splits are based on age as of 2004.

labeled firm distress, is about 5.5 percent in both groups. Economic distress driven by a large negative shock in the output market is therefore not a leading reason for freeze decision either.

They key distinction between freezing and non-freezing firms lies in the financial health of their DB plans. For every dollar in future liabilities, non-freezing firms have \$1.10 in assets. The same ratio — referred to as the funding ratio — is 1.05 for firms that ultimately freeze their plans. The PPA designates plans with funding ratios under 80 percent as being "at-risk" or distressed. Using the PPA's threshold, 20 percent of freezing firms are have distressed plans whereas the same rate is 15 percent for non-freezing firms. Funding deficiencies are particularly important from a cost management perspective because gaps must be closed to meet statutory requirements. Furthermore, once underfunded, plans are no longer buffered against financial market shocks the way overfunded plans are. Required contributions towards underfunded plans therefore become larger and more volatile in the face of market risk.

Table A.3 shows coefficients from a linear prediction model using freeze in the experiment year as the outcome and a variety of pre-event characteristics as predictors. The regressions are estimated on data pooling over a five year pre-period, thereby allowing for the inclusion of firm fixed effects. Columns 1 and 2 do not include firm fixed effects, while columns 3 and 4 do. The regressions show that a 1 percent improvement in the funding ratio lowers the likelihood of a future freeze by 2.5 percentage points. This partial effect is stable and statistically significant across all four specifications. Firm size is negatively correlated with future freezes, but the magnitude of the effect is small: a 1 percent increase in firm size lowers the likelihood of a future freeze by 0.2 to 0.5 percentage points. In specifications with firm fixed effects, employee age structure has no statistically significant impact on freezes and the magnitudes of the partial effects are negligible when expressed in proportional terms. DB plans that are collectively bargained are about 2.5 percentage points less likely to experience a freeze which implies that unions offer approximately the same protective effect as a one percent improvement in plan funding. Industry fixed effects, which are included in columns 2 an 4 have no appreciable impact on the estimated coefficients indicating that industry-specific factors are not important, conditional on the other predictors in the model.

A.3.2 Post-freeze changes

Figure A.2 compares the evolution of four variables between freezing and non-freezing firms before and after the freeze. Each panel plots coefficients from an event study regression using the specification described in equation (1.5.2) with firm-cohort-year level data. Note that the estimated coefficients are net of firm fixed effects and therefore remove time invariant unobserved heterogeneity between firms. In each panel, the horizontal axis represents the calendar year relative to the experiment year.

The upper row of Figure A.2 shows the difference in log of total employment and log of average pay between freezing and non-freezing firms. The estimated coefficients show that freezes lead to a persistent 2.5 percent reduction for both outcome variables — a gap which closes only after about 10 years. Post-freeze differences in size and pay between the two types of firms represent some combination of changes to the age or seniority composition of the firm's workforce after a freeze either through labor supply responses or changes in labor demand, although it is not clear from firm-level data alone what the role for each channel is.[14] The lower left panel of Figure A.2 shows that the fraction of DC covered workers starts to rise about 2 years prior to the DB freeze reflecting expanded DC plan eligibility, more generous DC match rates, or transitions to opt-out rather than opt-in DC enrollment.[15] In the period right around a DB freeze, DC coverage rates increase by about 5 percentage points off a baseline coverage rate of about 35 percent. In subsequent years, non-freezing firms gradually increase their DC coverage and catch up to the DC coverage rate prevailing at freezing firms. Whether the catch up occurs through soft-freezes that close existing DB plans to younger workers, or through more generous incentives for DC participation, the results shown here provide evidence that DB freezes accelerate the inevitable transition towards DC pension coverage within firms. Evidence for increased DC participation is important in explaining the extended labor force participation of some freeze-affected workers as it allows them to offset DB losses.

The lower right panel of Figure A.2 shows the change in the likelihood of a freezing firm to experience economic distress, which is defined as a reduction in employment of 30 percent or greater. The coefficient estimates from an unweighted regression (in blue) show that the immediate aftermath of a freeze induces a 1.5 percentage point increase in the probability of distress which lasts for two years. When the same regression is weighted by firm size (in red), thereby representing the change in the probability of freeze affected workers experiencing large employment contractions, the point estimates are economically and statistically insignificant. As such, it appears that distress is concentrated among smaller firms.

[14] The person-level analyses presented in Section 1.6 isolate labor supply factors by using propensity score methods to condition on the pre-freeze path of firm size which serves as a proxy for latent changes in output demand for the firm.

[15] DC coverage is measured as the ratio of the maximum number of participants across a firm's DC plans to total employment.

Table A.2: Pre-period firm and plan characteristics

	Non-freezing firms		Freezing firms	
Variable	Mean	Std. error	Mean	Std. error
Size	273.7	3.3	254.6	19.5
Average earnings ($)	69300	252	68010	520
Firm age	20.9	0.0	21.2	0.1
Multi-unit	0.273	0.001	0.250	0.003
Fraction workforce ≤ 45	0.579	<0.001	0.578	0.001
Fraction workforce [46,50]	0.117	<0.001	0.115	0.001
Fraction workforce [51,55]	0.110	<0.001	0.109	0.001
Fraction workforce [56,60]	0.092	<0.001	0.095	0.001
Fraction workforce [61,65]	0.057	<0.001	0.060	0.001
Fraction workforce [66,70]	0.024	<0.001	0.023	<0.001
Fraction workforce ≥ 71	0.022	<0.001	0.020	<0.001
Distressed firm	0.056	<0.001	0.057	0.001
DC plan offered	0.502	0.001	0.521	0.003
DC plan coverage rate	0.348	0.001	0.362	0.002
DB plan coverage rate	0.723	0.000	0.708	0.001
DB pension wealth/ptcp ($)	98910	372	101600	1612
DB pension accrual/ptcp ($)	14200	41	14690	145
Average benefit claim age	63.2	0.0	63.3	0.0
Collectively bargained plan	0.040	<0.001	0.030	0.001
Funding ratio	1.11	0.00	1.05	0.00
Distressed plan	0.159	0.001	0.193	0.002
Firm-experiment years	428000		28500	
Firms	22500		6500	

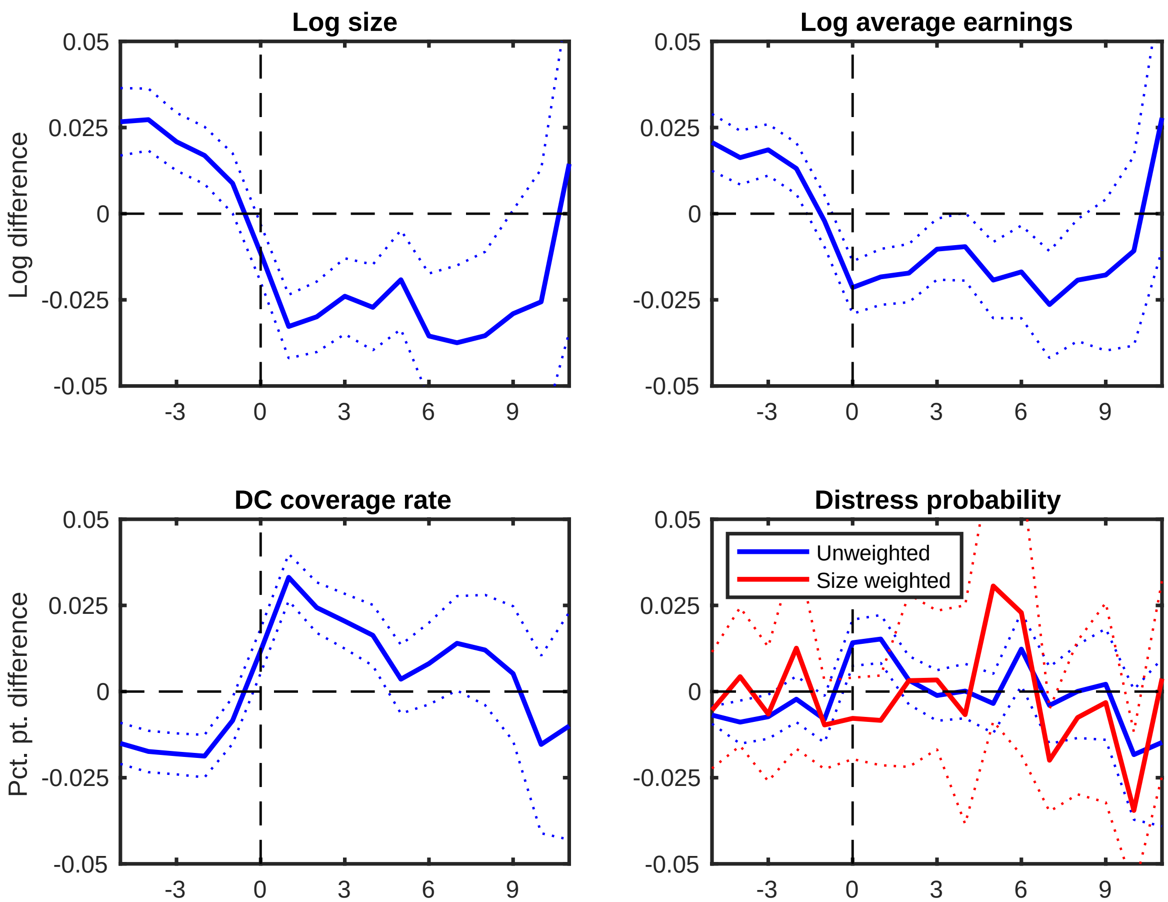
Notes: Statistics reported in the table average over the five year period preceeding any freeze activity. All dollar values are expressed in 2010 terms. Pension wealth per participant is computed as the present value of the liability owed to active participants divided by the number of active participants. Plan-years are coded as distressed if their ratio of assets to liabilities is under 80 percent — the threshold below which DB plans are considered "at risk" in the Pension Protection Act of 2006. Firm-years are coded as distressed if firm-wide year-on-year employment shrank by 30 percent or more.

Table A.3: Predictors of future freezes

Variables	(1)	(2)	(3)	(4)
Log funding ratio	-0.0264***	-0.02545***	-0.0258***	-0.0258***
	(0.0020)	(0.0020)	(0.0022)	(0.0022)
DB coverage rate	-0.0075**	-0.0029	0.0008	0.0009
	(0.003)	(0.0030)	(0.0036)	(0.0036)
Firm age	-0.00009	0.00001	0.00027	0.00034
	(1.04e-04)	(1.07e-04)	(5.49e-04)	(5.40e-04)
Log size	0.0026***	0.0021***	-0.0055**	-0.0056**
	(6.32e-04)	(6.51e-04)	(0.0024)	(0.0024)
Log average pay	0.00004	-0.00063	-0.00591***	-0.00598***
	(0.00114)	(0.00117)	(0.00182)	(0.00181)
Fraction workforce $\leq$ 45	0.0168*	0.0178*	-0.0141	-0.0143
	(0.0098)	(0.0099)	(0.0124)	(0.0124)
Fraction workforce [46,50]	0.0097	0.0119	-0.0115	-0.0117
	(0.01083)	(0.0109)	(0.013)	(0.013)
Fraction workforce [51,55]	0.0085	0.0114	-0.0217*	-0.022*
	(0.01068)	(0.0107)	(0.0131)	(0.0131)
Fraction workforce [56,60]	0.0223**	0.0251**	-0.0106	-0.0108
	(0.01087)	(0.0109)	(0.0129)	(0.0129)
Fraction workforce [61,65]	0.0253**	0.0272**	0.0037	0.0036
	(0.0116)	(0.0116)	(0.0131)	(0.0131)
Fraction workforce [66,70]	0.0024	0.0036	0.0069	0.0069
	(0.0124)	(0.0124)	(0.0123)	(0.0123)
DB plan collectively bargained	-0.0215***	-0.0257***		
	(0.0035)	(0.0036)		
Firm offers DC plan	0.0016	0.0030*		
	(0.0016)	(0.0016)		
Multi-unit firm	-0.0119***	-0.0069***		
	(0.0023)	(0.0024)		
Observations	456000	456000	456000	456000
Adjusted R-squared	0.015	0.016	0.362	0.362
Firm FE	No	No	Yes	Yes
Experiment year-calendar year FE	Yes	Yes	Yes	Yes
Industry FE	No	Yes	No	Yes
Number of firm clusters	23500	23500	23500	23500

Notes: Robust standard errors, clustered at the firm level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Regressions are estimated on a panel dataset that pools the five year period preceeding any freeze activity .

Figure A.2: Firm characteristics around freezes



Notes: Dotted lines show 95 percent confidence intervals which are based on standard errors clustered at the firm level. Horizontal axes show years relative to the experiment year. Firm-years are coded as distressed if firm-wide year-on-year employment shrank by 30 percent or more.

A.4 Propensity score re-weighting

For workers in cell $j(i)$, denote the probability of experiencing a freeze, or the propensity score, by $\hat{p}(\boldsymbol{z}_{j(i)})$. $\boldsymbol{z}_{j(i)}$ is a vector including all pre-period observations on log firm size, firm-level averages of log total pension wealth and log pension accruals per working participant, average benefit claim age from the pension plan, the age structure of the firm's workforce, cell-level annual earnings, retirement rates, labor force participation rates, and employer-to-employer transition rates. I also condition on state, gender, tenure, and earnings averaged over the workers LEHD history prior to $l-5$. $\hat{p}(\boldsymbol{z}_{j(i)})$ is estimated using logistic regression.

In the setting being considered in this paper, the parameter of interest is the average treatment effect on the treated (ATET) — i.e. the impact of pension freeze shocks on the labor supply of workers affected by those shocks. To estimate the ATET, each comparison group unit is re-weighted by $\frac{\hat{p}(\boldsymbol{z}_{j(i)})}{1-\hat{p}(\boldsymbol{z}_{j(i)})}$. Following Busso et al. (2014), the weights are first normalized to sum to 1 so that the number of weighted units in the comparison group is unaffected by the re-weighting procedure.

Constructing good counterfactuals for treated group units requires that comparison group units with the same $\hat{p}(\boldsymbol{z}_{j(i)})$ — i.e. the same ex-ante probability of experiencing the treatment — can be found in the sample. This requirement is referred to as the "common support condition" or the "overlap condition." Formally, the common support condition for the ATET parameter requires that $\hat{p}(\boldsymbol{z}_{j(i)}) < 1$ for all $j(i)$. In practice, the treatment and comparison groups in the data I use share a large region of common support and the maximum $\hat{p}(\boldsymbol{z}_{j(i)})$ is lower than 1.

A.5 Solution and estimation of the structural model

A.5.1 Calibrated parameters

Most of the calibrated first-step parameters are drawn from pension eligible respondents in the HRS. The HRS provides high quality information on DB pension wealth which is computed using summary plan descriptions (SPDs) obtained either directly from respondent's employers or from attachments to F5500 administrative records. Information from these documents is coded along with relevant data on respondent's past and future earnings projections and job tenure to calculate pension wealth at prospective retirement ages. These calculations are done using a software system known as the Pension Estimation Program (PEP).[16]

[16] See Fang et al. (2016) for more details.

Linked pension data are available for the 1992, 1998, 2004, and 2010 survey wave years of the HRS. While the 2004 and 2010 samples are most relevant because they coincide with the time period that I analyze, I rely exclusively on the 2010 linked pension sample for two reasons. First, unlike prior waves, the 2010 wave explicitly separates public and private sector plans. By focusing on respondents working in the private sector, I am able to align the survey data to reflect private sector pension provisions which constitute the relevant subset for the analysis. Second, the 2010 sample explicitly flags a small number of CB plans as being different than conventional DB plans. This distinction allows me to isolate the sub-sample where the evolution of pension accruals occurs under the status-quo — i.e. where participants do not experience a freeze. Frozen DB plans are not common in the HRS; only 6 respondents report having experienced one in the 2010 survey wave.

In what follows, I describe a variety of parameters that are calibrated using data on pension eligible respondents in the 2010 survey wave of the HRS.

DB pension wealth accruals

DB pension wealth in the PEP is calculated using the following formula

$$W_q^{DB} = \sum_{t=q}^{119} p_t \left(\frac{1 + COLA}{1+i} \right)^{t-T_0} B_{t|q}, \tag{A.1}$$

where W_q^{DB} is the present value of pension wealth at potential retirement date q, p_t is the probability of survival in period t conditional on being alive in period q, $COLA$ (cost of living adjustment) is the plan specific annual growth rate of payments (for most DB plans in the HRS $COLA = 0$), i is the nominal interest rate, and $B_{t|q}$ is the annual pension benefit in period t conditional on retiring in period q.[17] DB pension wealth is based on the maximum of wealth at the plan's normal retirement age, early retirement age, and vested deferred value of benefits. When respondents have wealth in multiple plans, I sum pension wealth over each plan. Nominal values of W_q^{DB} are then converted to 2010 dollars.

I average across respondents in the sample to compute real DB wealth at each prospective quit age. I use the mortality and real interest rate assumptions from the PEP to convert these values into their the annuity equivalents at each quit age which are denoted by $\{b_a^{DB}\}_{a=51}^{80}$.

[17] p_t are based on the 2010 version of gender-specific cohort mortality tables published by the Social Security Administration (SSA). The 2010 pension wealth calculations assume a nominal interest rate of 5.7 percent and an inflation rate of 2.8 percent according to economic assumptions detailed in the 2010 Annual Report of the Board of Trustees of the Old Age, Survivors, and Disability Insurance (OASDI) trust funds of the SSA.

Earnings

In the PEP, earnings from the 2010 survey wave are projected forward and backward for different quit dates using the following formula

$$\log(e_q) = \log(e_{2010}) + \alpha_m(q - 2010) + \beta_1 \text{age}_q + \beta_2 \text{age}_q^2, \tag{A.2}$$

α_m is nominal wage growth and β_1 and β_2 adjust for age-based changes in earnings. I estimate β_1 and β_2 separately for men and women using the within-job earnings histories for DB eligible workers in the HRS. α_m is set at 5 percent. I convert these values to 2010 dollars and then average the earnings projections across respondents by quit age to compute an age-dependent earnings function $\{e_a\}_{a=51}^{80}$.

Social Security benefits

For respondents who have yet to claim benefits, Social Security wealth in the HRS is computed by passing the respondent's administrative earnings record along with projected future earnings into the Social Security benefit formula. The benefit levels obtained from the formula are used to compute the present value of Social Security benefits at age 62, 65, and 70 (see Fang et al. (2016)). These values are reported in 2010 dollars. Because individuals cannot claim Social Security prior to age 62, I set Social Security wealth prior to that age as 0. I linearly interpolate Social Security wealth for ages between 62, 65, and 70. I use the mortality and real interest rate assumptions from the PEP to convert these values into their the annuity equivalents at age which are denoted by $\{b_a^{SS}\}_{a=51}^{80}$. Because Social Security benefits do not increase after age 70, I set $\{b_a^{SS}\}_{a=71}^{80} = b_{70}^{SS}$.

DC match function

The employer's DC match function, $m^e(m^w)$, is based on respondent reports of their own and employer contributions expressed as a percentage of earnings in the 2010 survey wave. Recall that m^e is the fraction of earnings that employers contribute to DC accounts and is a function of the employees' own contribution rate m^w. To obtain the match function, I first bin worker contribution rates into 0.01 sized intervals and then compute the median employer contribution rate within each interval which is denoted by $\tilde{m}^e$. I then regress binned worker contribution rates on median employer contribution rates for $m^w \in [.01, .065]$. Fitted values from the regression ($E[\tilde{m}^e|m^w]$) represent the estimated match function. Approximating the data for $m^w > .065$, I assume that $E[\tilde{m}^e|m^w > .065] = E[\tilde{m}^e|m^w = .065]$ which implies that employer matching contributions stop once workers contribute 6.5 percent of earnings. The estimated match function is

$$\hat{m^e}(m^w) = \begin{cases} .011 + .39m^w & \text{if } m^w \leq .065 \\ .0364 & \text{if } m^w > .065 \end{cases} \tag{A.3}$$

Initial DC wealth and non-pension wealth

I assume that initial DC balances (W^{DC}) and non-pension assets (A) are distributed log-normally. For each initial age (i.e. between 51 and 59), I simulate assets using log-normal parameters estimated from the distribution of A and W^{DC} for pension eligible respondents in the 2010 HRS. In these calculations, A is defined as household non-pension assets inclusive of spousal DC wealth, whereas W^{DC} is the respondent's DC wealth. log-normal parameters are estimated on the sample or respondents where A and W^{DC} are non-negative. I then use these parameter estimates to simulate W^{DC} and A for workers initially aged 51-59.

Additional parameters

The discount factor (β) and real interest rate (r) are based on values consistent with the PEP assumptions.[18] The maximum DC contribution limit, $\bar{C}$, is calibrated using 2010 IRS rules. Age specific mortality rates expressed as the probability of dying within one year, $\{p_a\}_{a=51}^{80}$, are based on the 2010 actuarial life table published by the Social Security Administration (SSA). I average the gender-specific rates into a combined rate for each age and then divide by p_{81} to impose a maximum age of 80 for all simulated individuals. Finally, I use the NBER TAXSIM calculator to compute income and payroll tax liabilities when solving the model.

Table A.4 provides a summary of all the calibrated parameters of the model and their sources.

A.5.2 Solution algorithm

The model does not have an analytical solution, so I solve it numerically by backward recursion. The state variables in the model are (a, e, g, A, b, W^{DC}) of which g, A, and W^{DC} are continuous and exhibit both within- and between-age heterogeneity. I discretize the continuous state variables over finite dimensional grids of size 6 for g, 20 for A, and 20 for W^{DC}. Grid points for A and W^{DC} are narrowly spaced for low values and widely spaced for high values. Iterating backward from the final age for a given value of $\boldsymbol{\theta} = (\sigma, \gamma, \phi, \rho, \sigma_v)$:

[18] The real interest rate is 2.9 percent which is based on the 2010 SSA OASDI trust fund report assumption. I assume that the annual discount factor is 0.97 to be consistent with a real interest rate of approximately 3 percent.

1. I compute the probability distribution over next period's work disutility which arises from randomness in the AR(1) component f. I employ the Rouwenhorst (1995) approximation for the AR(1) term to obtain $P(f' = f_{l'} | f = f_l)$ for $l, l' = 1, \ldots, 6$. I then use these probabilities to estimate

$$E_{f'}\left[\max\left\{V^R_{a+1}(a+1, b', A', W^{DC'}), V^W_{a+1}(X')\right\}\right] \tag{A.4}$$

At age 80, $V^W = V^R = 0$ as no workers live beyond that age and there are no bequests.

2. For each value of the state variables (within the current age iteration), I compute V^W and V^R and the associated decision rules as follows:

   (a) To obtain V^W, I compute the income and payroll tax rate faced by a worker for each potential $W^{DC'}$ choice given the earnings level for the current age. I then search over the $W^{DC'}$ and A' grid points to find the highest value of V^W . I impose the maximum contribution constraint ($\bar{C}$) and preclude decumulation of DC balances while working by setting the associated values of V^W to negative infinity. This grid search defines the decision rules for DC accumulation and non-pension saving (or dissaving) while working.

   (b) To obtain V^R, I compute the income tax rate faced by a retiree for each potential $W^{DC'}$ choice given the annuity income obtained by retiring at the current age. I then search over the $W^{DC'}$ and A' grid points to find the highest value of V^R. I preclude accumulation of DC balances in retirement by setting the associated values of V^R to negative infinity. This grid search defines the decision rules for DC decumulation and non-pension saving (or dissaving) while retired.

   (c) I compute the retirement decision for each value of the state variables by comparing V^W to V^R. If $V^W \leq V^R$, then the work decision is retirement. Else if $V^W > V^R$, then the work decision is to stay employed.

3. The solution algorithm terminates when value functions and decision rules have been obtained for each age.

I compute a separate set of value functions and decision rules by imposing pension freezes at ages 56-64. A worker who experiences a pension freeze at age a has $b^{DB}_{a+k} = b^{DB}_a$ for all $k > 0$. I apply earnings reductions as estimated from LEHD data for workers who are affected by freezes between the age of 56 and 64.[19] To the extent that career lengths change due to the

[19] Earnings losses occur in the year of the freeze and in the next four years. The estimates are -3.8 percent, -2.1 percent, -1.5 percent, -4.9 percent, and -1.6 percent. See the middle panel of Figure 1.6.

freeze, I assume that they do not affect the value of Social Security wealth. This assumption reflects the fact that most workers over 55 already have long work histories and do not accrue substantial increases in Social Security benefits through continued work. Other than the change to b^{DB} and the earnings path, all other parameters remain the same.

A.5.3 Estimation

Having solved the model for a given value of $\boldsymbol{\theta}$ by obtaining decision rules for each age, I simulate data as follows

1. I simulate initial assets and work disutility draws for 5000 individuals who are initially aged 51 to 59.[20] I apply decision rules from the no-freeze scenario to obtain work histories and asset accumulation paths to create a simulated control group. I use linear interpolation to infer the optimal decision rules for simulated values of A, W^{DC}, and g that lie between the grid points.

2. Next, I apply the freeze decision rules for the same population of individuals — i.e. individuals with the same initial assets and work disutility draws —starting five years after the initial age. Individuals in this exercise have the same work and asset accumulation choices as the simulation control group for the first five years, but have different work histories and asset accumulation choices once faced with a DB freeze. I call this sample the simulated treated group.

3. I compute two sets of moments using the simulated control and treated groups. The first set of moments is the average employment rate by age for the simulated control group. I compute the control group levels starting four periods prior to the freeze and lasting 6 periods after the freeze. The second set of moments is the difference in average employment rates between the two groups (the simulated treatment effect). I compute these differences for 12 periods starting from the period of the freeze.

Denote the vector of simulated moments by $\hat{\boldsymbol{h}}^S(\boldsymbol{\theta})$. Denote the analogous vector of observed moments from real-world (LEHD) data by $\hat{\boldsymbol{h}}^D$. The distance between the simulated and real-world moments is

$$\boldsymbol{m}(\boldsymbol{\theta}) = \hat{\boldsymbol{h}}^S(\boldsymbol{\theta}) - \hat{\boldsymbol{h}}^D. \tag{A.5}$$

[20] The share of individuals of each initial age is equal to the share of DB eligible working respondents in 2010 HRS.

The MSM estimate of $\boldsymbol{\theta}$ is given by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\, \boldsymbol{m}'(\boldsymbol{\theta})\boldsymbol{W}\boldsymbol{m}(\boldsymbol{\theta}) \tag{A.6}$$

where $\boldsymbol{W}$ is a weighting matrix.

I estimate $\hat{\boldsymbol{\theta}}$ using a two step procedure. In the first step, I set $\boldsymbol{W} = \zeta\boldsymbol{I}$ where $\boldsymbol{I}$ is the identity matrix and ζ is a scaling vector. The scaling vector re-weights treatment effect moments by a factor of 10 to give them approximately the same numerical importance as the employment trend moments. I make this adjustment because the treatment effect moments are economically very informative but are numerically an order of magnitude smaller than the employment trend moments. Denote by $\hat{\boldsymbol{\theta}}_1$ the parameter vector obtained using the first step weighting matrix which I estimate using the Nelder-Meade algorithm.[21] In the second step, I re-estimate the simulated moments 500 times holding $\hat{\boldsymbol{\theta}}_1$ fixed but re-drawing random components of the simulation that introduce sampling variability (i.e. g and the initial values of A and W^{DC}). Using this parametric bootstrap procedure, I compute the variance-covariance matrix of the simulated moments which is denoted by $\hat{\boldsymbol{S}}(\hat{\boldsymbol{\theta}}_1)$. I then estimate $\hat{\boldsymbol{\theta}}_2$, the final parameter estimate, by setting $\boldsymbol{W} = \hat{\boldsymbol{S}}^{-1}(\hat{\boldsymbol{\theta}}_1)$.

To obtain standard errors, I compute $\hat{\boldsymbol{D}} = \frac{\partial \boldsymbol{h}^S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} |_{\hat{\boldsymbol{\theta}}_2}$ numerically by using 10 small random deviations around $\hat{\boldsymbol{\theta}}_2$ and re-calculating $\boldsymbol{h}(\boldsymbol{\theta})$ at each perturbed value. I then average $\frac{\partial \boldsymbol{h}^S(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}$ across the perturbations to compute

$$\boldsymbol{Q} = (1 + N_S^{-1}) \left[\hat{\boldsymbol{D}}'\hat{\boldsymbol{S}}^{-1}(\hat{\boldsymbol{\theta}}_1)\hat{\boldsymbol{D}}\right]^{-1} \tag{A.7}$$

where $N_S = 1$ is the number of simulations associated with the estimate of $\hat{\boldsymbol{\theta}}_2$. The square root of the diagonal of $\boldsymbol{Q}$ is the vector of standard errors for $\hat{\boldsymbol{\theta}}_2$ (Duffie and Singleton (1993)).

[21] I use a variety of starting values to avoid the possibility of the parameter search procedure converging to a local minimum.

Table A.4: Calibrated parameters

Parameter	Value	Description	Source
r	0.029	Real interest rate	HRS PEP/SSA
β	0.97	Annual discount factor	$r \approx .03$
$\bar{C}$	\$54,500	Maximum combined contribution limit to DC plans (2010 level)	IRS
$\{\mu_a^A, \sigma_a^A\}_{a=51}^{59}$		Scale and shape parameters for non-pension wealth distribution for initial ages 51-59	HRS
$\{\mu_a^{W^{DC}}, \sigma_a^{W^{DC}}\}_{a=51}^{59}$		Scale and shape parameters for DC wealth distribution for initial ages 51-59	HRS
$E[\tilde{m}^e \vert m^w]$		Employer DC contribution rate expressed as a function of the worker's contribution rate	HRS
$\{b_a^{DB}\}_{a=51}^{80}$		Age-specific DB pension annuity	HRS
$\{b_a^{SS}\}_{a=51}^{80}$		Age-specific Social Security annuity	HRS
$\{p_a\}_{a=51}^{80}$		Age-specific mortality rate adjusted so that $p_{81} = 1$	SSA
$\{e_a\}_{a=51}^{80}$		Age-specific annual earnings	HRS
$\tau(\cdot)$		Federal income and payroll tax liabilities (2010 laws)	NBER TAXSIM

Notes: See Appendix A.5.1 for details.

A.6 Simulated pension freezes in the HRS

This Appendix describes how I construct the simulation underlying Figure 1.3. The simulations are based on the sample of DB eligible HRS respondents from the 2010 survey wave who are employed in the private sector and are not in hybrid or CB plans as of the survey date.[22] Using these data, I consider three age-specific components of compensation: DB wealth, earnings, and DC wealth.

For survey respondents, the PEP provides estimates of DB wealth at each potential quit date on the basis of equation (A.1). In addition, the PEP provides earnings projections at each potential quit date for each respondent on the basis of equation (A.2). DC wealth is reported by respondents as of the survey year (i.e. 2010). I use these measures along with self-reported own and employer contribution rates (m^w and m^e) to construct estimates of past and projected future values of DC wealth using the law of motion described in equation (1.3.4).[23] Notably, DC wealth in 2010 is 0 for approximately half of the sample. I then combine these estimates of DB wealth accruals, earnings, and DC wealth accruals to compute total compensation at each potential quit age using equation (1.3.7).[24] Finally, I average the data across respondents and quit dates to obtain the "no-freeze" path of compensation.

To simulate a hypothetical DB pension freeze for workers aged a^F, I make two changes. First, I assume that the annuity value of nominal DB wealth is frozen as of a^F. This is equivalent to receiving no new accruals either due to tenure increases or due to earnings growth.[25] Second, I assume that respondents with no DC wealth as of a^F — i.e. about half the sample — begins contributing to a hypothetical new DC plan starting at age a^F. I assume that contribution rates for these workers are equal to the sample averages of m^w and m^e for respondents with non-zero DC wealth. Respondents who have non-zero DC wealth are assumed to continue contributing at the same rate as they did prior to age a^F. Having defined post-freeze DB and DC wealth evolution, I combine the estimates of post-freeze DB wealth accruals earnings, and DC wealth accruals to compute total compensation at each potential quit age using equation (1.3.7). I then average these data across respondents and

[22]Frozen DB plans are not common in the HRS; only 6 respondents report having experienced one in the 2010 survey wave. I assume that all non-hybrid and non-CB plans in the 2010 sample are not frozen.

[23]I assume that respondents and their employers contribute at the same rate in all years. I rely on the RAND HRS files which convert respondent reports of own and employer contributions to percentages of earnings.

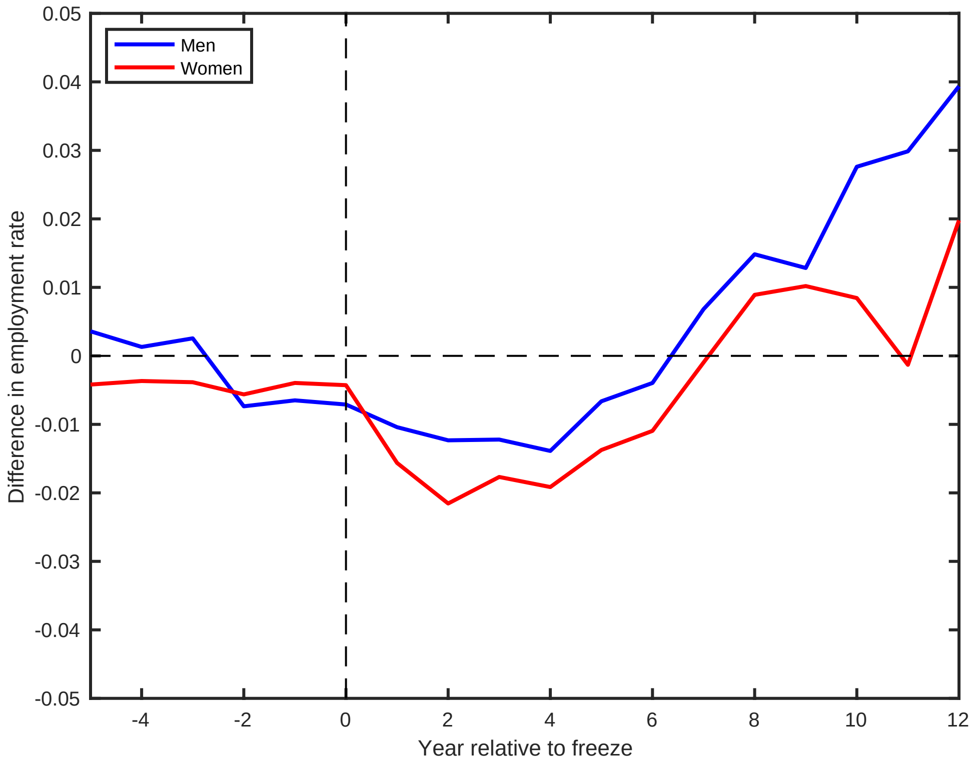
[24]All three components (i.e. DB wealth, earnings, and DC wealth) are converted to 2010 dollars. I do not account for taxes in these simulations because online tax calculators cannot be used within the restricted setting in which pension data are accessed.

[25]By preventing future growth in the nominal value of benefits, DB freezes have the effect of lowering the present value of DB wealth in each year subsequent to the freeze, thereby generating negative future accruals. This is because benefits stay fixed but the horizon over they can be collected falls.

quit dates to obtain the "Freeze at a^F" path of compensation. I conduct these calculations for different values of a^F to obtain different post-freeze paths of total compensation.

A.7 Supplementary tables and figures

Figure A.3: Impact of freezes on employment by gender (56-64 year-olds)



Notes: This figure shows the time path of the treatment effect of the freeze on employment rates for men and women separately. Workers are 56-64 years-old at the time of the freeze.

Table A.5: Pre-period summary statistics split by age group (without propensity score re-weighting)

	50-55			56-64			65-70		
Variable	Comp. mean	Diff.	p-value	Comp. mean	Diff.	p-value	Comp. mean	Diff.	p-value
				Worker characteristics					
Age	52.5	0.002	0.90	59.6	0.034	0.29	67.0	-0.011	0.68
Male	0.466	0.034	0.30	0.478	0.032	0.38	0.478	0.030	0.50
High school	0.226	0.005	0.77	0.231	0.000	0.99	0.288	0.006	0.75
Some college	0.329	-0.011	0.24	0.314	-0.010	0.26	0.292	-0.007	0.42
College or more	0.391	0.002	0.93	0.394	0.007	0.80	0.315	0.004	0.87
White	0.793	0.020	0.28	0.819	0.027	0.16	0.824	0.042	0.05
Black	0.093	-0.009	0.24	0.081	-0.013	0.05	0.074	-0.016	0.05
Hispanic	0.065	0.001	0.95	0.052	0.000	0.99	0.051	-0.006	0.46
Other race	0.049	-0.013	0.11	0.048	-0.014	0.09	0.051	-0.020	0.06
Earnings ($)	61,610	2051	0.61	62,240	1,243	0.73	48,130	-3915	0.10
Tenure at $l-5$	7.6	-0.817	0.07	8.2	-0.898	0.04	8.2	-0.760	0.07
Retired	0.020	0.002	0.17	0.054	0.004	0.23	0.188	0.008	0.50
In labor force	0.966	-0.003	0.16	0.930	-0.005	0.23	0.783	-0.010	0.49
Switched $l-5$ employer	0.034	0.025	0.01	0.030	0.024	0.02	0.027	0.006	0.07
				Pension and firm characteristics					
Log DB pension wealth/active participant	10.18	-0.037	0.91	10.27	-0.077	0.82	10.5	-0.028	0.87
Log DB pension accrual/active participant	7.73	-0.205	0.52	7.80	-0.264	0.41	7.98	-0.325	0.05
Pension plan claim age	62.7	-0.015	0.96	62.7	0.036	0.89	64.2	0.137	0.41
Log firm size	8.46	0.066	0.90	8.32	0.079	0.88	7.42	-0.273	0.55
Fraction workforce ≤ 45	0.584	0.007	0.63	0.570	0.005	0.74	0.553	-0.015	0.27
Fraction workforce [46,50]	0.146	-0.003	0.50	0.142	-0.001	0.68	0.136	0.000	0.99
Fraction workforce [51,55]	0.124	-0.002	0.60	0.128	-0.002	0.69	0.122	0.002	0.64
Fraction workforce [56,60]	0.086	-0.001	0.89	0.094	0.000	0.94	0.095	0.005	0.21
Fraction workforce [61,65]	0.041	0.000	0.98	0.045	0.000	0.96	0.061	0.005	0.20
Fraction workforce [66,70]	0.011	-0.001	0.62	0.012	-0.001	0.66	0.021	0.001	0.55
Fraction workforce ≥ 71	0.007	-0.001	0.46	0.008	-0.001	0.59	0.012	0.002	0.38
Comparison group workers		383000			373000			77000	
Treated group workers		60000			66500			11000	
Comparison group firms		7700			8600			4600	
Treated group firms		1500			1700			900	

Notes: Unless otherwise noted, statistics reported in the table average over the five year period preceeding any freeze activity. Pension wealth per active participant is computed as the present value of the liability owed to active participants divided by the number of active participants. Tenure is understated because the LEHD does not capture the complete history of an employer-employee relationship when states enter the dataset after a given employee-employer relationship is established. P-values for the difference between treatment and control groups are obtained by regressing the statistic of interest on a indicator variable for treatment status and clustering standard errors at the firm-level.

APPENDIX B

Appendix to Chapter 2

B.1 Decomposition of recession-induced losses

This appendix explains how we decompose recession-induced wage losses into three different components.

Start with the AKM/CHK decomposition (Equation (2.3.1)) in period t

$$\log(\text{wage}_{it}) = \alpha_i + \psi_j \mathbf{1}\{i \text{ works at } j \text{ in } t\} + \boldsymbol{x}'_{it}\beta + r_{it}.$$

Taking averages on both sides of Equation (2.3.1) in period t:

$$E\left[\log(\text{wage}_{it})\right] = E\left[\alpha_i\right] + E\left[\psi_j\right] + E\left[\boldsymbol{x}'_{it}\beta\right]. \tag{B.1}$$

Define the average wage in period t if the same group of individuals were to experience a 1 point increase in the unemployment rate at entry as

$$E\left[\log(\text{wage}^R_{it})\right] = E\left[\alpha^R_i\right] + E\left[\psi^R_j\right] + E\left[\boldsymbol{x}'_{it}\beta\right]. \tag{B.2}$$

In Equation (B.2), α^R_i and ψ^R_j represent estimated person and establishment fixed effects estimated for the same underlying individuals if they entered the labor market with a 1 point higher unemployment rate. Notably, person fixed effects in the AKM/CHK framework are subject to scarring and will be lower for otherwise identical individuals who, for exogenous reasons, have lower lifetime earnings.

Now, we can write $E\left[\log(\text{wage}_{it})\right] - E\left[\log(\text{wage}_{it}^{R})\right]$ as:

$$\underbrace{\beta_t^{\text{Wage}}}_{\text{\% wage differential}} = \underbrace{E\left[\alpha_i\right] - E\left[\alpha_i^R\right]}_{\text{\% due to non-employer factors}} + \underbrace{E\left[\psi_j\right] - E\left[\psi_j^R\right]}_{\text{\% due to employer-related factors}}. \tag{B.3}$$

Next, define

$$\beta_t^{\text{Non-employer}} = E\left[\alpha_i\right] - E\left[\alpha_i^R\right], \tag{B.4}$$

$$\beta_t^{\text{Employer}} = E\left[\psi_j\right] - E\left[\psi_j^R\right], \tag{B.5}$$

as the non-employer-specific and employer-specific components of recession-induced wage differentials. Estimating our main specification (Equation (2.4.1)) using CHK establishment fixed effects (ψ_j) on the left-hand-side yields $\left\{\hat{\beta}_0^{\text{Employer}}, \ldots, \hat{\beta}_{10}^{\text{Employer}}\right\}$.

Having defined recession-induced employer-specific losses, we now partition these losses into rents and compensating differentials using the decomposition in Sorkin (2018). Equation (2.3.3) splits establishment fixed effects into rents, which are explained by value, and compensation differentials which are orthogonal to value. Taking expectations on both sides, we can write:

$$E\left[\psi_j\right] = \pi E\left[V_j\right] - E\left[a_j\right]. \tag{B.6}$$

Establishment fixed effects for otherwise identical individuals who enter the labor market when unemployment rates are 1 point higher can be written as

$$E\left[\psi_j^R\right] = \pi E\left[V_j^R\right] - E\left[a_j^R\right]. \tag{B.7}$$

Subtracting (B.7) from (B.6), the employer-specific pay reduction is

$$\beta_t^{\text{Employer}} = \underbrace{\pi\left(E\left[V_j\right] - E\left[V_j^R\right]\right)}_{\text{\% due to rents}} - \underbrace{\left(E\left[a_j\right] - E\left[a_j^R\right]\right)}_{\text{\% due to amenities}}. \tag{B.8}$$

Next, define

$$\beta_t^{\text{Rent}} = \pi\left(E\left[V_j\right] - E\left[V_j^R\right]\right), \tag{B.9}$$

$$\beta_t^{\text{Amenity}} = E\left[a_j\right] - E\left[a_j^R\right], \tag{B.10}$$

Combining Equations (B.3) and (B.8) we can write

$$\begin{aligned}\beta_t^{\text{Wage}} &= \beta_t^{\text{Non-employer}} + \beta_t^{\text{Employer}} \\ &= \beta_t^{\text{Non-employer}} + \left(\beta_t^{\text{Rent}} - \beta_t^{\text{Amenity}}\right)\end{aligned} \tag{B.11}$$

Because β_t^{Wage} ,β_t^{Rent}, and β_t^{Amenity} are estimated for $t \in \{0, \ldots, 10\}$, we can recover $\beta_t^{\text{Non-employer}}$ using Equation (B.11).

Finally, define the present value of wages for the first decade of labor market experience as

$$PV = \bar{w}_0 + \bar{w}_1(1+r)^{-1} + \cdots + \bar{w}_{10}(1+r)^{-10}, \tag{B.12}$$

where $\bar{w}_t$ represents the average daily wage earned in year t. The PDV of wages for workers who face a 1 point increase in the unemployment rate at entry is

$$PV^R = \bar{w}_0(1-\beta_0^{\text{Wage}}) + \bar{w}_1(1-\beta_1^{\text{Wage}})(1+r)^{-1} + \cdots + \bar{w}_{10}(1-\beta_{10}^{\text{Wage}})(1+r)^{-10} \tag{B.13}$$

We use our estimates $\left\{\hat{\beta}_0^{\text{Wage}}, \ldots, \hat{\beta}_{10}^{\text{Wage}}\right\}$ to quantify the loss in the present value of wages attributable to a 1 point change in the unemployment rate. We then scale the resulting estimate by a one standard deviation increase in the unemployment rate which reflects a typical recession. Similar calculations with $\left\{\hat{\beta}_0^{\text{Employer}}, \ldots, \hat{\beta}_{10}^{\text{Employer}}\right\}$, $\left\{\hat{\beta}_0^{\text{Non-employer}}, \ldots, \hat{\beta}_{10}^{\text{Non-employer}}\right\}$, $\left\{\hat{\beta}_0^{\text{Rent}}, \ldots, \hat{\beta}_{10}^{\text{Rent}}\right\}$, $\left\{\hat{\beta}_0^{\text{Amenity}}, \ldots, \hat{\beta}_{10}^{\text{Amenity}}\right\}$ yield estimates of the loss in the PV of wages attributable to employer-specific factors, non-employer specific factors, rents, and amenities.

B.2 Accounting for measurement error in establishment values

When estimating equation (2.3.3), we find that almost 96 percent of the variation in establishment fixed effects is in the residual, which is substantially larger than the 72 percent estimate obtained by Sorkin using U.S. data. The key reason for this difference is that our establishment value estimates are derived from a 2 percent sample of workers — and therefore more likely to be affected by measurement error — while Sorkin's estimates are obtained using the full population of workers in several U.S. states.

To evaluate the quantitative implications of measurement error in our estimates, we randomly partition individuals in the SIAB into two groups and re-estimate the establishment values using moves within each partition separately. For each establishment, this exercise yields two different value estimates, each of which is based on an independent sample of worker moves. Denote the error-free value of establishment j by V_j^*; the split sample estimates V_j^1 and V_j^2 are given by

$$V_j^1 = V_j^* + u_j^1 \tag{B.1}$$
$$V_j^2 = V_j^* + u_j^2. \tag{B.2}$$

Because independent samples of worker moves are used to estimate V_j^1 and V_j^2, we assume that

$$E\left[u_j^1 u_j^2\right] = 0. \tag{B.3}$$

We then re-estimate equation (2.3.3) using an instrumental variables (IV) approach by employing V_j^1 as an instrument for V_j^2 and vice-versa. Assuming that each value estimate exhibits classical measurement error (i.e. that $E\left[V_j^* u_j^1\right] = E\left[V_j^* u_j^2\right] = 0$), one would expect OLS to yield estimates biased toward zero relative to IV.

Table B.1 presents estimates of π based on the pooled set of workers, as well as estimates based on sample of establishments for which we estimate two sets of value estimates based on independent sets of worker moves. Using OLS to estimate π on the pooled sample yields a coefficient of 0.212. Estimating π using OLS in the smaller sample of establishments for which we obtain two independent estimates of value, yields coefficients of 0.239 and 0.220 which are similar in magnitude to the pooled sample estimate. In contrast, the IV-based slope estimates are approximately two times larger than those obtained using OLS, a result which is consistent with classical measurement error induced attenuation.[1]

[1]The two IV-based estimates reverse the order of the instrumental and the instrumented variable which is arbitrary. Randomly partitioning the pooled sample reduces the number of workers available to estimate

We also note that our interpretation of Figure 2.6 does not change in the presence of measurement error. While the average relationship plotted in the dashed line would tilt upwards when measurement error is corrected for, so would the implied career paths, by roughly the same amount. This is the visual manifestation of the logic that recessions do not differentially send workers to firms with more or less measurement error in their estimated values.

Table B.**1:** Classical measurement error-induced attenuation in π

	Pooled Sample	Split Sample			
	OLS	OLS(1)	OLS(2)	IV(1)	IV(2)
$\hat{\pi}$	0.212	0.239	0.220	0.397	0.432
	(0.002)	(0.005)	(0.005)	(0.009)	(0.009)
R^2	0.04	0.06	0.05	0.03	0.004
N	171,640	59,762	59,762	59,762	59,762

Notes: All models include gender fixed effects. OLS(1) and OLS(2) use the value estimates obtained using each sample split. IV(1) and IV(2) reverse the order of the instrumental and instrumented variable which is arbitrary.

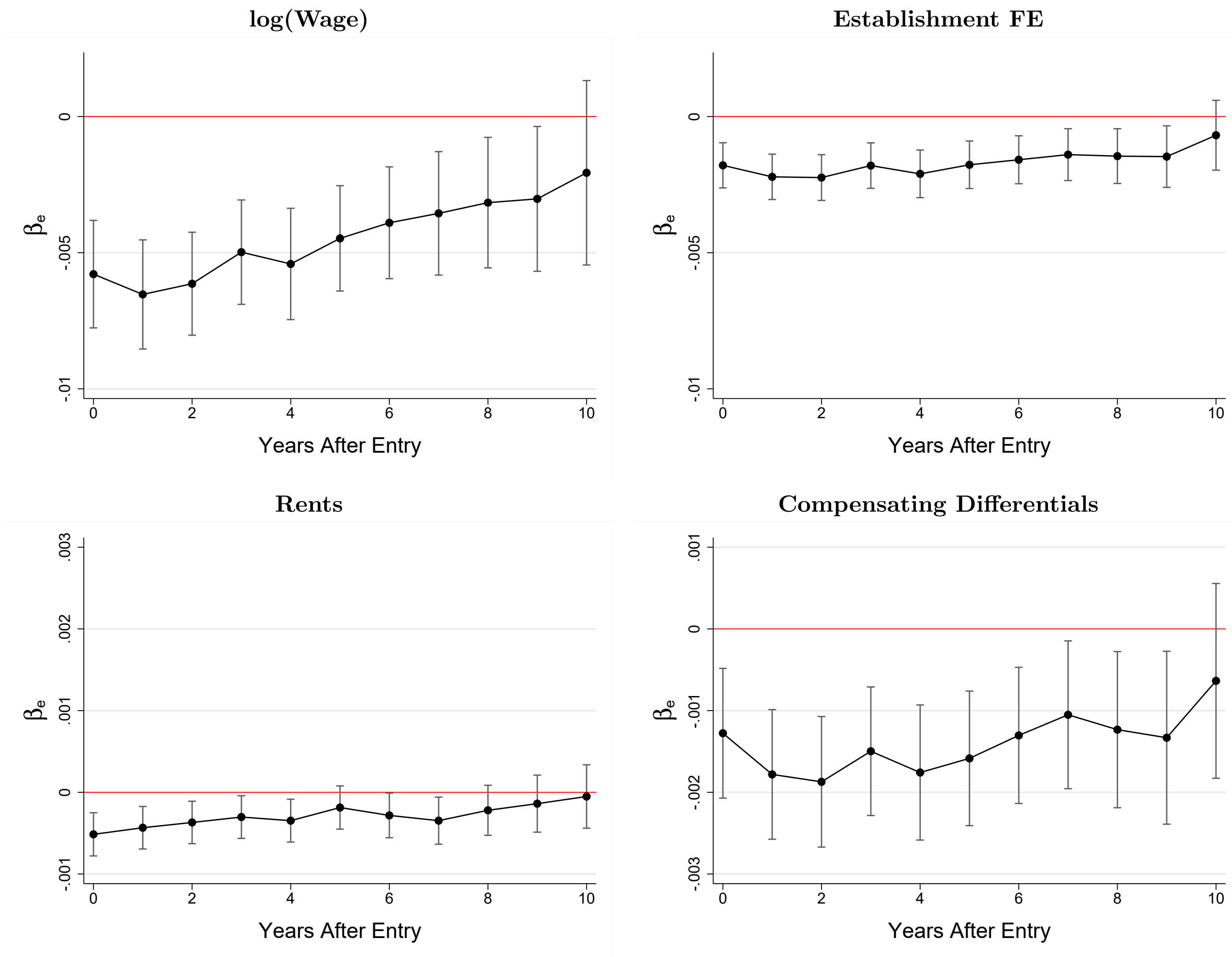
values for any given establishment, which disproportionately eliminates smaller establishments since they may no longer have sufficient flows to be in the strongly connected set. The slightly larger OLS-based estimates in the split sample versus the pooled sample are a by-product of this selectivity.

B.3 Additional analyses

B.3.1 Using actual training duration

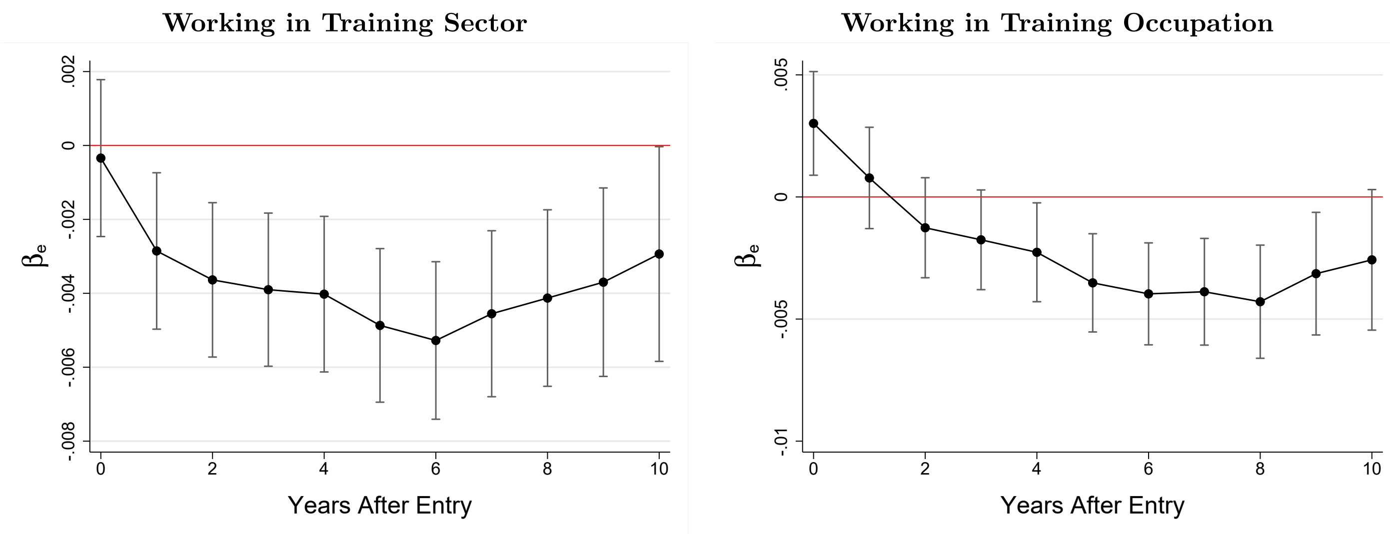
This appendix shows results obtained by re-estimating Equation (2.4.1) using actual rather than predicted year of entry. Actual year of entry is defined by the last day of vocational training.

Figure B.1: The effect of entry conditions (U_{osc}) on early career outcomes using actual year of entry



Notes: Sample size for each specification is 125,363. 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, predicted year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German indicator variable, and a female indicator variable.

Figure B.2: The effect of entry conditions (U_{osc}) on early career mobility using actual year of entry



Notes: Sample size for each specification is 125,363. 95% confidence intervals represented by dashed lines. U_{osc} is the unemployment rate in a given individual's training occupation and training state in the individual's year of entry. Controls included are potential experience fixed effects, year fixed effects, predicted year of entry fixed effects, training occupation fixed effects, state of training fixed effects, median wage paid by training firm during last year of training, age at start of training fixed effects, a German indicator variable, and a female indicator variable.

APPENDIX C

Appendix to Chapter 3

C.1 Constructing the training data set

Our training sample is composed of HRS-BR pairs generated by blocking the 1998 and 2004 waves of the HRS with the BR on 3-digit zip code, 10-digit phone number, telephone area code, and city-state. We choose these specific years for two reasons. First, 1998 and 2004 were years in which the HRS drew fresh cohorts of survey respondents. Second, the file structure of the BR changed in substantive ways in 2002. As such, using HRS cohorts before and after 2002 to train the model allows us to account for unobserved variation in the quality of data drawn from the BR.

Simple random sampling of pairs for human review would produce very few true matches thereby limiting the predictive ability of our model. Instead, we compute Jaro-Winkler (JW) scores for name and address similarity for each pair. We then divide the JW scores for name and address into 4 bins each, with grid points spaced closer together at the right tail of the score distribution. This binning exercise defines 16 strata from which we draw equally sized samples to obtain a total sample size of $N^T \approx 1000$ pairs. Because the bins are concentrated at the top of the JW name and address score distribution, this stratified sampling methodology substantially increases the share of true matches in the training data set relative to a simple random sample (see, e.g., Christen (2012)).

C.2 Model selection

Given that our training data set consists of approximately 2000 observations, estimating a model with a high dimension of predictors is likely to yield unstable parameter estimates.

To solve this dimensionality problem and, more importantly, to avoid over-fitting our model, we use ML tools to aid in prediction. While a complex model with many variables and interactions has the potential of reducing in-sample (training) errors substantially, this improvement is misleading because it considers the wrong model-fit criterion. To ensure that the model generalizes well, we consider out-of-sample (test) error which we estimate using 10-fold cross validation.

In our setting, the complexity of the prediction model is indexed by the dimension of the covariate vector. Reducing model complexity by shrinking the number of covariates increases the bias component of the test error, but has the potential to reduce the variance component substantially. In order to obtain a model with the optimal degree of complexity, we employ the Elastic Net (EN) shrinkage estimator. The EN estimator is the solution to the minimization problem posed in (C.1); l indexes observations in the training set, while p indexes regressors in the model:

$$\begin{gathered}\min_{\beta \in \mathbb{R}^p} \sum_{l=1}^{2N^T} \left(y_l - \Lambda \left(\beta_0 + \sum_{p=1}^{q} \tilde{x}_{lp} \beta_p \right) \right)^2 \\ \text{st: } \sum_{p=1}^{q} \beta_p^2 \leq t_1, \ \sum_{p=1}^{q} |\beta_p| \leq t_2 \end{gathered} \tag{C.1}$$

In (C.1), the typical least squares minimization criteria is supplemented with two constraints each of which constitutes a tuning parameter for the estimator. Together, these tuning parameters control the level of model complexity: t_1, as in Ridge Regression, sets a maximum threshold on the sum of squared values of the coefficients. The Ridge penalty term has the effect of controlling the variance component of test error by preventing any one predictor from exhibiting too strong of an effect on the outcome. This penalty is particularly important when some predictors are correlated. t_2, as in the LASSO, sets a maximum threshold on the sum of the absolute values of the coefficients. When this second constraint binds, some of the coefficients are set exactly to zero thereby shrinking the dimensionality of the model. The optimal prediction model is chosen by finding the pair of tuning parameters that jointly minimize out-of-sample error.

The EN estimator can also be conveniently summarized in Lagrangian form as shown in equation (C.2). The two tuning parameters discussed above are replaced by a Lagrange multiplier, $\lambda \in \mathbb{R}_+$, and a parameter $\alpha \in [0, 1]$ that controls the degree of mixing between the Ridge constraint and the LASSO constraint:

$$\min_{\beta \in \mathbb{R}^p} \sum_{l=1}^{2N^T} \left(y_l - \Lambda \left(\beta_0 + \sum_{p=1}^{q} \tilde{x}_{lp} \beta_p \right) \right)^2 + \lambda \sum_{p=1}^{q} \left(\alpha |\beta_p| + (1-\alpha) \beta_p^2 \right) \tag{C.2}$$

We obtain our establishment and employer prediction models by implementing the EN estimator using the `lassoglm` package in MATLAB. This particular implementation of the EN estimator takes a given value of α and finds the value of λ that delivers the lowest out-of-sample deviance. To obtain the best prediction model, we perform a grid search by iterating α from 0.05 to 0.95 in 0.05-unit increments. For each value of α, we obtain the model associated with the lowest test deviance estimate. The optimal model is the one with the lowest test deviance across all the values of α.

C.3 Sample selection induced by cutoffs

As noted in section 3.4.2.1 the data-driven cutoffs that we estimate and impose on the set of pairs prior to drawing multiple implicates sometimes results in 0 BR candidate matches for a given HRS job. For about 33 percent of employer matches and 8 percent of establishment matches where such a situation occurs, we have little confidence about having selected the right employer or establishment using our matching models.[1] This appendix examines the extent to which HRS respondents for whom we find at least one BR match differ from HRS respondents for whom we cannot find any suitable matches.

Table C.1 shows demographics, education, wages, annual hours of work, union membership, tenure, total labor market experience, and whether respondents work for public sector employers. The first column of the table shows characteristics for the full sample. The left panel shows characteristics for respondents matched to at least one employer (above cutoff) and respondents for whom no adequate BR candidate employer is available (below cutoff). The right panel shows the same information for establishment matches. Concentrating first on the left panel reveals some key differences: Nonwhite and foreign-born respondents are more likely to be unmatched. Respondents that are unmatched have about two years less in tenure with their current employer and have slightly lower lifetime labor market experience. Finally, unmatched workers are substantially less likely to be employed in the public sector. These differences point at employer attachment as an important source

[1] While employer matches are easier to confirm than establishment matches, the ROC-based optimal cutoffs trade off sensitivity and specificity. Furthermore, the employer and establishment match models are estimated independently. Thus, the share of HRS jobs with inadequate BR employer candidates (33 percent) is larger than the share of HRS jobs with inadequate BR establishment candidates (8 percent).

of signal strength in the data on employer names and addresses (obtained primarily for pension characteristics in the HRS). Furthermore, because public employers are more likely to maintain unified pension plans, it is easier to obtain sharper matching of public sector employees using our method.[2]

Patterns of selection in the right hand panel are similar to those highlighted above: Respondents who are nonwhite, foreign-born, have lower tenure, less labor market experience, and are more likely to be employed outside the public sector are less likely to be matched to any establishment. In addition, union membership is predictive of higher quality matches. On the whole, these statistics show that information elicited from survey responses may be garbled in ways that are correlated with individual characteristics. While our cutoff-based procedure filters away these sources of noise and hones in on higher quality matches for the majority of the sample, it does not refine information that is already garbled. Addressing this concern ultimately requires reliance on sharper identifiers such as EINs.

[2]Private employers often offer multiple pension plans with a variety of names. While reporting information about their pension plans, respondents may provide pension plan names that differ in small but meaningful ways from the employers name as it would appear in the BR. This source of variability could reduce the accuracy of our matching algorithm.

Table C.1: Characteristics of matchable and non-matchable respondents

		Employer match		Establishment match	
Variable	Full sample	Above cutoff	Below cutoff	Above cutoff	Below cutoff
N	5700	3700	1900	5200	450
Male	0.446	0.432	0.472	0.439	0.531
Age	56.8	56.7	57.0	56.7	58.2
Native born	0.837	0.865	0.782	0.843	0.758
White	0.643	0.684	0.565	0.647	0.596
Black	0.236	0.210	0.287	0.235	0.249
Other race	0.121	0.107	0.148	0.118	0.155
Schooling (years)	13.3	13.5	13.0	13.4	12.5
Wage ($/hr)	29.5	28.9	30.6	29.0	34.8
Hours	1902	1901	1905	1902	1906
Union	0.143	0.139	0.151	0.147	0.096
Tenure (years)	10.2	10.7	9.2	10.3	7.5
Experience (years)	32.5	32.9	31.9	32.6	30.2
Public employer	0.247	0.280	0.184	0.258	0.108

Notes: This table is based on the set of working HRS respondents in the 2010 wave who provided names and addresses of their employers. Some HRS respondents report hourly wages directly. Others report compensation at daily, weekly, monthly, or annual levels. When compensation is reported at a different level than hourly, we convert it using the respondent's report of how many hours per week and weeks per year worked. Public sector employment is coded by the HRS based on the report of the employer name elicited from the respondent.

C.4 Measurement error and nonresponse bias

Define $R_i = 1$ if the HRS respondent reports a value for firm size and $R_i = 0$ if they do not. Denote the decile of the log hourly wage of respondent i by d_i. The expectation of true log firm size conditional on the decile of log wages is

$$E\left[s_{ij}^*|d_i\right] \tag{C.1}$$

The log of self-reported firm size conditional on the decile of log wages is

$$E\left[s_{ij}|d_i,\, R_i = 1\right] \tag{C.2}$$

The log of true firm size conditional on the decile of log wages among those who do respond is

$$E\left[s_{ij}^*|d_i,\, R_i = 1\right] \tag{C.3}$$

Measurement error is given by subtracting (C.3) from (C.2):

$$\underbrace{E\left[v_i|d_i,\, R_i = 1\right]}_{\text{Measurement error}} = E\left[s_{ij}|d_i,\, R_i = 1\right] - E\left[s_{ij}^*|d_i,\, R_i = 1\right] \tag{C.4}$$

Decompose term (C.1) by writing

$$E\left[s_{ij}^*|d_i\right] = p^d E\left[s_{ij}^*|d_i, R_i = 1\right] + (1 - p^d) E\left[s_{ij}^*|d_i, R_i = 0\right] \tag{C.5}$$

where $p^d = P\left[R_i = 1|d_i\right]$ is the conditional response probability. Subtracting $E\left[s_{ij}^*|d_i, R_i = 1\right]$ from both sides of equation (C.5) yields the following expression for bias due to nonresponse

$$\underbrace{E\left[s_{ij}^*|d_i\right] - E\left[s_{ij}^*|d_i, R_i = 1\right]}_{\text{Nonresponse bias}} = (1 - p^d)\left\{E\left[s_{ij}^*|d_i, R_i = 0\right] - E\left[s_{ij}^*|d_i, R_i = 1\right]\right\} \tag{C.6}$$

Positive values of the left side of equation (C.6) imply that non-responders work at larger employers than do responders since $1 - p^d \in (0, 1)$. The converse is true for negative values of the left side.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abowd, J. M. and O. C. Ashenfelter (1981): "Anticipated Unemployment, Temporary Layoffs, and Compensating Wage Differentials," in *Studies in Labor Markets*, National Bureau of Economic Research, Inc, NBER Chapters, 141–170.

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67, 251–333.

Abowd, J. M., B. E. Stephens, L. Vilhuber, F. Anderson, K. L. McKinney, M. Roemer, and S. Woodcock (2009): "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators," in *Producer Dynamics: New Evidence from Micro Data*, ed. by T. Dunne, J. B. Jensen, and M. J. Roberts, University of Chicago Press, 149–230.

Abowd, J. M. and M. H. Stinson (2013): "Estimating Measurement Error in Annual Job Earnings: A Comparison of Survey and Administrative Data," *Review of Economics and Statistics*, 95, 1451–1467.

Akerlof, G. and L. F. Katz (1989): "Workers' Trust Funds and the Logic of Wage Profiles," *Quarterly Journal of Economics*, 104, 525–536.

Altonji, J. G., L. B. Kahn, and J. D. Speer (2016): "Cashier or Consultant? Entry Labor Market Conditions, Field of Study, and Career Success," *Journal of Labor Economics*, 34, 361–401.

Ameriks, J., J. S. Briggs, A. Caplin, M. Lee, M. D. Shapiro, and C. Tonetti (2020): "Older Americans Would Work Longer If Jobs Were Flexible," *American Economic Journal: Macroeconomics*, 12, 174–209.

Azariadis, C. (1975): "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, 83, 1183–1202.

Bailey, M., C. Cole, M. Henderson, and C. Massey (2017): "How Well Do Automated Methods Perform in Historical Samples? Evidence From New Ground Truth," NBER Working Paper No. 24019.

Beaudry, P. and J. DiNardo (1991): "The Effect of Implicit Contracts on the Movement of Wages Over the Business Cycle: Evidence from Micro Data," *Journal of Political Economy*, 99, 665–688.

Belin, T. R. and D. B. Rubin (1995): "A Method for Calibrating False-Match Rates in Record Linkage," *Journal of the American Statistical Association*, 90, 694–707.

Berkovic, J. and S. Stern (1991): "Job Exit Behavior of Older Men," *Econometrica*, 59, 189–210.

Bloom, N., F. Guvenen, B. S. Smith, J. Song, and T. von Wachter (2018): "Inequality and the Disappearing Large Firm Wage Premium," *American Economic Association Papers and Proceedings*, 108, 317–322.

Bovbjerg, B. D., C. A. Jeszeck, C. Ford, I. Johnson, L. Moy, M. Ramage, J. Applebaum, C. Winslow, G. Kuehneman, B. Friedman, M. Swearingen, M. Mayfield, S. Bernstein, and W. Vance (2008): "Defined Benefit Pensions: Freezes Affect Millions of Participants and May Pose Retirement Income Challenges," Tech. rep., United States Government Accountability Office.

Brown, C. (1980): "Equalizing Differences in the Labor Market," *Quarterly Journal of Economics*, 94, 113–134.

Brown, C. and J. Medoff (1989): "The Employer Size-Wage Effect," *Journal of Political Economy*, 97, 1027–1059.

Brown, K. M. (2013): "The Link Between Pensions and Retirement Timing: Lessons from California Teachers," *Journal of Public Economics*, 98, 1–14.

Burgette, L. F. and J. P. Reiter (2010): "Multiple Imputation for Missing Data via Sequential Regression Trees," *American Journal of Epidemiology*, 172, 1070–1076.

Burtless, G. and J. P. Quinn (2002): "Is Working Longer the Answer for an Aging Workforce?" Tech. rep., Center for Retirement Research at Boston College.

Busso, M., J. DiNardo, and J. McCrary (2014): "New Evidence on the Finite Sample Properties of Propensity Score Reweighting and Matching Estimators," *Review of Economics and Statistics*, 96, 885–897.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *Quarterly Journal of Economics*, 128, 967–1015.

Christen, P. (2008a): "Automatic Record Linkage Using Seeded Nearest Neighbour and Support Vector Machine Classification," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Association for Computing Machinery, 151–159.

——— (2008b): "Automatic Training Example Selection for Scalable Unsupervised Record Linkage," in *Proceedings of the 12th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining*, Springer-Verlag, 511–518.

——— (2012): *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*, Springer-Verlag.

Cochinwala, M., V. Kurien, G. Lalk, and D. Shasha (2001): "Efficient Data Reconciliation," *Information Sciences*, 137, 1–15.

Coffin, M. and S. Sukhatme (1997): "Receiver Operating Characteristic Studies and Measurement Errors," *Biometrics*, 53, 823–837.

Coile, C. and J. Gruber (2007): "Future Social Security Entitlements and the Retirement Decision," *Review of Economics and Statistics*, 89, 234–246.

Davis, S. J. and T. von Wachter (2011): "Recessions and the Costs of Job Loss," *Brookings Papers on Economic Activity*, 1–72.

Decressin, A., T. Hill, K. McCue, and M. Stinson (2009): "The Role of Fringe Benefits in Employer and Workforce Dynamics," in *Producer Dynamics: New Evidence from Micro Data*, ed. by T. Dunne, J. B. Jensen, and M. J. Roberts, University of Chicago Press, 473–505.

Duffie, D. and K. J. Singleton (1993): "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–952.

Elfeky, M. G., V. S. Verykios, and A. K. Elmagarmid (2002): "TAILOR: A Record Linkage Toolbox," in *Proceedings 18th International Conference on Data Engineering*, Institute of Electrical and Electronics Engineers, 17–28.

Fang, C., C. Brown, and D. Weir (2016): "Cohort Changes in Social Security Benefits and Pension Wealth," Michigan Retirement Research Center Working Paper WP 2016-350.

Fellegi, I. P. and A. B. Sunter (1969): "A Theory for Record Linkage," *Journal of the American Statistical Association*, 64, 1183–1210.

Ferrie, J. P. (1996): "A New Sample of Males Linked from the 1850 Public Use Micro Sample of the Federal Census of Population to the 1860 Federal Census Manuscript Schedules," *Historical Methods*, 29, 141–156.

Fortini, M., B. Liseo, A. Nuccitelli, and M. Scanu (2001): "On Bayesian Record Linkage," *Research in Official Statistics*, 4, 185–198.

Freeman, R. B. (1978): "Job Satisfaction as an Economic Variable," *American Economic Review*, 68, 135–141.

French, E. (2005): "The Effects of Health, Wealth, and Wages on Labour Supply and Retirement Behaviour," *Review of Economic Studies*, 72, 395–427.

French, E. and J. B. Jones (2011): "The Effects of Health Insurance and Self-Insurance on Retirement Behavior," *Econometrica*, 79, 673–732.

Friedberg, L. and A. Webb (2005): "Retirement and the Evolution of Pension Structure," *Journal of Human Resources*, 40, 281–308.

Gelber, A. M., A. Isen, and J. Song (2016): "The Effect of Pension Income on Elderly Earnings: Evidence from Social Security Full Population Data," Unpublished working paper.

Gelber, A. M., D. Jones, D. W. Sacks, and J. Song (2017): "Using Non-Linear Budget Sets to Estimate Extensive Margin Responses: Method and Evidence from the Social Security Earnings Test," NBER Working Paper No. 23362.

Goda, G. S., J. B. Shoven, and S. N. Slavov (2009): "Removing the Disincentives in Social Security for Long Careers," in *Social Security Policy in a Changing Environment*, ed. by J. Brown, J. Liebman, and D. Wise, University of Chicago Press, 21–38.

Goldstein, H., K. Harron, and A. Wade (2012): "The Analysis of Record-Linked Data Using Multiple Imputation with Data Value Priors," *Statistics in Medicine*, 31, 3481–3493.

Gustman, A., O. Mitchell, A. Samwick, and T. L. Steinmeier (2000): "Evaluating Pension Entitlements," in *Forecasting Retirement Needs and Retirement Wealth*, ed. by O. Mitchell, P. B. Hammond, and A. M. Rappaport, University of Pennsylvania Press, 310–326.

Gustman, A. L. and T. L. Steinmeier (2015): "Effects of Social Security Policies on Benefit Claiming, Retirement, and Saving," *Journal of Public Economics*, 129, 412–437.

Gutman, R., C. C. Afendulis, and A. M. Zaslavsky (2013): "A Bayesian Procedure for File Linking to Analyze End-of-Life Medical Costs," *Journal of the American Statistical Association*, 108, 34–47.

Gutman, R., C. J. Sammartino, T. C. Green, and B. T. Montague (2014): "Error Adjustments for File Linking Methods Using Encrypted Unique Client Identifier (eUCI) with Application to Recently Released HIV+ Prisoners," *Statistics in Medicine*, 35, 115–129.

Haltiwanger, J., H. Hyatt, E. McEntarfer, L. Sousa, and S. Tibbets (2014): "Firm Age and Size in the Longitudinal Employer-Household Dynamics Data," Tech. rep., Center for Economic Studies Working Paper 14-16.

Hamman, M. K., D. Hochfellner, J. M. Nunley, and C. Ruhm (2016): "Peer Effects and Retirement Decisions: Evidence from Pension Reform in Germany," Unpublished working paper.

Hammill, B. G., A. F. Hernandez, E. D. Peterson, G. C. Fonarow, K. A. Schulman, and L. H. Curtis (2009): "Linking Inpatient Clinical Registry Data to Medicare Claims Data Using Indirect Identifiers," *American Heart Journal*, 157, 995–1000.

Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993): "Earnings Losses of Displaced Workers," *American Economic Review*, 83, 685–709.

Jarmin, R. S. and J. Miranda (2002): "The Longitudinal Business Database," Center for Economic Studies, U.S. Census Bureau, Working Paper 02-17.

Jarosch, G. (2015): "Searching for Job Security and the Consequences of Job Loss," Unpublished working paper.

Kahn, L. (2010): "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy," *Labour Economics*, 17, 303–316.

Kotlikoff, L. J. and D. A. Wise (1985): "Labor Compensation and the Structure of Private Pension Plans: Evidence for Contractual versus Spot Labor Markets," in *Pensions, Labor, and Individual Choice*, ed. by D. A. Wise, University of Chicago Press, 55–88.

Krolikowski, P. (2017): "Job Ladders and Earnings of Displaced Workers," *American Economic Journal: Macroeconomics*, 9, 1–31.

Krueger, A. B. and J.-S. Pischke (1992): "The Effect of Social Security on Labor Supply: A Cohort Analysis of the Notch Generation," *Journal of Labor Economics*, 10, 412–437.

Kuppe, A. M., B. Lorig, A. Stohr, and H. Schwarz (2013): "Training Regulations and How They Come About," Tech. rep., Federal Institute for Vocational Education and Training, Bonn.

Lahiri, P. and M. D. Larsen (2005): "Regression Analysis With Linked Data," *Journal of the American Statistical Association*, 100, 222–230.

Laitner, J. and D. Silverman (2012): "Consumption, Retirement, and Social Security: Evaluating the Efficiency of Reform that Encourages Longer Careers," *Journal of Public Economics*, 96, 615–634.

Larsen, M. D. (2004): "Record Linkage Using Finite Mixture Models," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, ed. by A. Gelman and X.-L. Meng, John Wiley, 309–318.

Lawson, E. H., C. Y. Ko, R. Louie, L. Han, M. Rapp, and D. S. Zigmond (2013): "Linkage of a Clinical Surgical Registry with Medicare Inpatient Claims Data Using Indirect Identifiers," *Surgery*, 153, 423–430.

Lazear, E. P. (1979): "Why Is There Mandatory Retirement?" *Journal of Political Economy*, 87, 1261–1284.

——— (1981): "Agency, Earnings Profiles, Productivity, and Hours Restrictions," *American Economic Review*, 71, 606–620.

Lazear, E. P. and R. L. Moore (1988): "Pensions and Turnover," in *Pensions in the U.S. Economy*, ed. by Z. Bodie, J. B. Shoven, and D. A. Wise, University of Chicago Press, 163–190.

Lucas, R. E. (1977): "Hedonic Wage Equations and Psychic Wages in the Returns to Schooling," *American Economic Review*, 67, 549–558.

LUMSDAINE, R. L., J. H. STOCK, AND D. A. WISE (1992): "Three Models of Retirement: Computational Complexity versus Predictive Validity," in *Topics in the Economics of Aging*, ed. by D. A. Wise, University of Chicago Press, 21–60.

MACLEAN, J. (2013): "The Health Effects of Leaving School in a Bad Economy," *Journal of Health Economics*, 32, 951–964.

MACURDY, T. E. (1981): "An Empirical Model of Labor Supply in a Life-Cycle Setting," *Journal of Political Economy*, 89, 1059–1085.

MANOLI, D. S. AND A. WEBER (2016): "Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions," *American Economic Journal: Economic Policy*, 8, 160–182.

MCFADDEN, D. (1989): "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica*, 57, 995–1026.

MCKINNEY, K., A. GREEN, L. VILHUBER, AND J. ABOWD (2017): "Total Error and Variability Measures with Integrated Disclosure Limitation for Quarterly Workforce Indicators and LEHD Origin Destination Employment Statistics in On The Map," Center for Economic Studies, U.S. Census Bureau, Working Paper 17-71.

MCLAUGHLIN, K. J. AND M. BILS (2001): "Interindustry Mobility and the Cyclical Upgrading of Labor," *Journal of Labor Economics*, 19, 94–135.

MENG, X.-L. (1994): "Multiple-Imputation Inferences with Uncongenial Sources of Input," *Statistical Science*, 9, 566–573.

MONAHAN, A. (2012): "Statutes as Contracts - The California Rule and its Impact on Public Pension Reform," *Iowa Law Review*, 97, 963–1029.

MUNNELL, A. H. AND M. SOTO (2007): "Why are Companies Freezing their Pensions?" Working Paper 2007-22. Center for Retirement Research, Boston College.

MURRAY, J. S. (2018): "Multiple Imputation: A Review of Practical and Theoretical Findings," *Statistical Science*, 33, 142–159.

NEAL, D. (1995): "Industry-Specific Human Capital: Evidence from Displaced Workers," *Journal of Labor Economics*, 13, 653–677.

OI, W. Y. AND T. L. IDSON (1999): "Firm Size and Wages," in *Handbook of Labor Economics 3B*, ed. by O. C. Ashenfelter and D. Card, North Holland, 2165–2214.

OKUN, A. (1973): "Upward Mobility in a High-Pressure Economy," *Brookings Papers on Economic Activity*, 207–261.

OREOPOULOS, P., T. VON WACHTER, AND A. HEISZ (2012): "The Short- and Long-Term Career Effects of Graduating in a Recession," *American Economic Journal: Applied Economics*, 4.

Oyer, P. (2006): "Initial Labor Market Conditions and Long-Term Outcomes for Economists," *Journal of Economic Perspectives*, 20, 143–160.

——— (2008): "The Making of an Investment Banker: Stock Market Shocks, Career Choice, and Lifetime Income," *Journal of Finance*, 63, 2601–2628.

Rauh, J. D., I. Stefanescu, and S. P. Zeldes (2020): "Cost Saving and the Freezing of Corporate Pension Plans," NBER Working Paper No. 27251.

Reiter, J. P. (2005): "Using CART to Generate Partially Synthetic Public Use Microdata," *Journal of Official Statistics*, 21, 441–462.

Rinz, K. (2019): "Did Timing Matter? Life Cycle Differences in Effects of Exposure to the Great Recession," Working Papers 19-25, Center for Economic Studies, U.S. Census Bureau.

Rogerson, R. and J. Wallenius (2013): "Nonconvexities, Retirement, and the Elasticity of Labor Supply," *American Economic Review*, 103, 1445–1462.

Rosen, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *Journal of Political Economy*, 82, 34–55.

——— (1985): "Implicit Contracts: A Survey," *Journal of Economic Literature*, 23, 1144–1175.

——— (1986): "The Theory of Equalizing Differences," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and R. Layard, Elsevier, vol. 1, 641–692.

Rouwenhorst, K. G. (1995): "Asset Pricing Implications of Equilibrium Business Cycle Models," in *Frontiers of Business Cycle Research*, ed. by T. F. Conley, Princeton University Press, 294–330.

Rubin, D. B. (1976): "Inference and Missing Data," *Biometrika*, 63, 581–592.

——— (1987): *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley.

——— (1996): "Multiple Imputation After 18+ Years," *Journal of the American Statistical Association*, 91, 473–489.

Rust, J. and C. Phelan (1997): "How Social Security and Medicare Affect Retirement Behavior In a World of Incomplete Markets," *Econometrica*, 65, 781–831.

Schwandt, H. and T. von Wachter (2016): "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets," in *Youth Labor Markets*, National Bureau of Economic Research, Inc, NBER Chapters.

Setoguchi, S., Y. Zhu, J. J. Jalbert, L. A. Williams, and C.-Y. Chen (2014): "Validity of Deterministic Record Linkage Using Multiple Indirect Personal Identifiers," *Circulation: Cardiovascular Quality and Outcomes*, 7, 475–480.

Sorkin, I. (2018): "Ranking Firms Under Revealed Preference," *Quarterly Journal of Economics*, 133, 1331–1393.

Steorts, R. C., R. Hall, and S. E. Feinberg (2016): "A Bayesian Approach to Graphical Record Linkage and De-duplication," *Journal of the American Statistical Association*, 111, 1660–1672.

Stinson, M. (2003): "Technical Description of SIPP Job Identification Number Editing, 1990-1993 SIPP Panels," SIPP Technical Paper, U.S. Census Bureau.

Stock, J. H. and D. A. Wise (1990): "Pensions, the Option Value of Work, and Retirement," *Econometrica*, 58, 1151–1180.

Tancredi, A. and B. Liseo (2011): "A Hierarchical Bayesian Approach to Record Linkage and Population Size Problems," *Annals of Applied Statistics*, 5, 1553–1585.

vom Berge, P., M. Konig, and S. Seth (2013): "Sample of Integrated Labour Market Biographies 1975-2010 (SIAB 7510), FDZ data report," Tech. rep., Research Data Center (FDZ) of the Federal Employment Agency (BA) at the Institute for Employment Research (IAB), Nuremberg.

von Wachter, T. and S. Bender (2006): "In the Right Place at the Wrong Time - The Role of Firms and Luck in Young Workers' Careers," *American Economic Review*, 96, 1679–1705.

Warren, J. L., C. N. Klabunde, D. Schrag, P. B. Bach, and G. F. Riley (2002): "Overview of the SEER-Medicare Data: Content, Research Applications, and Generalizability to the United States Elderly Population," *Medical Care*, 40, IV3–IV18.

Youden, W. J. (1950): "Index for Rating Diagnostic Tests," *Cancer*, 3, 32–35.

Zelinksy, E. A. (2000): "The Cash Balance Controversy," *Virginia Tax Review*, 19, 683–762.

Zou, H. and T. Hastie (2005): "Regularization and Variable Selection via the Elastic Net," *Journal of the Royal Statistical Society Series B*, 67, 301–320.