Title
Essays on Unemployment Insurance and Risky Behavior

Author
Schnorr, Geoffrey

Publication Date
2021

Peer reviewed|Thesis/dissertation

Essays on Unemployment Insurance and Risky Behavior

By

GEOFFREY C. SCHNORR

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Economics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

____

Marianne Bitler, Chair

____

Monica Singhal

____

Brendan Price

____

Marianne Page

Committee in Charge

2021

Copyright notice for Chapter 2:

Acknowledgements

I am indebted to my committee for their advice, support, and generosity. I thank Marianne Bitler for guiding me through this process and for pushing me to prioritize my interests; Monica Singhal, for keeping me motivated and for her persistent focus on making the work better; Brendan Price, for his attention to detail and for helping me to understand my comparative advantages; and Marianne Page, for her positivity and for many conversations and opportunities which helped me transition from coursework to research.

For helpful comments which improved the work in this dissertation, I thank George Bulman, Kitt Carpenter, Scott Carrell, Jonathan Cohen, Chloe East, Emiliano Huet-Vaughn, Krzysztof Karbownik, and Guangli Zhang. For assistance in accessing and understanding the data used in Chapter 1, I thank Muhammad Akhtar, Cathy Barratt, Juan Barrios, Amy Faulkner, Erika Knutson, Jose Sanchez, and Marc Stockton. I also want to thank Till von Wachter and Janey Rountree for helping to facilitate relationships that made Chapter 1 possible. To Eunju Lee, Lester Lusher, and Rebecca Taylor: Thank you for showing me the importance of great coauthors.

I could not be happier with my decision to attend UC Davis for graduate school. I tell everyone who will listen (and some who won't) that this is a special place. To all of the students, staff, and faculty in the Economics Department, thank you for making this community what it is. Special thanks to Briana Ballis, Annie Hines, Konstantin Kunze, Marina Lovchikova, Lester Lusher, Derek

Rury, and Justin Wiltshire for their friendship.

My path to a PhD was long and atypical. Peter Bach and Partha Deb were critical in helping me to find, and start down, that path. I am very grateful to both of them.

Finally, I'd like to thank my family. Mom and Dad, for your love and life lessons which have helped me so much; Patty and Brian, for your love and kindness; Zach, for being a great brother (and a decent barkeeper); and Annie, for being the best companion a human could ask for.

Most importantly, I'd like to thank Alyssa. Without your love and support none of this would have been possible. Thank you for taking this leap with me.

Abstract

This dissertation contributes to literatures on the design and effects of two important public policies: Unemployment insurance and the minimum legal drinking age. I provide two unique perspectives on the design of unemployment insurance programs. First, by studying an oft-ignored policy parameter which influences benefit generosity, and second by analyzing an understudied margin along which these programs affect work incentives. I extend the literature on the minimum legal drinking age by empirically testing the hypothesis that increases in alcohol consumption at the drinking age have causal spillover effects on the alcohol consumption of younger peers.

In chapter 1, I analyze a previously unexamined policy parameter used in unemployment insurance (and other social insurance) programs. Since unemployment insurance replaces a percentage of prior earnings while a claimant is out of work, policymakers must define a base period from which prior earnings are measured. A base period structure common in the Unemployment Insurance (UI) program in the United States has two important implications. First, for claimants with volatile enough earnings, this base period structure creates "benefit risk"—a job loss at the wrong time implies lower benefit amounts. Second, since base periods are determined by the claim filing date, claimants can partially avoid the negative effects of this risk by strategically timing their claims. Using several new sources of administrative data from California's UI program, I provide four key results on this topic. First, I demonstrate that exposure to benefit risk is widespread. Of roughly 22 million claimants in my sample, over 8 million are exposed to some level of benefit risk. Second,

using a bunching approach I demonstrate that roughly 3% of affected claimants strategically delay their claims after a job loss in order to receive higher benefits. Third, I provide evidence that information frictions are a key barrier preventing more widespread use of this strategic response. Finally, I use a dynamic model of job search and Unemployment Insurance to show that the private welfare costs of benefit risk are large. After accounting for claim-timing responses, the average claimant would trade 4% of their expected Unemployment Insurance benefits to eliminate exposure to benefit risk. This number rises substantially among young and especially low-income claimants.

In chapter 2, which is joint work with Lester Lusher and Rebecca L.C. Taylor, we provide causal evidence of an ex ante moral hazard effect of UI by matching plausibly exogenous changes in UI benefit duration across state-weeks during the Great Recession to high-frequency productivity measures from individual supermarket cashiers. Estimating models with day and cashier-register fixed effects, we identify a modest but statistically significant negative relationship between UI benefits and worker productivity. This effect is strongest for more experienced and less productive cashiers, for whom UI expansions are especially relevant. Additional analyses from the American Time Use Survey reveal a similar increase in shirking during periods with increased UI benefit durations.

Chapter 3, written with Eunju Lee, uses data on sibling pairs near the minimum legal drinking age to provide causal estimates of peer effects in alcohol consumption. Following prior work on other outcomes, we exploit the discontinuous increase in alcohol consumption of the older sibling at the legal drinking age in a regression discontinuity design. Our preferred point estimates imply that the number of binge drinking days reported by the younger sibling decreases by 27% of the mean at the cutoff. While our estimates are somewhat imprecise, we are able to consistently rule out leading positive estimates of peer effects in alcohol consumption. Our research design provides estimates which are interpretable as the causal effect of the peer's alcohol consumption. This is in

contrast to most prior work which instead identifies the causal effect of exposure to the peer. We explain how this distinction matters for policy.

Contents

Chapter 1

Claim Timing and Unemployment Insurance Benefit Generosity

1.1 Introduction

Many social insurance programs replace some percentage of prior earnings while a claimant is away from work during a shock (e.g., unemployment, disability, or the birth of a child). A large literature in economics has studied the optimal design of such programs, focusing on the optimal level of this replacement rate (Baily, 1978; Chetty, 2006). In this paper I analyze the implications of a different program parameter, the "base period" (BP). Largely ignored in the social insurance literature, the BP defines the time period from which prior earnings used to calculate benefits are measured. I show empirically and theoretically that this seemingly innocuous program characteristic can have substantial implications for social insurance claimants.

BPs are important for at least three reasons. First, for claimants with enough earnings volatility, benefit eligibility will vary across BPs. These changes are often dramatic and this implies that program rules expose some claimants to a particular type of variability in their benefits that I refer to as "benefit risk"—if their qualifying event (e.g., job loss) occurs at the "wrong" time they will receive lower benefits. Second, in many programs a claimant's BP is a function of the date on which the claimant chooses to file their claim. This creates a take-up decision on the *intensive* margin. Affected claimants can self-select into more generous benefits by strategically timing their claims. Third, claimants will vary in their exposure to benefit risk and are likely to vary in their ability to respond to it. Since benefit risk is driven by earnings volatility, more exposed claimants are likely to be less advantaged (Hardy and Ziliak, 2014).[1] Within the set of exposed claimants, barriers such as limited knowledge of the relevant program rules and imperfect control over claim-timing (due to, e.g., behavioral factors or liquidity constraints) are likely to prevent some claimants from making claim-timing adjustments.[2] Heterogeneous exposure and heterogeneous claim-timing responses to benefit risk each have the potential to alter the targeting properties of social insurance programs.

In California's Unemployment Insurance (UI) program, the empirical context for this paper, UI benefits are determined by earnings in a BP defined as the first four of the last five completed calendar quarters as of the claim date. It follows that claimants with volatile enough earnings histories will receive different benefits on a claim filed late in quarter q instead of early in quarter $q+1$. Similar BP structures are used in UI programs in every other state, as well as Paid Family Leave and Temporary Disability Insurance programs in several states. The purpose of this paper is to quantify the magnitude of benefit risk in the context of California's UI program, estimate its causal effect on the timing of UI claims, and determine the extent to which exposure and responsiveness to benefit risk are heterogeneous across different types of claimants. I utilize a new administrative dataset which includes the universe of UI claims filed in California (CA) between 1/1/2000 and 12/31/2019, as well as matched worker-firm quarterly earnings records for the universe of UI covered workers in CA from 1995-2019.[3] This setting is useful because of the size and richness of the data.

[1] Past research (e.g., Hardy and Ziliak, 2014), has found that income volatility is concentrated at the top and bottom of the income distribution. However, social insurance programs typically cap payments at a maximum, insulating high-income claimants from benefit risk regardless of their earnings volatility.

[2] These barriers are analogous to similar constraints which have been found to explain incomplete take-up of various social programs on the extensive margin (Currie, 2006).

[3] This data has been acquired through a partnership between the Employment Development Department (EDD)—the state government agency in CA which administers the UI program—and the California Policy Lab.

Over 40 million claims are observed along with detailed information on the claimants and their firms.

My empirical analyses begin by quantifying exposure to benefit risk among UI claimants in CA. To measure benefit risk I compare the UI benefits a claimant would receive if their claim were filed in the quarter of their layoff to the benefits they would receive if their claim were filed in the next quarter. This measure is a useful starting point since it aligns with the claim-timing choice that the claimant faces—a UI claim cannot be filed before the job loss occurs and a job loser is very unlikely to delay their claim more than one full quarter. Benefit risk is found to be extremely common and is often very large. 38% of the relevant claims in the data were filed by workers whose available benefits over the life of the claim would change if they delayed claiming until the quarter after their layoff. Many claimants face dramatically large benefit changes. 5% of all claimants would see their benefits increase by $107 or more per week. Another 5% would see their benefits decrease by $30 or more. I also demonstrate that claimants more exposed to these claim-timing incentives are broadly less advantaged. This is important because policymakers may be interested in targeting benefits towards such claimants.

Visual evidence of bunching in claim date distributions strongly suggest that *some* of these claims are strategically delayed in order to receive higher benefits. Among claimants with an incentive to delay claiming until the quarter after their layoff (i.e., those whose UI benefits would be more generous if they delayed their claims until the next BP), claim date distributions show missing masses prior to the BP (quarter) change and excess masses immediately at the BP change. This bunching behavior is more pronounced among claimants with larger incentives to delay (larger increases in benefits with the new BP) and among claimants who are laid off closer to the end of a calendar quarter.

I use bunching methods to quantify these responses and "count" claimants who strategically delay their claims. I find that between 2.5% and 5.4% of claimants incentivized to delay their claims to receive more generous benefits do so. While this is a meaningful proportion, it is perhaps more surprising how many claimants do not strategically delay—at least 94.6% of affected claimants *do not* bunch, effectively choosing the lower benefit level.

In order to investigate heterogeneity in these claim-timing responses across different types of UI claimants, and provide additional support for a causal interpretation of the bunching results, I develop an alternative regression-based approach. Specifically, I regress an indicator for whether the claimant delays their claim until the next BP on their incentive to delay (parameterized by the change in benefits between the two BPs) utilizing two separate identification strategies. First, I implement a selection-on-observables approach which sequentially adds sets of controls to the baseline model. Second, I exploit variation in the claim-timing incentives driven solely by large changes to the UI benefit schedule in CA during the early 2000s. These policy changes differentially affected claimants based on their prior earnings histories, were very large, and their effects (on any outcomes) have yet to be estimated in the literature. The two identification strategies produce similar results which are broadly in line with bunching methods—there is a moderately-sized but meaningful claim-timing response to the change in benefit generosity between base periods.

This more flexible approach also demonstrates that these responses are heterogeneous in one key expected dimension: the length of time the claimant needs to wait to receive the new BP. Unsurprisingly, claimants laid off very early in a calendar quarter (who would need to delay their claims several months in order to reach the new BP) are virtually unresponsive to these incentives. I also investigate heterogeneity along several other dimensions including a simple measure of predicted unemployment duration. This is an important dimension because delay is costly, and paying this cost may not be worthwhile for claimants experiencing short unemployment spells. I show that claim-timing responses are stronger in exactly the groups that are more exposed to benefit risk. This suggests that these claimants may be able to effectively undo some of the negative effects of benefit risk exposure. However, there is no clear pattern of heterogeneity by predicted unemployment duration. I also use similar approaches to show that claimants incentivized to claim sooner—because their benefits decrease with the new BP—do not engage in such strategic behavior.

Strategic claim-timing responses can be thought of as a take-up decision on the intensive margin. Viewed in this light, barriers to take-up are another group of potential explanations for the relatively limited amount of strategic claim-timing observed. Information frictions are one such barrier often found to be important in the wider literature on social programs (e.g., Mastrobuoni, 2011; Chetty, Friedman and Saez, 2013; Armour, 2018; Barr and Turner, 2018; Finkelstein and Notowidigdo, 2019). Given the complex program rules involved with the claim-timing decision that I study, similar barriers may be relevant. In fact, the agency which administers UI in CA routinely notifies *some* claimants about upcoming BP changes. Every calendar quarter roughly 8,000 claimants who would have seen their benefits increase (by any amount) had they delayed their claim until the following week are notified of this fact and given the option to revisit their claim-timing decision. Roughly 400,000 claims in my sample received this information and were given the option to switch to the higher-benefit BP ex post. Just under 150,000 of these claims, roughly 39%, were delayed. In other words, among claimants who were incentivized to delay their claim, but failed to do so initially, 39% changed their decision when given the opportunity and made aware of the exact incentives that they faced. This suggests that incomplete information is a key barrier to take-up in this setting.

Finally, I provide a simple back-of-the-envelope calculation of the private welfare costs of benefit risk. I do this by adapting a standard dynamic model of job search and UI from Schmieder and Von Wachter (2016) to include benefit risk. Using the model, I compare the expected utility of claimants given the current structure of the UI program (i.e., with benefit risk) to their expected utility in a hypothetical alternative system without benefit risk. I define a risk premium as the percentage reduction in expected UI benefits (in the no benefit risk system) that would make the claimant indifferent between the two systems. Using a combination of observed (e.g., the variation in benefits across base periods) and assumed parameters (e.g., counterfactual layoff dates), I calculate a risk premium for each claim in the data. These calculations can be made with and without allowing for strategic claim-timing responses. Without claim-timing responses the average claimant in the data would trade 6.4% of their expected UI benefits to remove benefit risk. Allowing for claim-timing responses reduces this average risk premium to 4%. While highly stylized, this exercise demonstrates that the differences in benefit generosity between adjacent base periods has meaningful welfare consequences for UI claimants.

This paper makes two important contributions to the literature. First, I demonstrate that the structure of base periods in social insurance programs can have important implications for program claimants. Due to earnings volatility, a commonly used base period structure exposes a large number of primarily less-advantaged claimants to benefit risk. Using a simple model, I show that UI claimants incur a meanginful welfare cost from this risk. A small body of existing work on social programs has called attention to the interaction of program design with earnings volatility. For example, estimating the ability of transfer programs to smooth periods of income instability (Hardy, 2017), and analyzing the interaction between earnings volatility, the design of recertification periods, and program churn (Prell, 2008; Pei, 2017). To my knowledge, this is the first work to highlight the connection between time aggregation in benefit formulas and benefit risk in the context of social insurance programs. This adds to a growing body of evidence on the role of time aggregation in social program benefit determinations more generally (e.g., Prell, 2008; Graves, 2012; Shore-Sheppard, 2014; Pei, 2017; Hong and Mommaerts, 2021) and in tax assessment (e.g., Vickrey, 1939; Milton and Mommaerts, 2020). Second, I demonstrate that claimants exposed to benefit risk in California's UI program strategically time their claims to take-up additional benefits. Although several papers have analyzed the extensive margin take-up decision in UI (whether or not to claim) from both theoretical and empirical perspectives (Blank and Card, 1991; McCall, 1995; Anderson and Meyer, 1997; Ebenstein and Stange, 2010; Auray, Fuller and Lkhagvasuren, 2019), no existing work has identified or analyzed the claim-timing decision that I study.[4] Since the relevant parameters of CA's UI program are broadly similar to those used by UI programs in many other US states (and several other social insurance programs), each of these findings has implications which extend beyond my empirical setting.

The remainder of this paper is structured as follows. Section 1.2 describes the UI program in CA, with a specific focus on the relevant aspects of benefit schedules. Section 1.3 summarizes the data sources used in my empirical analyses. Section 1.4 presents some simple measures of exposure to benefit risk (including the magnitude of this risk and how it varies by certain claimant characteristics), and discusses a simple model which demonstrates the welfare implications of benefit risk. Section 1.5 demonstrates that some claimants strategically time their claims to avoid the negative effects of benefit risk, and quantifies this behavior. Sections 1.6 and 1.7 investigate heterogeneity in this strategic behavior. Section 1.8 provides some simple back-of-the-envelope calculations of the private welfare cost benefit risk. Finally, section 1.9 concludes.

1.2 Institutional Context

1.2.1 UI Benefit Levels in California

The UI program in the US is run by individual states. States have the discretion to set benefit levels and durations (with some constraints), among other program parameters, but in every state

[4] Similar claim-timing decisions in other social insurance programs have also not yet been studied, with one exception. A large related literature on claim-timing in the Social Security program has attempted to explain why many retirees do not time their Social Security claims to maximize benefit receipt (see, e.g., Coile et al., 2002; Sass, Sun and Webb, 2013; Henriques, 2018).

benefits are determined based on prior earnings in a “Base Period” (BP) of four completed calendar quarters. The remainder of this section will describe the specific rules which determine BPs, benefit levels, and benefit durations in the State of California.

As shown in Figure 1.1 a UI claimant’s BP is defined as the first four of the last five completed calendar quarters prior to the beginning date of their claim. The beginning date of the claim is typically set as the Sunday prior to the claim filed date. In other words, if a claim begins in quarter q the BP for that claim is the four quarter period beginning in quarter $q-5$ and ending in quarter $q-2$. If the same claimant instead begins their claim in quarter $q+1$, their BP spans quarters $q-4$ through $q-1$. The generosity of UI benefits are defined by two parameters, the weekly benefit amount (WBA), and the maximum number of weekly payments that can be made during the life of the claim (potential benefit duration, or PBD). The maximum $ amount payable during the life of the claim (maximum benefit amount, or MBA) is defined as the WBA multiplied by the PBD. If a claimant temporarily regains employment before exhausting their UI benefits, the claim can be reopened at any time during the 52 week period following the start date of the claim.

WBAs and PBDs are functions of earnings in the claimant’s BP, specifically their total base period earnings, or Base Period Wages (BPW), and their earnings in the highest earning quarter of the base period, or High Quarter Wages (HQW). The specific functions used are:

$$WBA = \min\left\{HQW \cdot \frac{1}{13} \cdot RR,\, WBA_{max}\right\}$$

Where RR is the replacement rate (% of pre-claim earnings being replaced by UI benefits). While these formulas appear complex, they do have straightforward interpretations. The WBA is equal to a proportion, RR, of the average weekly wage in the high earning quarter, up to some WBA_{max}.

Since my analyses focus on variation in WBA, additional information on the determination of PBDs is included in Appendix 1.A.

Finally, to be UI eligible, claimants must meet minimum earnings requirements ($HQW \geq$ \$1,300, or $HQW \geq$ \$900 and $BPW \geq$ \$1,125).

1.2.2 Policy Changes

Replacement rates and maximum $WBAs$ in California changed four times during the time period covered in my analyses. As shown in Figure 1.2 these four changes occurred on 1/1/2002, 1/1/2003, 1/1/2004, and 1/1/2005 and differentially affected workers based on their prior earnings histories. These changes were instituted as part of California's Senate Bill 40, passed on 10/1/2001. In each case, new claims beginning on or after the date of policy change received the new RR and WBA. Claims filed before January 1st of the relevant year received the prior year's WBA schedule. As described in more detail in the section 1.2.3, this is helpful for my purposes as it creates claim-timing incentives which differentially affect claimants based on their earnings histories and layoff dates that are driven solely by the policy change.

These policy changes were significant, with maximum WBAs increasing by \$100 (from \$230 to \$330, +43%) on 1/1/2002, and by \$40 in each of the remaining years—to \$370 on 1/1/2003 (+12%), \$410 on 1/1/2004 (+11%), and finally \$450 on 1/1/2005 (+10%). To my knowledge, no existing research has studied the effects of these policy changes (on any outcomes).

1.2.3 The Claim-Timing Decision

With the formulas from section 1.2.1 in mind we can consider some specific examples of earnings histories that would result in a claimant's benefit level and/or duration differing between two adjacent BPs. Table 1.1 demonstrates how different earnings histories translate into different claim-timing incentives. In each case I consider a claimant laid off in quarter q and show the WBA that the claimant would receive if their claim were filed in quarter q vs $q+1$. The example earnings histories are chosen to highlight the existence of four different sources of variation in claim-timing incentives:

1. The magnitude of earnings volatility experienced by the claimant in the 5 quarters preceding the quarter of the job loss
2. The timing of this earnings volatility
3. Whether or not this earnings volatility pushes a claimant across the maximum WBA threshold
4. Whether the layoff occurs in a quarter of a policy change (i.e., the quarter before a new RR and/or WBA_{max} are applied to new claims)

In Table 1.1, comparing claimant 1 to claimant 2, demonstrates the clearest source of variation in claim-timing incentives. Claimant 1 has no earnings volatility in the relevant quarters so that BP_1 ($q1$-$q4$) provides the same benefits as BP_2. Claimant 2 on the other hand can increase their WBA by waiting for BP_2, since for this claimant $HQW_1 = \$10k < HQW_2 = \$15k$. Comparing claimant 2 to claimant 3 demonstrates the importance of both the magnitude and the timing of earnings volatility within the five relevant quarters. Despite having the same five quarterly earnings amounts as claimant 2, claimant 3 has no claim-timing incentive due to the ordering of those amounts. Defining the magnitude of earnings volatility as the standard deviation of the five quarterly earnings amounts, claimant 4 has earnings volatility identical in magnitude to claimant 2. However, claimant 4's earnings amounts are such that they are always eligible for the maximum WBA. Finally, claimant 6 is an example of a claimant whose claim-timing incentives are driven entirely by a policy change (here the increase in WBA_{max} for claims effective on or after 1/1/2005).

Claimant 5 has an identical (relevant) earnings history to claimant 6 but was laid off one quarter earlier, not exposed to the policy change, and has no incentive to delay their claim.

1.3 Data

In my analyses I utilize administrative data from the UI program in the State of California for the years 1995-2019. Specifically, I combine three administrative datasets maintained by the State of California's Employment Development Department (EDD)[5]: Quarterly earnings records (1995-2019), the Quarterly Census of Employment and Wages (QCEW, 2000-2019), and UI claims microdata (2000-current).

The quarterly earnings records are administrative earnings records which exist in each state and were originally developed to administer the UI program (e.g., to determine benefit eligibility). Earnings are recorded at the firm-employee-quarter level. Each record includes a masked individual identifier, a masked UI account number (unique employer-level identifer used by EDD), and the total earnings paid to the employee in the relevant quarter for each UI-covered job in the State of California.

The QCEW data are administrative records of establishment-level earnings and employment which, much like the quarterly earnings records, exist in each state and are used to administer the UI program (e.g., to facilitate calculation of UI payroll tax liability). The QCEW data includes information on industry (6-digit NAICS code), size (number of employees), and location for all business establishments in CA with UI-covered employees, and is linkable to the quarterly earnings files via the UI account number.

UI claims microdata consists of a variety of information collected or produced by EDD in order to process UI claims. The UI claims microdata contains the universe of UI claims filed in CA on

[5]EDD is the government agency in CA which adminsters, among other programs, UI, Temporary Disability Insurance, and Paid Family Leave.

or after 1/1/2000 and includes a variety of detailed information about each claimant, their work history, their benefit eligibility, and the UI payments that they receive. Key information used in my analyses includes the claimant's self-reported last worked date, the date on which the claim was filed, a masked individual indentifier linkable to the quarterly earnings records, and detailed claimant demographics. In total the data contains roughly 41 million UI claims.

Similar datasets from other states have been used by other researchers.[6] However, the data used in this analysis is unique in two key ways. First, due to both the size of the state of California and the long time period covered, the dataset used in this paper is an order of magnitude larger than those previously used in the literature. Second, the UI claims data used in this analysis is abnormally rich, notably including exact last worked and claim-filing dates which are needed to study the questions of interest. A subset of these data has been used in a series of policy briefs written by myself and a team of researchers at the California Policy Lab (Bell et al., 2020).

The goal of my empirical analyses is to quantify both exposure to benefit risk and resulting strategic claim-timing behavior in the California UI program. The starting point for my analyses is the UI claims microdata. I exclude claims for which I cannot observe key information necessary to quantify exposure to benefit risk and/or claim-timing responses. Notably, my measures of benefit risk and claim-timing will require me to observe the date that the claimant lost their job and their quarterly earnings amounts in the six calendar quarter period ending with the quarter of their job loss. Therefore, I drop claimants with missing last worked dates, claims filed after 12/31/2019 (whose pre-claim earnings are not fully observed), claims for various special types of UI which follow different benefit rules or are otherwise different from typical UI claimants (e.g., the Disaster Unemployment Assistance program, and the Short Time Compensation program), claims filed by

[6] For example, several researchers have used administrative datasets from state UI systems including Florida Johnston (2021), Missouri Card et al. (2015); Johnston and Mas (2018), New York Meyer and Mok (2014), and Ohio Leung and Pei (2020). Several earlier papers also made use of the Continuous Wage and Benefit History data, which combined administrative data from the UI programs of several states during the 1970s and 80s (e.g., Anderson and Meyer, 1997; Landais, 2015).

workers who either reside out of state or were denoted by EDD as having out-of-state wages in the relevant pre-claim time period (since out of state wages are used to determine benefit eligbility but are not observable in the data I use), and claims which would be monetarily ineligible in either of the two BPs of interest in my analyses. This leaves me with 22.2 million claims,[7] the characteristics of these claims are described in Table 1.2, Table 1.3, and Table 1.4.

1.4 Benefit Risk

In this section I begin by outlining the measures of benefit risk that I use in my analyses and describing how I calculate those quantities in the data. Next, I demonstrate that exposure to benefit risk in my setting is both substantial and highly concentrated among more disadvantaged groups. Finally, I present a simple framework to contextualize the potential welfare cost of benefit risk for UI claimants.

1.4.1 Measurement

Linking the UI claims microdata to the quarterly earnings files via the masked identifier allows me to observe complete earnings histories for every UI claimant. I use these earnings histories to calculate benefits that the claimant would be eligible for in two different base periods: the base period if their claim were filed during the quarter in which they reported working last (denoted BP_1), and the base period if their claim were filed in the following quarter (denoted BP_2). I define a simple measure of exposure to these BP-driven changes in benefit generosity as the change in WBA between these two adjacent BPs. The WBA if the claim were filed in the quarter *after* the layoff, denoted WBA_2, minus the WBA if the claim were filed in the quarter of the layoff, WBA_1.

[7]In a subset of my analyses, presented in section 1.8, I am interested in the variability of benefits over a larger set of *four* (instead of two) BPs, and therefore use a weaker restriction that the claimant is eligible in at least one of those four BPs. That sample includes 24 million claims.

I will refer to this measure throughout as ΔWBA.[8] Using the quarterly earnings data, the last worked date, and the benefit formulas described in section 1.2.1, I calculate WBA_1 and WBA_2 for each claim in the sample.

ΔWBA will treat an additional \$1 in UI benefits equally for all claimants, regardless of their prior income level. In order to provide a second measure scaled by prior earnings I define two replacement rates, RR_1 and RR_2, where again the subscript denotes the base period in which the claim is filed. In each case the numerator of the RR is the WBA in the relevant BP and the denominator is a measure of earnings *across both* BPs. Specifically, the denominator is the average weekly earnings in the highest earning quarter of the five quarter period spanning both BPs. ΔRR is therefore interpretable as the percentage point change in the replacement rate received by the claimant across these two BPs.[9]

As a first step in my analysis I identify all claimants in the full sample of 22.2 million claimants who are exposed to any amount of benefit risk as measured by $\Delta WBA \neq 0$ (or equivalently $\Delta RR \neq 0$). Figure 1.3 shows the distribution of this mesaure of benefit risk in the subsample of 8.3 million claimants (38% of the full sample) who have some change in their WBA (or RR) between the two base periods ($\Delta WBA \neq 0$).[10] The bottom panel of the figure shows the same histogram in terms of ΔRR. Further, the magnitude of this risk is often large for each of these two measures. For example, 10% of the claimants included in the top panel of Figure 1.3, or 3.8% of the full sample, would see their WBA increase by at least \$129 if their claim were delayed until the quarter after their layoff. Each panel also demonstrates that a small but meaningful portion of benefit risk exposure entails the loss of all benefits if a claim is filed "too early" or "too late." In each bin the

[8] I similarly define $\Delta MBA = MBA_2 - MBA_1$ and $\Delta PBD = PBD_2 - PBD_1$. However, I focus on the WBA based measure in my descriptive analyses. Various results reported below for ΔWBA are also reported for these measures in Appendix 1.C.

[9] Note that these replacement rates are *not* the replacement rates that are defined by the benefit functions in section 1.2.1 (i.e., the slopes for values of WBA below WBA_{max} in Figure 1.2), since those replacement rates use the HQW from a single BP as the denominator.

[10] Figure 1.A1 shows similar results for ΔMBA and ΔPBD

portion of the bar accounted for by this group is shaded in black. The remainder (in gray) consists of claimants who are eligible in each BP, but for different benefit levels (i.e., each panel of the figure consists of two histograms which are "stacked," not overlaid).

These measures of benefit risk are of interest because they create incentives for claimants to alter the timing of their claims in order to take-up additional benefits. It is worth noting that these measures underestimate benefit risk created by the BP structure in the UI program for two reasons. First, these measures are limited to workers with realized UI claims and to the two potential BPs available to those claimants conditional on the realized date of their job loss. Second, the ΔWBA and ΔRR variables ignore exposure to benefit risk that operates through variation in PBD. These measures therefore also ignore the PBD extensions which occur in CA during downturns and have been studied, for example, by Farber, Rothstein and Valletta (2015) and Rothstein (2011). By increasing the total amount of benefits at stake, these extensions may amplify benefit risk and the associated claim-timing incentives. I abstract away form these sources of benefit risk for simplicity.

1.4.2 Benefit Risk Exposure is Concentrated in Disadvantaged Groups

We would expect exposure to benefit risk to vary with various claimant characteristics for two reasons. First, benefit risk exposure is primarily driven by earnings volatility. Second, since WBAs are increasing in prior earnings and capped at a maximum amount, the highest earning claimants will remain unexposed to benefit risk even with substantial amounts of earnings volatility. Since prior research has found that income volatility is concentrated at the top and bottom of the income distribution, this implies that benefit risk is likely to be concentrated among the low-income (Hardy and Ziliak, 2014).

To visualize the relationship between benefit risk and claimant characteristics, I take the absolute value of ΔWBA (or ΔRR) and graph the average values of these amounts across groups defined by prior earnings, age, completed education, and race/ethnicity. Results are shown in Figure 1.4 and Figure 1.5. In each case, results demonstrate that less advantaged claimants are more exposed to benefit risk. Among the highest earning UI claimants in the sample (roughly the top two deciles of earnings in the 5 pre-claim quarters that makeup BP_1 and BP_2) there is virtually no expsoure to benefit risk.[11] In contrast, the average claimant in the bottom two deciles of this prior earnings measure is exposed to a \$47 change (in absolute value) in WBA between the two adjacent BPs. Younger and less educated claimants are also differentially exposed to benefit risk, although less dramatically. Considered in terms of RR the differences are even more stark. The highest earning claimants see virtually no change in RR between the two BPs, while the bottom deciles of claimants by prior earnings see a 16pp increase in the proportion of prior earnings replaced by UI.

These results uncover a previously unknown source of inequity in a key social insurance program. UI benefits received by lower-income, younger, less-educated, and minority claimants are volatile. The extent to which UI is able to protect such claimants from the risk of lost earnings during unemployment will depend upon whether their job loss and claim-filing dates occur at a better or worse point in their earnings history. Since policy parameters similar to BPs are used in all other wage-replacing social insurance programs, these results also imply that similar inequities are likely to exist in other settings.

1.4.3 The Private Welfare Costs of Benefit Risk

To provide a simple measure of the welfare costs of benefit risk, I adapt a standard dynamic model of job search and UI from Schmieder and Von Wachter (2016) to include benefit risk by intoducing uncertainty over UI benefit levels, denoted by b. Using this model, I compare the expected utility of a representative claimant given some level of benefit risk, $\mathbb{E}[U(b)]$, to the utility

[11] The small amount of benefit risk exposure at the high end of the earnings distribution is primarily driven by high earners who lose their jobs in the quarter before one of the policy changes described in section 1.2.2.

that that claimant would receive if benefit risk were removed while holding the *expected* benefit level constant, denoted $U(\mathbb{E}[b])$. Under the assumption that the claimant is risk averse, concave utility implies that $U(\mathbb{E}[b]) > \mathbb{E}[U(b)]$. Finally, I define a risk premium rp as the drop in consumption in the no benefit risk case that equalizes these two values: $U(\mathbb{E}[b] - rp) = \mathbb{E}[U(b)]$. This risk premium provides a simple measure of the private welfare cost of benefit risk which I take to the data later in the paper.

The model consists of a representative UI claimant who begins an insured unemployment spell at time t = 0. The model is in discrete time and ends at time T. While unemployed, the worker consumes $c_{u,t} = b_t + A$,where b_t is the unemployment benefit and A is income from other sources. The unemployed worker also exerts job search effort s_t at cost $\psi_t(s_t)$, remaining unemployed at time t with probability S_t. Upon reemployment consumption is $c_e = w - \tau$, where τ is a lump sum tax that finances UI benefits.

Unemployment benefits are set to b until $t = P$, at which point benefits are exhausted, i.e. $b_t = 0$ for $t \geq P$. I adapt the model to accomodate benefit risk by assuming that b is stochastic: $b \sim F(b)$. This distribution is discrete, with up to K possible values occurring with probability q_k, so that $\sum_{k=1}^{K} q_k = 1$ and $\sum_{k=1}^{K} q_k b_k = \mathbb{E}[b]$.

After accounting for benefit risk, the claimant's expected utility is:

$$\mathbb{E}[U] = \sum_{k=1}^{K} q_k \left(\sum_{t=0}^{P} S_t u(b_k + A) + \sum_{t=P+1}^{T} S_t u(A) + \sum_{t=0}^{T} (1 - S_t) u(w - \tau) - \sum_{t=0}^{T} S_t \psi_t(s_t) \right) \tag{1.1}$$

The claimant's risk premium rp solves:

$$\sum_{t=0}^{P} S_t u(\mathbb{E}[b] + A - rp) + \sum_{t=P+1}^{T} S_t u(A) + \sum_{t=0}^{T} (1 - S_t) u(w - \tau) - \sum_{t=0}^{T} S_t \psi_t(s_t) = \mathbb{E}[U] \tag{1.2}$$

This framework makes clear that there exists some private welfare cost of benefit risk so long as claimants are risk averse and claimants face some uncertainty over their UI benefit levels (i.e. $\exists k$ s.t. $q_k \neq 0$ and $b_k \neq \mathbb{E}[b]$). Earlier in this section I quantified exposure to benefit risk using simple measures which directly connect to the claim-timing decisions I will study later in the paper. The risk premium measure presented here serves as a useful complement to those earlier measures since it allows me to directly analyze the normative implications of benefit risk. In section 1.8, I will use this framework to provide some simple back of the envelope calculations of the private welfare cost of benefit risk.

1.5 Do Claimants Strategically Delay?

In this section I present results which demonstrate that some claimants respond to benefit risk by strategically timing their claims so that they receive more generous benefits. For simplicity, I focus on the effect of ΔRR on a simple measure of claim-timing: An indicator for whether the claim was filed in the first week of BP_2 (i.e., the first week of the quarter after the layoff) and focus on claimants with either no incentive to time their claim ($\Delta RR = 0$) or some incentive to delay ($\Delta RR > 0$).

1.5.1 Visual Evidence

As a starting point, I present descriptive evidence in Figure 1.6 which shows the fraction of claimants in claim-date bins around BP changes in each of four groups defined by $\Delta RR = RR_2 - RR_1$. Groups are (1) claimants with $\Delta RR = 0$ and (2)-(4) defined by terciles of ΔRR among claimants with $\Delta RR > 0$. In the figure, I bin claim dates at the weekly level and center them around the closest BP change, so that week zero is the first week in which the new BP is effective (i.e., the week beginning with the first Sunday in a quarter). There is clear and substantial bunching at the BP change among claimants with large incentives to wait for the next BP. This is strong evidence that some claimants are aware of and responsive to these incentives. However, these bunching results abstract from a second useful source of variation in the data: The distance between the layoff date and the BP change.

We would expect that these claim-timing responses to benefit risk are decreasing in this distance. To provide visual evidence for this, Figure 1.7 graphs several distributions of the number of weeks between layoff and claim dates. Each panel is limited to claimants laid off some number of weeks before the BP change, and shows these distributions in two groups: Claimants with $\Delta RR = \$0$ (no incentive to delay), and claimants with $\Delta RR > 0$ (some incentive to delay). In each panel there is a clear spike in the claim-date distribution for the second group in exactly the week where the BP changes. While this spike is very small for long wait times, it clearly grows as the layoff date moves closer to the BP change.

1.5.2 How Many Claimants Delay?

The results described in section 1.5.1 strongly suggest that claimants are responsive to these incentives in deciding when to file their claim. However, they do not allow me to quantify the response or to do inference. To address these issues, I use a standard bunching approach to estimate the number of claimants with $\Delta RR > 0$ who strategically delay their claims. As described in Kleven (2016) for a more general setting, the approach uses the distribution of claim dates away from the date on which the BP changes to estimate a counterfactual distribution in the absence of any incentive to delay. By comparing the observed distribution of the claim dates to this counterfactual distribution around the "notch," we can quantify the bunching response—effectively counting the number of claims that were strategically delayed until the start of the next BP.

I start by estimating the following regression among claims with $\Delta RR > 0$, using data binned to the day of claim, d, where d is centered at the nearest BP change as in Figure 1.6:

$$claims = \sum_{i=0}^{p} \beta_i^p d + \sum_{k=-14}^{13} \gamma_k \cdot 1\{d = k\} + \sum_{j=1}^{6} \delta_j \cdot 1\{dow(d) = j\} + \epsilon \tag{1.3}$$

Where *claims* is the number of claims filed on day d, δ_j are coefficients on day-of-week dummies with Sundays the excluded category[12], p is the polynomial order used to fit the distribution, and days -14 to 13 make up the "manipulation" region. I determine the manipulation region in an ad hoc manner by visually inspecting the distribution. While more automated approaches exist and are common in the literature, the combination of diffuse bunching and a relatively discrete running variable in my setting make these approaches difficult to implement.

Next, I estimate the counterfactual distribution as the fitted value from this regression excluding the manipulation region dummies:

$$\hat{claims} = \sum_{i=0}^{p} \hat{\beta}_i^p d + \sum_{j=1}^{6} \hat{\delta}_j \cdot 1\{dow(d) = j\} \tag{1.4}$$

Finally, I sum the gaps between the counterfactual distribution and the empirical distribution of claim dates for each claim date above the notch and within the manipulation region. This is

[12]Likely due to a combination of claimant behavior and EDD processes, the claim date distribution exhibits substantial day-of-week effects, with relatively few claims filed on weekends. This is akin to the "round number bunching" issue common in tax bunching settings (see, e.g., Kleven and Waseem (2013)).

the "excess mass," or the number of claims that were strategically delayed. Standard errors are calculated via bootstrap.

The key threat to the validity of this approach is that the distribution outside of the manipulation region may not serve as an adequate counterfactual. This would occur if, for example, there were some other reason unrelated to the change in benefit generosity for claimants to bunch at day zero. To investigate, and if necessary correct for, this concern we can exploit the group of claims with no incentive to bunch (i.e., $\Delta RR = 0$). First testing for bunching in this group, and second, exploiting variation within the $\Delta RR > 0$ group to demonstrate that the amount of bunching grows with incentives to bunch.

Figure 1.8 shows, for both the $\Delta RR > 0$ and the $\Delta RR = 0$ groups, the empirical distribution of claim dates, the counterfactual distribution estimated as described above, the bunching estimate (cumulative distance between empirical and counterfactual distributions in bunching region), a bootstrapped standard error for the bunching estimate, and the bunching estimate as a percentage of all claims in the relevant group. In the $\Delta RR > 0$ group, we see a clear excess mass of claims on the right of the cutoff and a clear missing mass on the left much like Figure 1.6. I estimate that 278,000 claims, or 5.4% of claims with $\Delta RR > 0$, bunch—i.e. are strategically delayed until the BP changes at day zero. However, we see in the bottom panel for the $\Delta RR = 0$ that bunching also occurs in the control group, although it is much less substantial at 2.5% of claims.

One potential explanation for the presence of bunching in the control group is demonstrated in Figure 1.A2, which shows that layoffs are concentrated in the last week of a quarter in my sample (regardless of ΔRR values). Since, as shown in Figure 1.7, the vast majority of claims are filed in the week after the job loss, this would lead to an excess mass of claims in the first week of a base period in all ΔRR groups. Regardless of the underlying reason, a reassuring feature of the bunching apparent in the "treatment" group is that the magnitude of the bunching behavior grows with incentives to bunch. In Figure Figure 1.A3 I implement the same bunching approach used above in groups defined by terciles of ΔRR values (excluding claims with $\Delta RR = 0$). In the first tercile ($\Delta RR < .06$), 4.1% of claimants bunch. This number grows to 4.8% in the second tercile ($.06 \leq \Delta RR < .29$) and 7.4% in the third ($\Delta RR \geq .29$).

The bunching results provide strong evidence that between 2.8% and 5.4% of claimants with $\Delta RR > 0$ strategically delay their claims. Perhaps more interesting is the implication that at least 94.6% *do not* strategically delay, effectively choosing the lower benefit level. In the remainder of this section I provide some additional evidence to support the bunching results before turning to the question of why some claimants engage in this strategic behavior while many others do not.

1.5.3 An Alternative Approach

The bunching results from section 1.5.2 quantify the number of claimants that strategically delay claiming. However, both the diffuseness of the bunching and the existence of bunching among claimants with no incentive to delay call for additional evidence to support this interpretation. Further, the bunching results abstract away from the dramatic variation in incentives *within* the subset of claimants with $\Delta RR > 0$. Figure 1.6 suggests that strategic claim delay is very responsive to the magnitude of ΔRR within this group. Finally, Figure 1.7 clearly shows the importance of a second key dimension of heterogeneity in strategic delay—the distance between the job loss date and the next BP—which is also ignored by the simple bunching approach. To address these issues, I develop an alternative and more flexible approach to quantify strategic claim-timing responses by estimating regressions of the following form:

$$delay_c = \sum_{-12}^{-1} \beta_1^\tau 1\{week_c = \tau\} + \sum_{-12}^{-1} \beta_2^\tau 1\{week_c = \tau\} \cdot \Delta RR_c$$
$$+ \beta_3 \Delta RR_c + X_c' \beta_4 + \gamma_{d(c)} + \psi_{q(c)} + \epsilon_c \quad (1.5)$$

Subscripts denote claims (c), week of layoff (τ, in event time relative to the BP change[13]), weekday of layoff (d), and quarter of layoff (q). The outcome is some measure of waiting for the quarter *after* the layoff to claim (i.e., proxies for a claimant "choosing" the new BP), X_c is a large vector of covariates (e.g., demographics, employment history, etc) which are interacted with week dummies in all specifications unless otherwise noted, γ_d and ψ_q are weekday and quarter of layoff FEs. The coefficients of interest are β_3 and the β_2^τs.

For my main results I use an indicator for whether or not the claim was filed in the first week of the new BP as the outcome. Figure 1.9 shows results for a baseline regression, estimated among all claimants in the analysis sample with $\Delta RR \geq 0$ who are eligible in both BPs (the latter restriction is made to lessen concerns about extensive margin effects). In Figure 1.9 I start with a simple specification that includes only a control for ΔPBD and then progressively add sets of additional covariates to the model. The sets of controls are:

1. "Baseline": week of layoff dummies, layoff date FEs (quarter-year and day of week), and ΔPBD

2. "Demographics": Completed education, gender, age, ethnicity, citizenship status, and 3-digit zip code[14]

3. "Prior Earnings": Average quarterly earnings in the five completed calendar quarters pre-

[13] BPs typically include 13 weeks but are sometimes 14 weeks long. I define τ to range from 1 to 13, where 13 includes claimants laid off 13 or 14 weeks before a BP change and is the omitted category.

[14] To keep the number of regressors manageable I do not interact the zip code dummies with week dummies.

claim and a measure of the magnitude of "effective" earnings volatility in these quarters (more detail on this below)

4. "Pre-separation Employer": Reason for job loss, indicator for whether the claimant expects a recall to the separating employer, size of separating employer (# employees and # establishments), average earnings of employees at separating employer during quarter of sepration, and sector (two-digit NAICS code) of separating employer[15]

Each claimant has five quarterly earnings amounts which make up BP_1 and BP_2 and therefore influence ΔRR. A simple way to control for earnings volatility would be to include some measure of the dispersion of these five quarterly earnings amounts in equation 1.5. However, earnings volatility is only relevant for ΔRR if it occurs at earnings levels below those which correspond to maximum benefits. To measure only the volatility in earnings that matters for ΔRR I instead calculate $WBAs$ for each of the four possible combinations of four earnings amounts from this set of five quarters. Two of these WBA values are WBA_1 and WBA_2, the remaining do not correspond to any possible BP (or claim). My measure of "effective" earnings volatility is the standard deviation of these four $WBAs$.

Results are stable across models and suggest that a 10pp increase in ΔRR leads to a 1.4-1.8pp increase in the probability of filing a claim in the first week of the next BP among claimants who lost their jobs 2-4 weeks before the BP change. This effect fades out as the distance between the layoff date and the BP change widens. As shown in Table 1.3, 7% of all claimants file their claim in the first week of the quarter following their layoff, so the marginal effects for claimants laid off towards the end of a quarter are clearly meaningfully large. While often statistically significantly different from zero, the marginal effects among claimants laid off very early in a quarter are very small (or zero). This makes sense as only those claimants with extremely large ΔRR values should

[15] I do not interact the NAICS sector dummies with week dummies.

be willing to delay their claim a full calendar quarter. A causal interpretation of these results relies on the assumption that there are no unobserved confounding factors and this assumption is ultimately untestable. However, coefficients are remarkably stable as controls are added and this is reassuring. In the next subsection I briefly describe several related specifications which establish the robustness of these results and support a causal interpretation.

1.5.4 Robustness & Threats to Identification

Alternative Specifications

To demonstrate robustness, I vary the specifications from section 1.5.3 in several dimensions. First, using ΔWBA as the RHS variable of interest instead of ΔRR in Figure 1.A5. Second, in all of the aformentioned figures I include claims with $\Delta RR = 0$, but these claimants are very different from claims with $\Delta RR > 0$ and it may be preferable to drop them (e.g., to differentiate this approach from the bunching by focusing on variation in incentives within the $\Delta RR > 0$ group). The (a) panels of Figure 1.A6 and Figure 1.A4 do this for each treatment variable. Coefficients are notably less stable when controls are added, but the general pattern remains.

One potentially concerning pattern in these results is the large drop in the effect of ΔRR on claim delay for claimants laid off in the last week of a BP relative to the second to last week. There are two competing explanations for this pattern. First, as can be seen in Figure 1.7, the vast majority of claimants file their claims either in the week of their layoff or the week after. This could explain the pattern in claim-timing responses since it implies that the vast majority of claimants laid off in week -1 would have filed their claim in week 0 anyway. Second, the pattern could reflect claimants with larger ΔRR values waiting longer to claim for reasons unrelated to the benefit change between BPs. An alternative outcome robust to these concerns is an indicator for whether the claimant waited at least two weeks after their layoff to file their claim. For this outcome we should expect to see a zero effect among claimants laid off in the last week of a BP since they are only incentivized to delay their claim by one week. Results using this alternative claim delay measure as an outcome are shown in Figure 1.A7. These results are less stable as groups of control variables are added and at times produce counterintuitive negative effects of ΔRR on claim delay. However, the broad patterns are the same: Claimants laid off early in the quarter show very small to zero responses to these incentives, claimants laid off late in the quarter are substantially more likely to delay their claim if doing so provides higher benefits as measured by ΔRR. Importantly, the estimate for layoffs occuring in week -1 is zero.

Exploiting Policy Driven Variation

As described above, causal interpretations of the results in section 1.5.3 rely on selection-on-observables assumptions. An alternative approach is to exploit policy-driven variation in claim-timing incentives shown in Figure 1.2 and described in section 1.2.2. Each policy change consists of a change to the WBA schedule, specifically an increase in the WBA_{max} value and in some cases an increase in the replacement rate. These changes differentially affect claimants based on their HQW values. Claimants with higher prior earnings amounts receive larger replacement rate increases.

To isolate the variation in claim-timing incentives driven by these policy changes I assign each claim to an earnings group based on the HQW of the claimant in each of the two possible BPs[16] and add earnings group level fixed effects to equation 1.5. In other words, I estimate the following equation:

[16] The most flexible version of this would require one earnings group for each possible pair of WBAs implied by the HQWs. To make this more tractable I divide each HQW into 26 bins and assign the average change in benefits within the bin.

$$delay_c = \sum_{-12}^{-1} \beta_1^{\tau} 1\{week_c = \tau\} + \sum_{-12}^{-1} \beta_2^{\tau} 1\{week_c = \tau\} \cdot \Delta RR_c$$
$$+ \beta_3 \Delta RR_c + X_c' \beta_4 + \gamma_{d(c)} + \psi_{q(c)} + \delta_g + \epsilon_c \quad (1.6)$$

Where the subscripts denote earnings groups (g) and quarter of layoff (q). For this approach I also limit the sample to claimants who were laid off between 1/1/2000 and 12/31/2007.

Figure 1.10 shows results from this regression using an indicator for whether or not the claim was filed in the first week of the next BP as the outcome, and comparable selection-on-observables results from Figure 1.9, both including the most complete set of controls. Results from the two approaches are broadly similar. Estimates from the diff-in-diff approach imply that an increase of 10pp in ΔRR has moderate effects on claim delay among claimants laid off early in the quarter (a roughly 1.5pp increase in the outcome among claimants laid off in the first 3 weeks of a quarter) that grow to nearly 5pp for claimants laid off 2-3 weeks before the BP change. Results are broadly similar to those in section 1.5.3, although effects are generally larger. Potential explanations for larger responses to policy-change-driven incentives to delay include increased salience (as described in section 1.2.2 these policy changes where announced several months ahead of time) and differences between the groups of claimants exposed to the two types of variation (earnings volatility driven variation is concentrated among the young and low-income while policy change driven variation is more broadly distributed).

Sensitivity of these results to alternative specifications and tests for differential pre-trends in an event study setup are available in Appendix 1.B.

Endogenous Layoff Dates

Both Figure 1.6 and Figure 1.7 consider claim-timing behavior taking the layoff date as given. However, it is also possible for claimants and/or their employers to manipulate layoff dates in response to these incentives. On the claimant side, this could occur through negotiation between the worker and their employer. If the worker is aware of their benefit risk in advance they may be able to influence their employer to alter the timing of the separation. On the employer side, experience rated UI payroll taxes provide an incentive for firms to minimize receipt of UI benefits by their former employees. Experience rating of UI payroll taxes has been shown to influence hiring (Johnston, 2021), firm location decisions (Guo, 2021), and employer challenges to former employees' UI benefit eligibility (Anderson and Meyer, 2000; Lachowska, Sorkin and Woodbury, 2021). This characteristic of the UI system may also induce employers to systematically time layoffs when benefit eligibility is lowest. As a first pass at investigating this margin of response to benefit risk, I produce a bunching figure analogous to Figure 1.6 where the X-axis is the layoff date relative to the new BP. Figure 1.A2 provides little evidence for strategic timing of layoff dates around BP changes.

In a complementary approach I also reestimate all regressions described in section 1.5.3 in a sample of claims that result from mass layoff events. The logic for this exercise is that employers may have less ability to strategically time layoffs in such circumstances, both because mass layoff events are often thought of as relatively unexpected shocks that firms would not have been able to plan, and because ΔRR is likely to vary substantially across workers suffering from the same mass layoff event—making it difficult for the affected firm to time layoffs accordingly. This mass layoff sample consists of a subset of the full sample of 22.2m claims, which appear likely to have occurred due to a mass layoff event at their separating employer. Specifically, I consider a firm to have experienced a mass layoff event if at least 20% of the firm's employees (as measured in the first month of the quarter in the QCEW) file a UI claim and report a last worked date that falls in the same one week period. 3.4m claims are included in the mass layoff sample. Results using this sample are shown in the (c) panels of Figure 1.A4 and Figure 1.A6.

1.6 Who Strategically Times Their Claim?

In ths section I provide a brief narrative overview of dimensions along which we would expect claim-timing responses to vary. I then investigate these potential heterogeneous effects empirically.

1.6.1 Who Should We Expect to Strategically Time Their Claim?

While the claim-timing decision is clearly complex, there are several key dimensions along which claim-timing responses should be expected to vary. First, and most simply, we would expect that claimants who expect to have longer unemployment spells would be more likely to delay. This is because, relative to claiming immediately, delay involves an initial temporary *decrease* in UI benefits received (from WBA_1 to \$0) while the claimant waits to file their claim, followed by an increase from WBA_1 to WBA_2 while the claimant receives benefits. If the spell is too short, delay may cause the claimant to experience a decrease in total benefits received despite the increase in WBA. Second, we might expect claimants to respond differently to different changes in their WBA between BPs. For example, they might ignore small changes while responding strongly to very large ones. Similarly, they may respond asymmetrically so that a postitive value of ΔWBA induces claim-delay while a negative value of ΔWBA does not induce claimants to claim sooner. This might occur if, for example, claim delay is less costly than speeding up the claiming process.

1.6.2 Heterogeneity in Claim-Timing Responses

To investigate heterogeneity in the strategic behavior described in section 1.5, I make three changes to equation 1.5. First, I limit the sample to job losses occurring in the last 5 weeks of a calendar quarter, since that is the range in which claim-timing responses are concentrated as per, e.g., Figure 1.9. Second, I drop interactions between τ (the number of weeks between the job loss and BP change) and ΔRR. This simplifies exposition and allows me to report only one coefficient per subgroup—the average effect of ΔRR. Third, I multiply the estimated coefficient by $\overline{\Delta RR}/\overline{delay}$ so that the reported values are interpretable as elasticities evaluated at the means. This ensures that variation in the effect of ΔRR on claim delay across subgroups is isolated from variation in the distribution of ΔRR and claim delay across subgroups.

The resulting specification is:

$$delay_c = \sum_{\tau=-4}^{-1} \beta_1^{\tau} 1\{week_c = \tau\} + \beta_2 \Delta RR_c + X_c' \delta + \psi_{q(c)} + \epsilon_c \tag{1.7}$$

This equation is estimated via OLS with cluster robust standard errors at the layoff-quarter level. The coefficient of interest is $\hat{\beta}_2$. Results are shown in Figure 1.11 which report $\hat{\beta}_2 \cdot \left(\overline{\Delta RR}/\overline{delay}\right)$, with standard errors calculated by the delta method.

Two broad patterns stand out in these plots. First, groups which are more exposed to benefit risk (as shown in Figure 1.4) are more responsive to the incentive to delay. Since the results in these plots are elasticities, this is not driven by the differing magnitude of those incentives across groups unless the response is nonlinear. This is important because it suggests that the groups most negatively affected by benefit risk may actually be those who are most likely to avoid some of the negative consequences of this risk by strategically delaying their claims. Second, claimants are much more responsive to these incentives during the time period encompassing the policy changes described in section 1.5.4, than afterwards,[17] but effects are otherwise relatively consistent over time. This is surprising since we would expect that claimants expecting long unemployment spells—such as those losing their jobs during the Great Recession—would be more responsive to these claim-timing incentives.

To further investigate heterogeneity in claim-timing respones by expected unemployment duration, I produce a simple nonparametric prediction of unemployment duration for each claim in my sample. These predictions exploit information on the number of weekly UI payments received by each claimant. A complication with this proxy for expected unemployment duration is that realized unemployment duration (the outcome that I predict) is differentially censored. Different claimants often have different PBD values, meaning that the maximum number of weekly payments a claimant can receive varies. To deal with this I predict the probability that each claimant receives at least 13 weekly payments, since 13 is the minimum PBD in CA throughout the time period covered by my sample. Since this is meant to be suggestive, my predictions correspond to the average value of this outcome in cells defined by claimant age, gender, completed education, recall status, industry, tenure at separating employer, and quarter of job loss. Finally, I bin claimants into 9 groups based on the values of this predicted probability and estimate the same regression shown above in subgroups defined by these groups. Results shown in Figure 1.12 are somewhat consistent with the hypothesized effects. Groups with the lowest and highest predicted probabilities of suffering long unemployment spells fit the hypothesized heterogeneity—claimants with short predicted unemployment spells are less responsive and those with long predicted spells are more responsive. However, responses are very similar across the remaining groups.

[17] This is not suprising in light of the results described in section 1.5.4

1.6.3 Claimants Do Not "Hurry Up"

Taken together, the results in Figure 1.6, Figure 1.7, Figure 1.9, and Figure 1.11 show that claimants with an incentive to delay are more likely to do so if that incentive (e.g., ΔRR) is larger, if the "cost" of delay (e.g., the length of time between the job loss and the new BP where benefits increase) is smaller, if the claimant is low-income, if the claimant has less completed education, if the claimant is young, or if the claimant is Black. This heterogeneity is broadly consistent with a framework where claimants are weighing the costs and benefits of delay and responding accordingly. However, there is one key exception to this pattern.

I have focused thus far in this section on the subset of claimants with an incentive to delay claiming. However, as noted in section 1.4, there are also a meaningful number of claimants who have an incentive to speed up their claim—i.e., claimants whose benefits will decrease in the new BP relative to the current one ($\Delta RR < 0$). To investigate this, Figure 1.13 plots the distribution of weeks elapsed between job loss and claim filing for two groups of claimants, a control group with $\Delta RR = 0$ and a treatment group with $\Delta RR < 0$ (much like Figure 1.7, with the $\Delta RR < 0$ group replacing the $\Delta RR > 0$ group). The plot is reproduced for several subsamples, each limited to claimants losing their job some number of weeks before the BP change. In these figures there is no evidence that claimants with an incentive to speed up their claims are differentially likely to file their claims before the BP changes and their benefits decrease.

The first finding highlighted in this section is that 5.4% of claimants with $\Delta RR > 0$ strategically delay their claims. Perhaps more interesting, is the implication that 94.6% *do not*, effectively choosing the lower benefit level. This evidence that such strategic behavior does not occur at all in the other direction is important because it helps us to better understand this latter group. While some of these claimants may have simply decided that delay is not worthwhile, this cannot be the only explanation.

1.7 Information as a Barrier to Strategic Claim-Timing Responses

Information frictions are often found to be important barriers to take-up in the wider literature on social programs (e.g., Mastrobuoni, 2011; Chetty, Friedman and Saez, 2013; Armour, 2018; Barr and Turner, 2018; Finkelstein and Notowidigdo, 2019) and are likely to be especially relevant here given the complexity of the choice that claimants face. To study information frictions in this context, I exploit a subset of UI claimants who are explicitly informed about these incentives by the UI agency in California as part of the normal claim-processing steps undertaken by the agency.

Every calendar quarter roughly 8,000 claimants who would have seen their benefits increase (by any amount) had they delayed their claim until the following week are notified of this fact by the agency and given the option to revisit their claim-timing decision. In my sample, these informed claims include every claim filed on day -7 to -1 in the top panel of Figure 1.8. For the typical claimant, a claim begins on the Sunday prior to the date that they file their claim. Whichever quarter that Sunday falls in determines their BP. An exception is made for claimants who file their claim in the last week of a BP, but would have seen their benefits increase (by any amount) if they had instead waited until the next Sunday where the new BP would become effective. These claimants are notified by the agency that their benefits would have been different if they had delayed their claim one more week, made aware of the exact change in benefits they would have been eligible for, and given the opportunity to revisit their decision (i.e., to delay their claim by one week after it had already been filed).

Table 1.5 presents some simple descriptive statistics on the 398,546 claims in my sample that received this information and option to switch to the higher-benefit BP ex post. Just over 154,000 of these claims, roughly 39%, were delayed. In other words, among claimants who were incentivized to delay their claim, but failed to do so initially, 39% changed their decision when given the opportunity and made aware of the exact incentives that they faced.

1.8 The Welfare Cost of Benefit Risk

The two key results shown so far in this paper are that earnings volatility exposes some UI claimants to large and costly benefit risk, and that some exposed claimants are able to reduce the negative consequences of benefit risk by strategically timing their claims. This section brings these two results together within the theoretical framework laid out in section 1.4.3. This allows me to quantify the private welfare cost of benefit risk accounting for claim-timing responses.

The starting point for this exercise is equation 1.2, which defines the risk premium in terms of the claimants expected utility in states of the world with and without benefit risk:

$$\sum_{t=0}^{P} S_t u(\mathbb{E}[b] + A - rp) + \sum_{t=P+1}^{T} S_t u(A) + \sum_{t=0}^{T} (1 - S_t) u(w - \tau) - \sum_{t=0}^{T} S_t \psi_t(s_t) = \mathbb{E}[U]$$

Where $\mathbb{E}[U]$ is defined as:

$$\mathbb{E}[U] = \sum_{k=1}^{K} q_k \left(\sum_{t=0}^{P} S_t u(b_k + A) + \sum_{t=P+1}^{T} S_t u(A) + \sum_{t=0}^{T} (1 - S_t) u(w - \tau) - \sum_{t=0}^{T} S_t \psi_t(s_t) \right)$$

In earlier sections I focused on a simple two-BP measure of benefit risk which was directly tied to the claim-timing decision that claimants face. Here I widen that view in order to provide a fuller picture of the welfare costs of benefit risk. Figure 1.14 visualizes the approach that I take. For each claim in my sample, I start with the observed job loss date from the claims data. First, I allow for variability in the week of the job loss by assuming that the actual week was drawn from an approximately normal distribution centered at the actual week, and bounded above and below by one calendar quarter.[18] Next, I allow for variability in how long each claimant waits

[18] Specifically, I use a binomial distribution with 26 trials and a probability of succss of 0.5, since binomial distribu-

after their job loss to file their claim by applying the empirical probability distribution of the number of weeks between job loss and claim filing. In order to account for the role of strategic claim-timing responses, I retrieve this distribution for seven separate groups of claimants based on their ΔRR values: those with $\Delta RR = 0$, three groups defined by terciles of ΔRR if $\Delta RR < 0$, and three groups defined by terciles of ΔRR if $\Delta RR > 0$. Combining the observed layoff week, the probability distribution of counterfactual layoff weeks occurring ≤ 1 quarter earlier or later, and the probability distribution of time elapsed between the job loss and the claim, I calculate the probability that each of four potential BPs is realized. (These probabilities correspond to q_k in equation 1.1 where $k = \{1, 2, 3, 4\}$.)

With these four possible BPs identified for each claim in the data, the goal is to solve for each claimant's risk premium. To do this, I also need information on claimant preferences, other sources of consumption while unemployed (A), the probability of remaining unemployed in each period (S_t), and consumption while employed. I make the following assumptions and simplifcations. First, I follow related work and assume that claimants have constant relative risk aversion (CRRA) preferences with coefficients of relative risk aversion equal to 3 (Luttmer and Samwick, 2018; Caldwell, Nelson and Waldinger, 2020). Second, by defining the risk premium as a portion of consumption while the claimant is receiving UI benefits (and not, e.g., a proportion of all period consumption) I have implicitly removed the role S_t and of unemployment duration in general—I return to the implications of this assumption below. Third, I rely on estimates of the change in consumption at UI benefit exhaustion from Ganong and Noel (2019) and Rothstein and Valletta (2017) to back out an estimate of the proportion of consumption while unemployed financed by non-UI sources.[19]

tions with a probability of success close to 0.5 are approximately normal and I want to ensure that the counterfactual job loss weeks are no more than one calendar quarter after the actual job loss so that I can observe enough pre-claim earnings history to determine counterfactual benefits.

[19] Specifically, I assume that the change in consumption at benefit exhaustion is equal to the level of UI benefits. With this assumption, results from Ganong and Noel (2019) suggest that non-UI sources of consumption are roughly 5.7 times UI benefit levels.

Finally I directly measure benefit eligibility in each of the four potential BPs using the claimant's earning history.

I now am able to plug in each of these values to equation 1.2 and solve for rp for each claim in the data. To simplify interpretation, I will scale each risk premium by the claimant's expected benefit level $\mathbb{E}[b]$. This implies that the risk premia I report are interpretable as the percent of expected UI benefits that a claimant would trade in order to remove benefit risk.

Figure 1.15 displays two separate cumulative distributions of risk premia in my sample calculated as described above. In the first, I use the empirical distribution of the number of weeks between job loss and claim-filing among claimants with $\Delta RR = 0$ *for all claimants* (i.e., to calculate risk premia without strategic claim-timing responses). In the second, I allow the claim-timing distribution to vary with ΔRR as described above (risk premia *with* strategic claim-timing responses). Results suggest that the private welfare cost of benefit risk is moderate but meaningful. Without strategic claim-timing responses, the average claimant in my sample would trade 6.4% of their expected UI benefits to eliminate their exposure to benefit risk. With strategic claim-timing responses this number falls to 4%.

Both distributions have long right-tails. Unsurprisingly, many claimants (nearly 60%) have risk premia equal to zero. This is not suprising given that many claimants have no benefit risk exposure, either because their earnings are not volatile or their earnings are always high enough to be eligible for the maximum WBA. On the other hand, there is a relatively small but important group of claimants with *very* large risk premia—the 90th percentile of the risk premium distribution is 16.4% without strategic claim-timing and 8% after allowing for strategic claim-timing. These results make clear that benefit risk exposure has important normative implications for UI claimants, and that, while these welfare costs are reduced by claim-timing repsonses, they remain important after accounting for such responses.

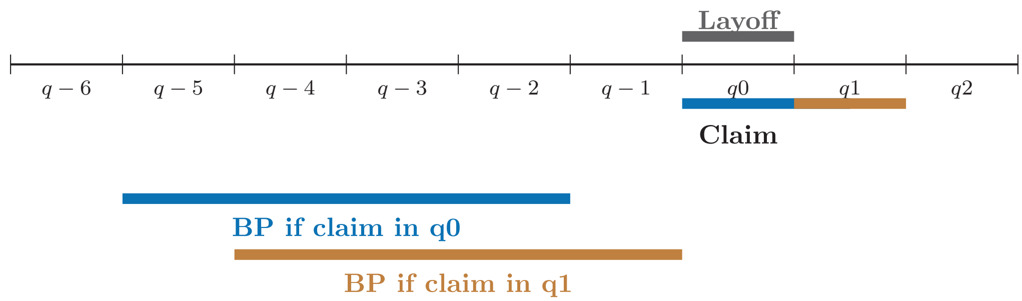
Several important caveats are worth mentioning. First, as mentioned above this model is only superficially dynamic. This can be reframed either as an assumption that all claimants are exhausting benefits, or as scaling the risk premia by the amount of UI benefits each claimant expects to receive. Second, I have implicitly assumed that strategic claim-timing adjustments are costless in these calculations. This is very unlikely to be true, implying that these risk premia estimates are likely *underestimates* for any claimants who take advantage of the option to strategically delay their claims. Third, claimants are assumed to be identical in all dimensions except for ΔRR. This is important since substantial heterogeneity is likely to exist in other dimensions (such as non-UI sources of consumption during unemployment, or unemployment duration), and this heterogeneity is likely to be strongly correlated with exposure to benefit risk (e.g., the broadly more disadvantaged group exposed to benefit risk is likely to have less access to non-UI sources of consumption during unemployment and longer unemployment spells, both implying larger welfare costs from benefit risk).

1.9 Conclusions

I demonstrate that base periods, a parameter previously ignored by the literature on social insurance programs, can have substantial implications for program claimants. Base periods define the pre-claim time period from which earnings are measured in order to calculate benefit eligibility. Using data from California's UI program, a commonly used base period structure is shown to expose many claimants to a previously unexamined form of risk, which I call "benefit risk." Benefit generosity varies, sometimes dramatically, for such claimants based on when their job loss occurrs. Claimants exposed to benefit risk are broadly less-advantaged than those who are not. I show theoretically that benefit risk reduces the value of UI and empirically that claimants engage in strategic claim-timing behavior in order to partially reduce the negative effects of this risk.

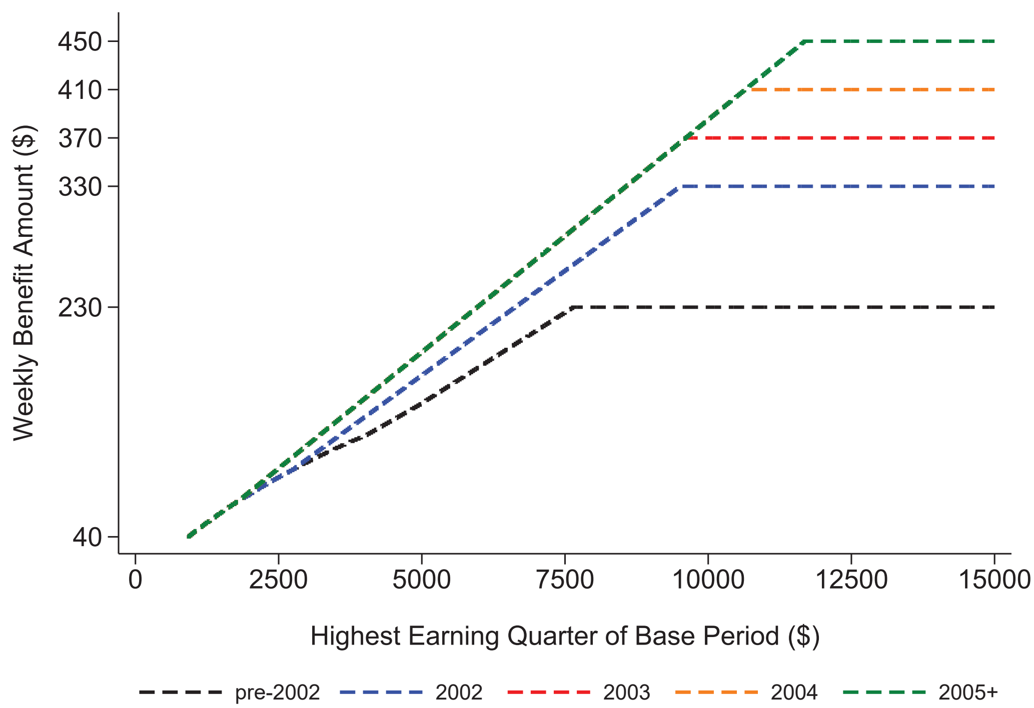
1.10 Figures and Tables

Figure 1.1: Base Periods by Quarter of Claim



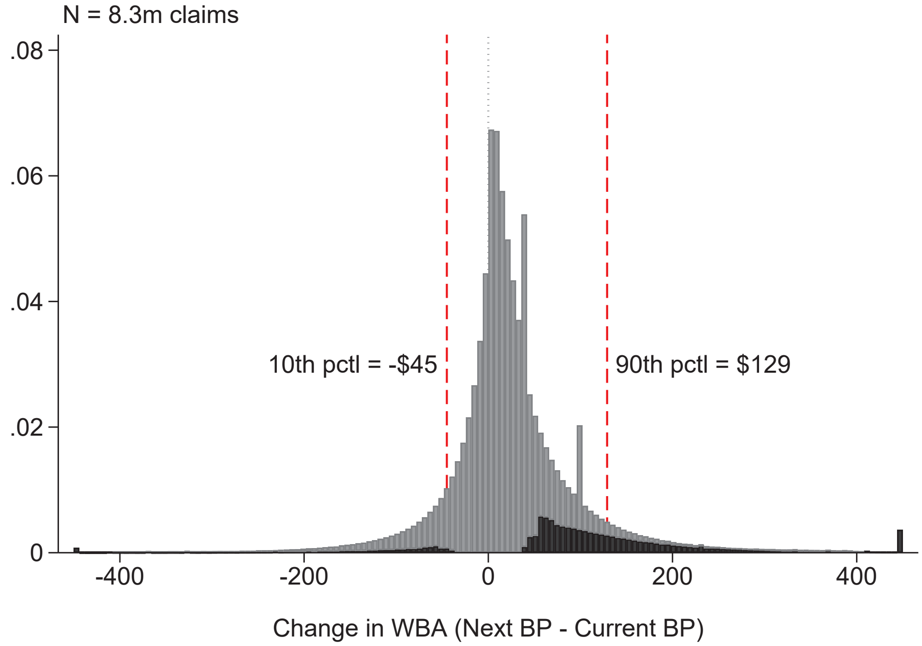
Notes: WBAs and PBDs are calculated as functions of earnings in the base period (BP). Base periods are continuous periods of four completed calendar quarters, determined based on the claim date as shown.

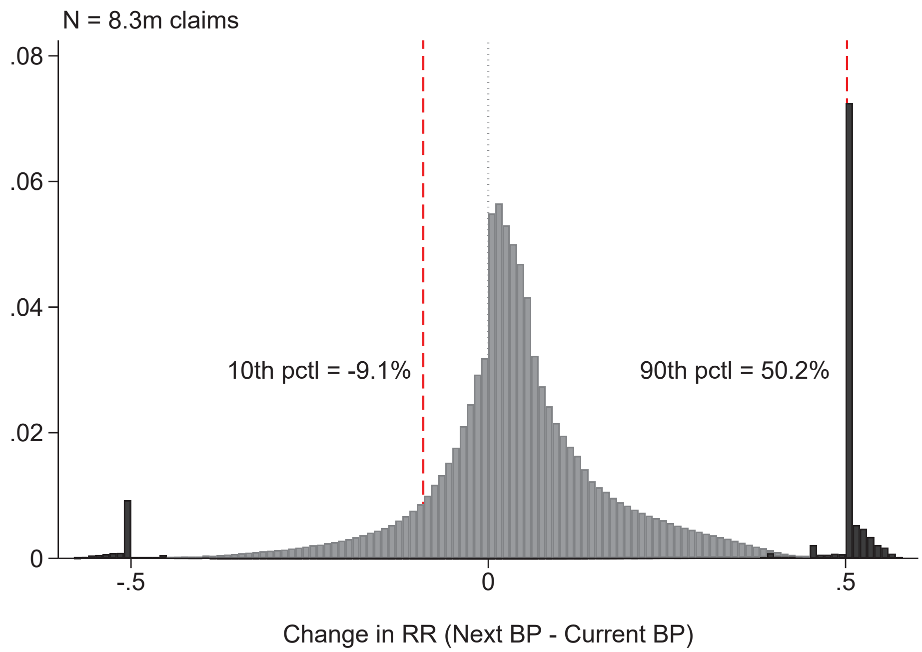
Figure 1.2: WBA Schedule in California



Notes: Beginning with the 2002 schedule, each WBA schedule is effective for new claims made on or after January 1st of the relevant year.

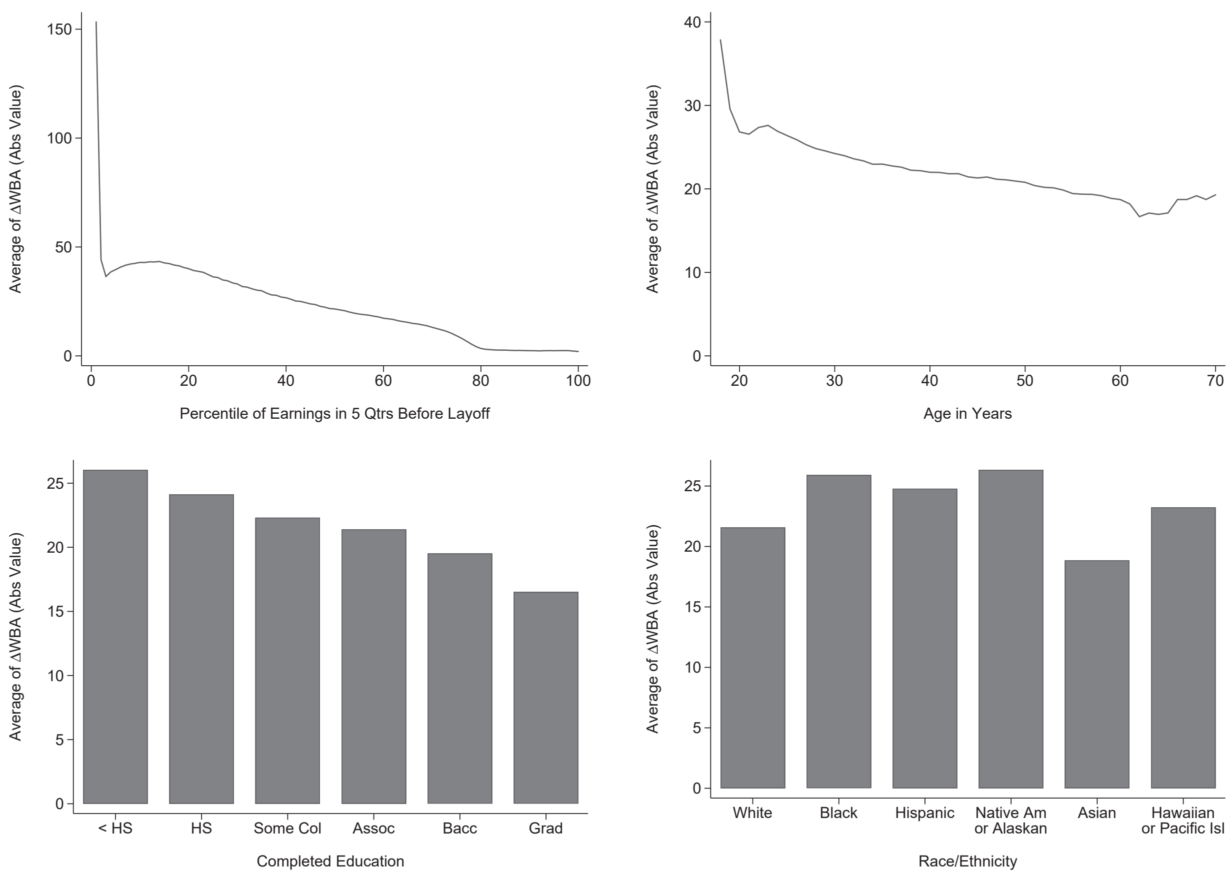
Figure 1.3: Distribution of WBA Change With New Base Period





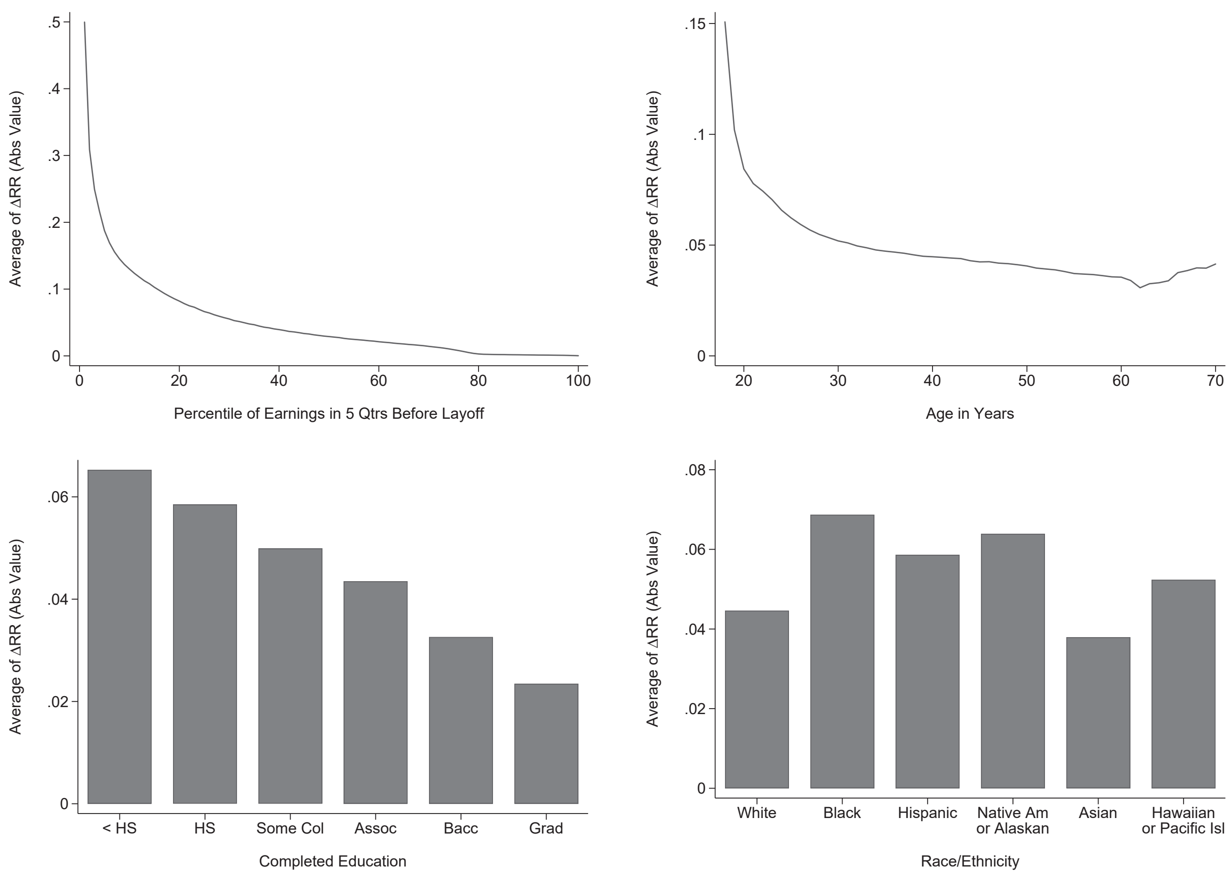
Notes: Histograms of $\Delta WBA = WBA_2 - WBA_1$ and $\Delta RR = RR_2 - RR_1$. Where subscripts denote the BP if the claim is filed in the same quarter as the layoff (1) and the quarter after the layoff (2). Analysis sample limited to claimants with ΔWBA (or RR) $\neq 0$. In each bin, claimants that lose eligibility entirely in one of the two BPs are shaded black while claimants who are eligible in both BPs, but for different benefit amounts, are shaded gray (i.e., the histograms are "stacked" and not overlaid).

Figure 1.4: Heterogeneous Exposure to Benefit Risk



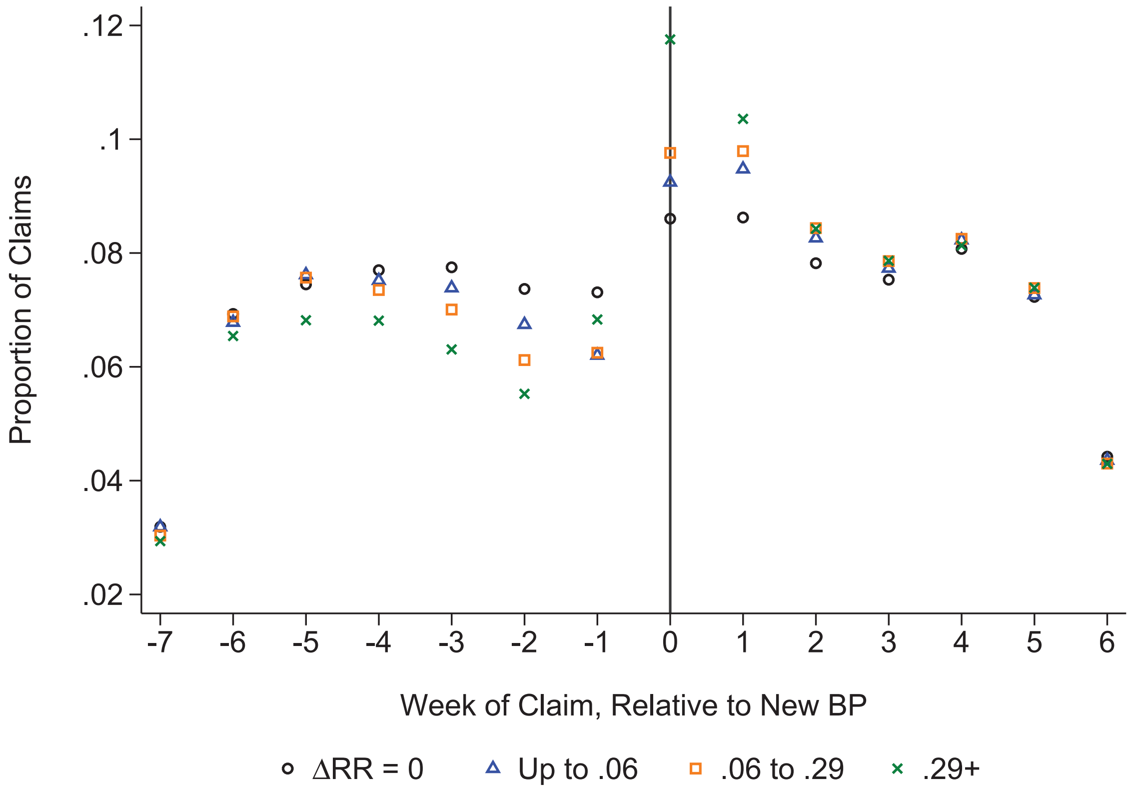
Notes: Each panel displays the mean value of $abs(\Delta WBA)$ in bins defined by the x-axis in the full sample of 22.2 million claims. Completed education, ethnicity, and date of birth are self-reported by the claimant to EDD when the claim is filed. Age is calculated as of the date of the layoff.

Figure 1.5: Heterogeneous Exposure to Benefit Risk



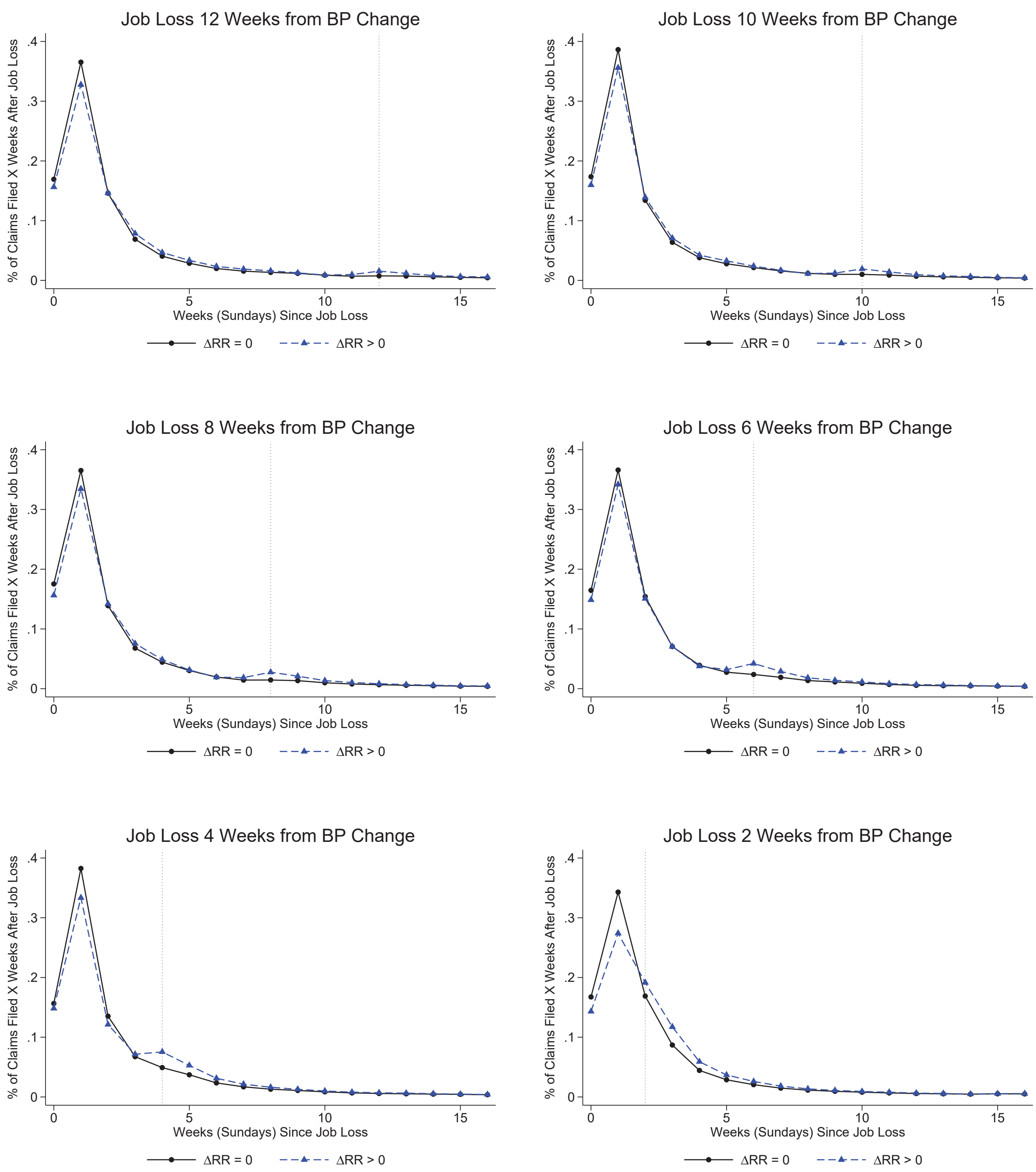
Notes: Each panel displays the mean value of $abs(\Delta RR)$ in bins defined by the x-axis in the full sample of 22.2 million claims. Completed education, ethnicity, and date of birth are self-reported by the claimant to EDD when the claim is filed. Age is calculated as of the date of the layoff.

Figure 1.6: Bunching at Preferred Claim Dates



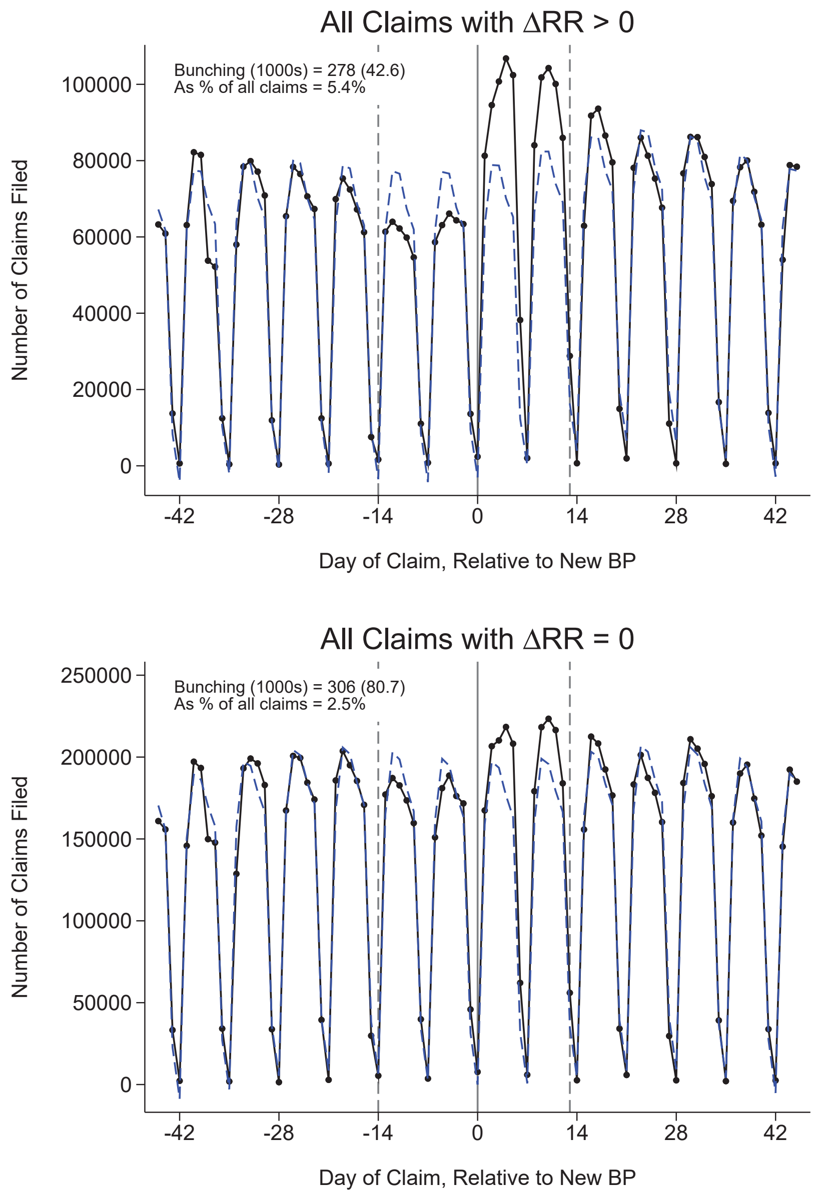
Notes: Bins are claim dates, relative to the closest BP change. For example, week -1 is the week before the BP changes (week ending on the Saturday before the first Sunday of a quarter).

Figure 1.7: Bunching at Preferred Claim Dates, by Layoff Date Relative to BP Change



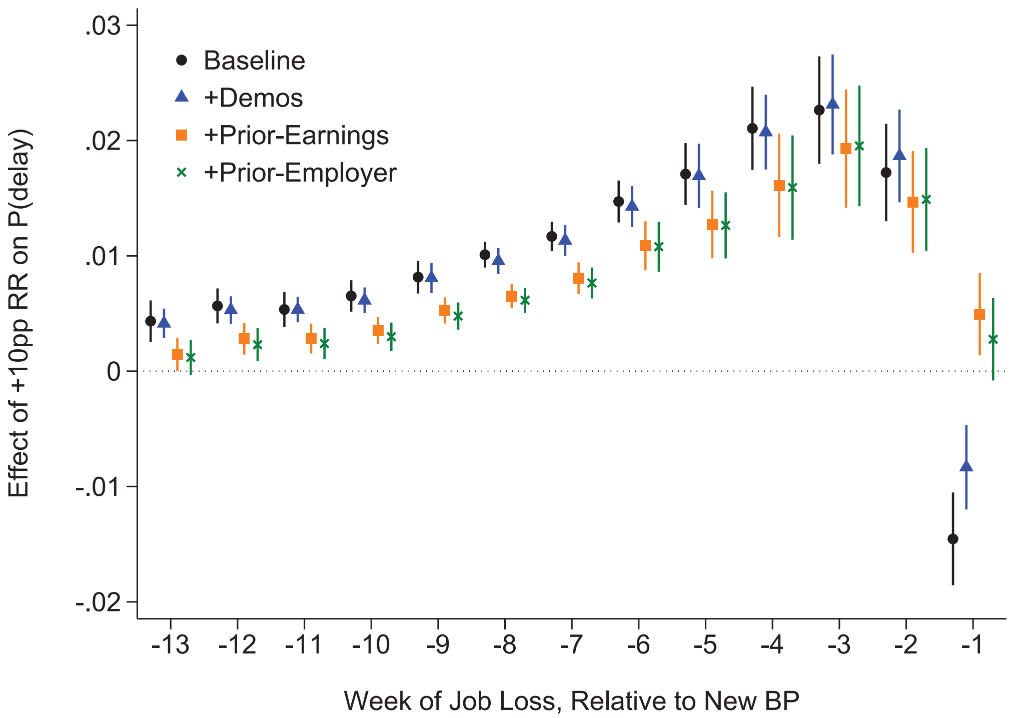
Notes: Each panel shows distributions of the number of weeks between layoff and claim filing with (blue) and without (black) incentives to delay their claim until the BP change.

Figure 1.8: Bunching Estimates



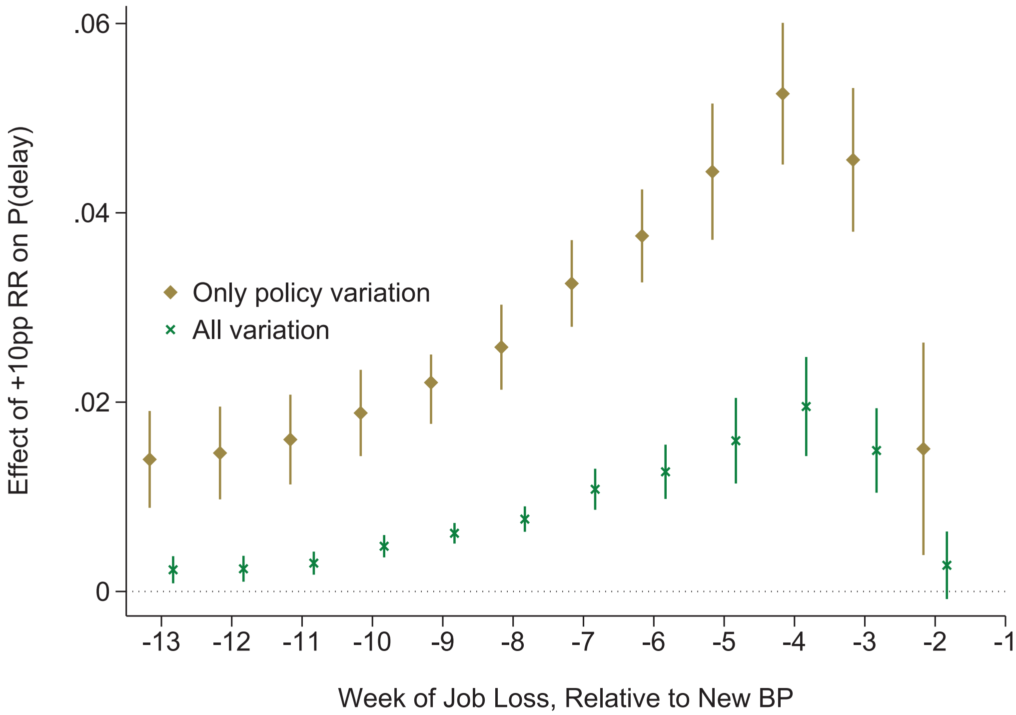
Notes: Panels show distributions of claim dates, centered at the first day of the "new" BP (first Sunday of the quarter after the claimant's last worked date). Black solid lines represent actual distributions, blue dashed lines represent counterfactual distributions which are estimated as described in section 1.5.2. Vertical dashed lines represent the excluded region used to estimate the counterfactual distribution. Estimates of the number (with bootstrapped standard errors) and percent of claims that are delayed are shown in the top left of the figure.

Figure 1.9: Effects of ΔRR on P(claim filed in first week of next qtr)



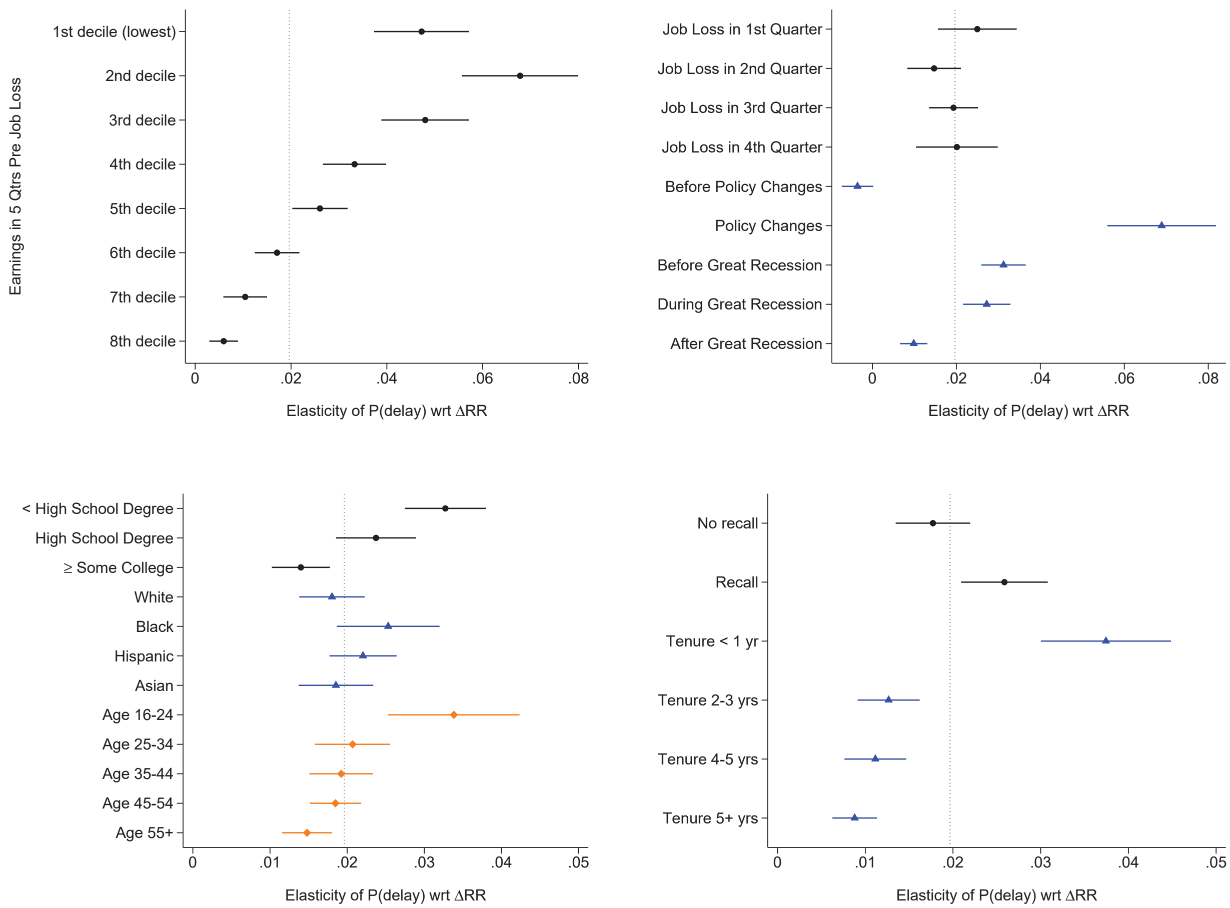
Notes: Estimates of the marginal effect of an extra 10pp in ΔRR on the probability of filing a claim in the first week of the quarter after the layoff. (i.e., $\hat{\beta}_2^{\tau} + \hat{\beta}_3$ from equation 1.5.) Colors denote separate models, estimated with sequentially more complete sets of controls. Each model is estimated via OLS with cluster-robust standard errors at the layoff-quarter level. Each model includes quarter and weekday of layoff FEs, week of layoff (relative to the BP change) dummies, and a control for ΔPBD. "Demos" includes completed education, gender, age, ethnicity, citizenship status, and 3 digit zip code. "Earnings" includes average quarterly earnings totals in the 5 calendar quarters that span the two possible BPs and a measure of "effective" earnings volatility in the same period as described in section 1.5.3. "Pre-separation Employer" refers to the separating employer and includes the reason for job loss, an indicator for whether a recall is expected, firm size, average quarterly earnings of coworkers during the layoff quarter, and sector (two-digit NAICS).

Figure 1.10: Effects of ΔRR on P(claim filed in first week of next qtr), Generalized Difference-in-Difference vs. Selection on Observables



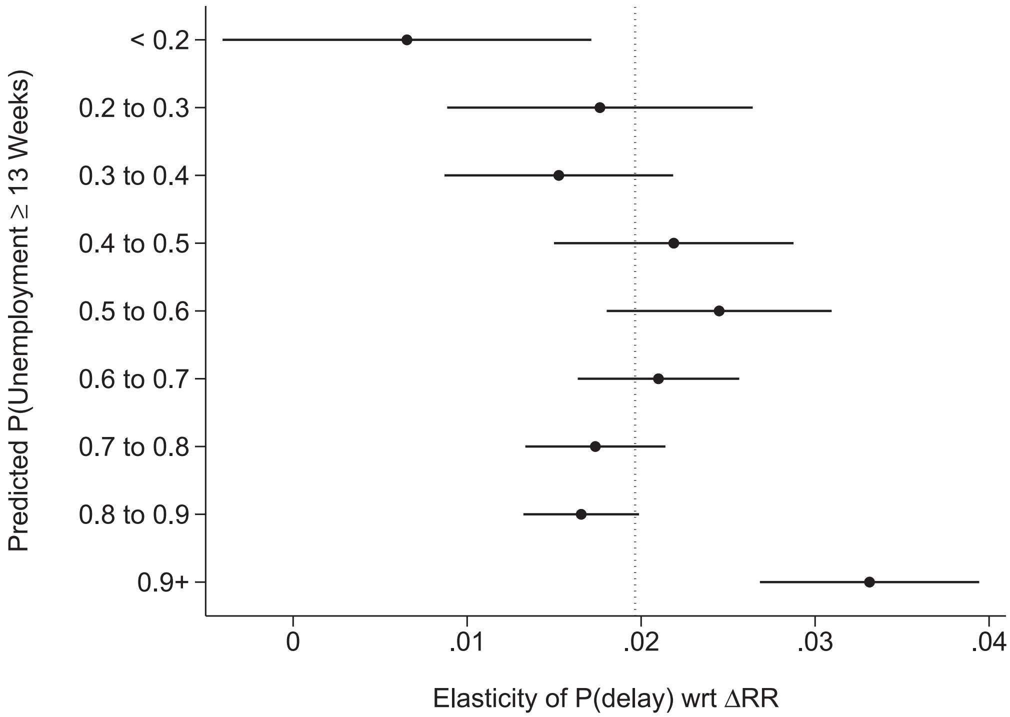
Notes: Estimates of the marginal effect of an extra 10pp in ΔRR on the probability of filing a claim in the first week of the quarter after the layoff, i.e., $\hat{\beta}_2^\tau + \hat{\beta}_3$ from equations 1.5 (selection-on-observables) and 1.6 (difference-in-differences). The difference-in-difference model is estimated in the subset of claims filed by workers laid off between 1/1/2000 and 12/31/2006. Both models are estimated via OLS with cluster-robust standard errors at the layoff-quarter level. Each model includes the most complete set of controls.

Figure 1.11: Heterogeneity in the Effect of ΔRR on P(claim filed in first week of next qtr)



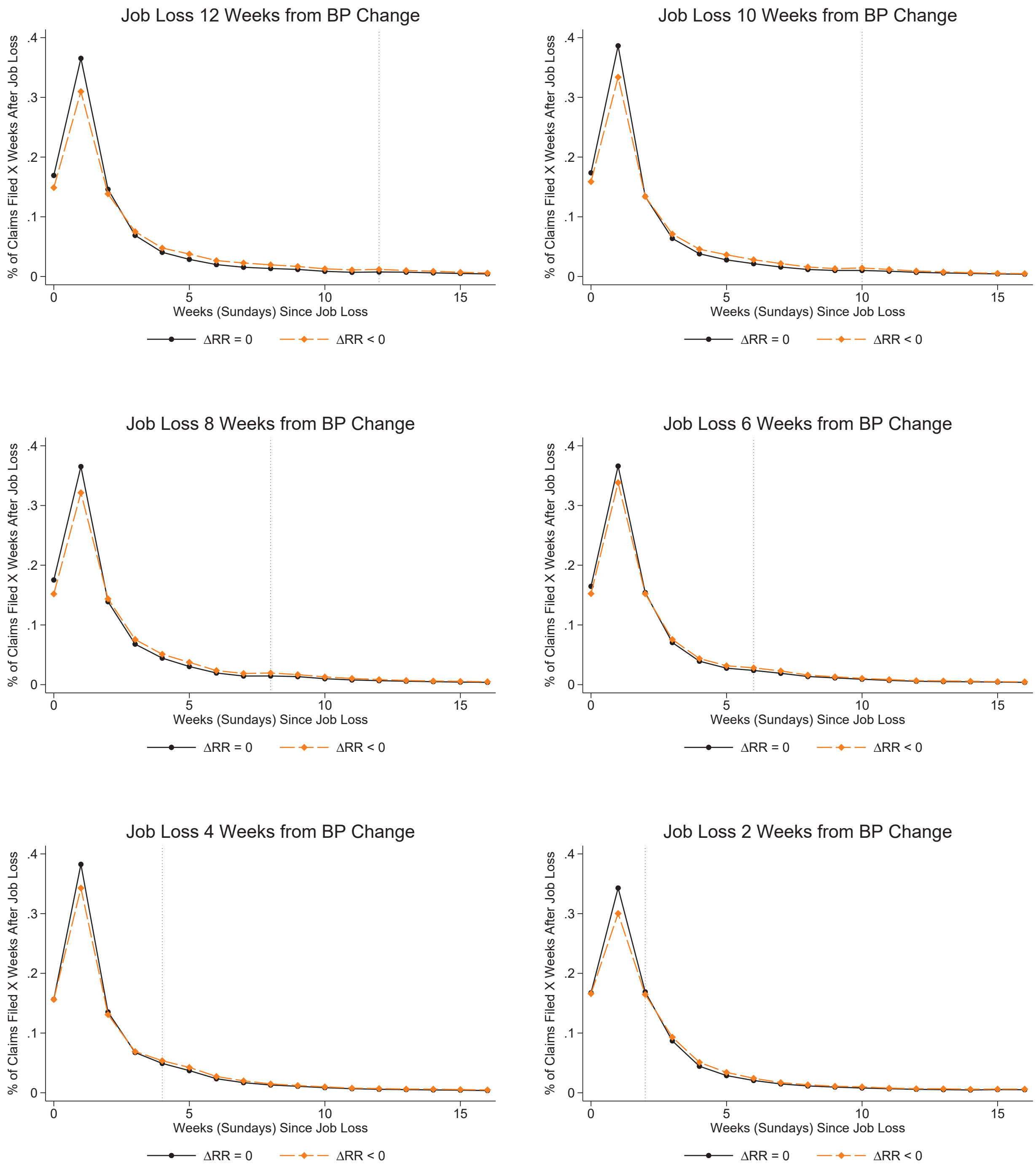
Notes: Estimates of the marginal effect of the elasticity of claim delay (here measured as the probability of filing a claim in the first week of the quarter after the layoff) with respect to Δ RR. The elasticity is calculated as $\hat{\beta}_2$ from equation 1.8 multiplied by $\overline{\Delta RR}/\overline{delay}$ so that the reported values are interpretable as elasticities evaluated at the means. This ensures that variation in the effect of ΔRR on claim delay across subgroups is isolated from variation in the distribution of ΔRR and claim delay across subgroups. Standard errors are calculated via the delta method. Each row is coefficient from single model, limited to a subsample of claimants. All models are limited to claimants with $\Delta RR \geq 0$ and job losses occurring in the last 5 weeks of a BP. For reference, each panel includes a dotted vertical line which corresponds to the point estimate of the same elasticity among all claimants with $\Delta RR \geq 0$ and job losses occurring in the last 5 weeks of a BP.

Figure 1.12: Heterogeneity in the Effect of ΔRR on P(claim filed in first week of next qtr) by Predicted Unemployment Duration



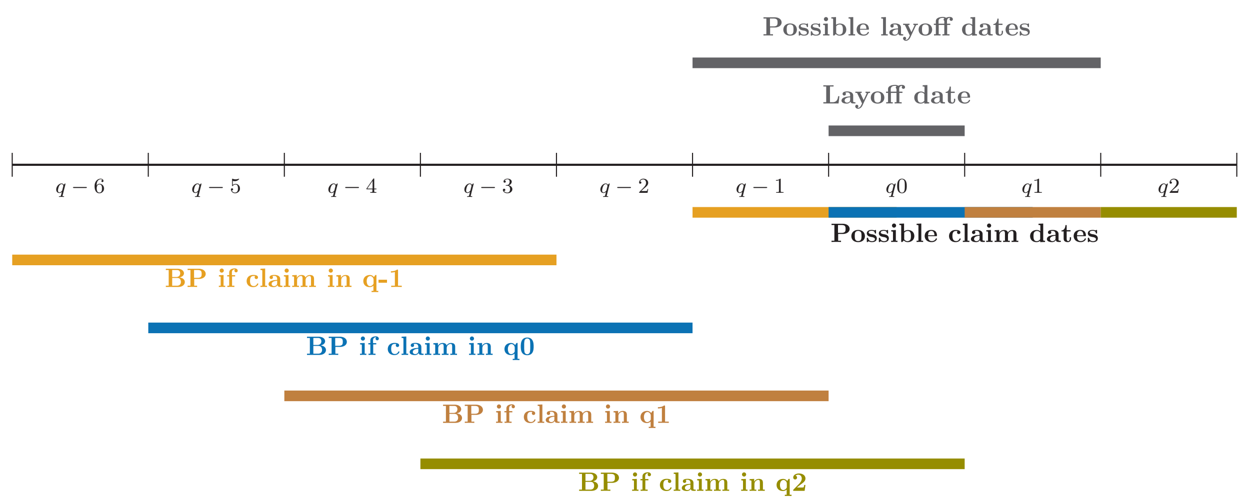
Notes: Estimates of the marginal effect of the elasticity of claim delay (here measured as the probability of filing a claim in the first week of the quarter after the layoff) with respect to Δ RR. The elasticity is calculated as $\hat{\beta}_2$ from equation 1.8 multiplied by $\overline{\Delta RR}/\overline{delay}$ so that the reported values are interpretable as elasticities evaluated at the means. This ensures that variation in the effect of ΔRR on claim delay across subgroups is isolated from variation in the distribution of ΔRR and claim delay across subgroups. Standard errors are calculated via the delta method. Each row is coefficient from single model, limited to a subsample of claimants defined by predicted unemployment duration (probability of a 13+ week spell), predicted as described in section 1.6.2. All models are limited to claimants with $\Delta RR \geq 0$ and job losses occurring in the last 5 weeks of a BP. For reference, each panel includes a dotted vertical line which corresponds to the point estimate of the same elasticity among all claimants with $\Delta RR \geq 0$ and job losses occurring in the last 5 weeks of a BP.

Figure 1.13: Claimants do Not “Hurry Up”



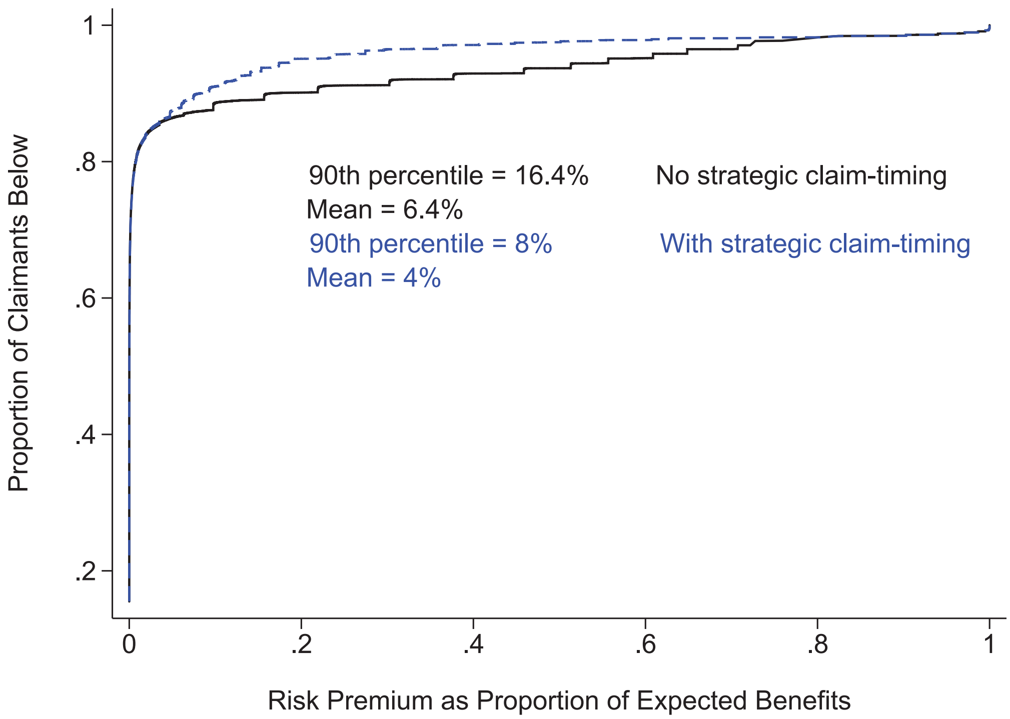
Notes: Each panel shows distributions of the number of weeks between layoff and claim-filing with (blue) and without (black) incentives to expedite their claim until the BP change.

Figure 1.14: Illustration of Potential Base Periods Considered in Risk Premia Calculations



Notes: Each claimant's risk premium is based on benefit eligibility under four possible base periods, and the probability that each of those base periods occurs. This figure shows an example of those four base periods using a timeline. The starting point is a claimant's actual layoff date ($q0$). I then allow for variability in week of layoff using an assumed distribution. (Binomial w/ 26 trials and prob success = 0.5 to ensure that the possible layoff weeks are +/- 1 qtr from the actual and roughly normally distributed). Next, to determine the probability that the claimant waits X weeks before claiming, I use the empirical distribution for full sample. Together these two distributions imply 4 potential BPs as shown. Finally I calculate benefit eligibility in each BP, filling in missing earnings info in the layoff quarter by extrapolation (assume avg weekly earnings in qtr paid for remainder of qtr).

Figure 1.15: Cumulative Distribution of Risk Premia



Notes: Figure displays cumulative distributions of risk premia calculcated as described in section 1.8. Blue distribution (dashed line) is calculated with strategic claim-timing by allowing the distribution of time between job loss and claiming to vary with with ΔRR. Black distribution (solid line) is calculated without strategic claim-timing.

Table 1.1: Examples of Earnings Volatility Driven Benefit Risk

Claimant	$q-5$	$q-4$	$q-3$	$q-2$	$q-1$	q (layoff qtr)	WBA_1	WBA_2
1	$10k	$10k	$10k	$10k	$10k	2005q3	$385	$385
2	$10k	$10k	$10k	$10k	$15k	2005q3	$385	$450
3	$10k	$10k	$10k	$15k	$10k	2005q3	$450	$450
4	$15k	$15k	$15k	$15k	$20k	2005q3	$450	$450
5	$15k	$15k	$15k	$15k	$15k	2004q3	$410	$410
6	$15k	$15k	$15k	$15k	$15k	2004q4	$410	$450

Notes: Table displays hypothetical earnings histories chosen to demonstrate various sources of claim-timing incentives. For each claimant the table shows 5 quarterly earnings amounts for the 5 *completed* calendar quarters precieding the quarter of the layoff, WBA_1 (the WBA received by the claimant if the claim is filed in the layoff quarter), WBA_2 (WBA if the claim is filed in the quarter after the layoff).

Table 1.2: Claimant Demographics

	All mean/sd	$\Delta WBA = 0$ mean/sd	$\Delta WBA > 0$ mean/sd	$\Delta WBA < 0$ mean/sd
Age	38.60	39.76	35.91	38.64
	(13.09)	(12.94)	(12.92)	(13.33)
Female	0.45	0.43	0.47	0.48
	(0.50)	(0.50)	(0.50)	(0.50)
White	0.36	0.39	0.32	0.30
	(0.48)	(0.49)	(0.47)	(0.46)
Black	0.09	0.08	0.10	0.09
	(0.28)	(0.27)	(0.30)	(0.28)
Hispanic	0.37	0.34	0.41	0.44
	(0.48)	(0.47)	(0.49)	(0.50)
< HS diploma	0.19	0.16	0.23	0.25
	(0.39)	(0.37)	(0.42)	(0.43)
Completed Education:				
HS diploma	0.30	0.29	0.34	0.33
	(0.46)	(0.45)	(0.47)	(0.47)
Some college	0.27	0.27	0.27	0.27
	(0.44)	(0.44)	(0.44)	(0.44)
Associates	0.06	0.06	0.05	0.05
	(0.23)	(0.24)	(0.22)	(0.22)
Bachelors	0.13	0.16	0.09	0.08
	(0.34)	(0.37)	(0.29)	(0.27)
> Bachelors	0.04	0.06	0.02	0.02
	(0.21)	(0.24)	(0.15)	(0.13)
N	22,211,577	13,880,919	6,035,894	2,294,764

Notes: ΔWBA is defined as $WBA_2 - WBA_1$, where subscripts denote BP if claim is filed in quarter of layoff (1) or quarter after layoff (2). Completed education, ethnicity, and date of birth are self-reported by the claimant to EDD when the claim is filed. Age is calculated as of the date of the layoff.

Table 1.3: Claimant Earnings, Benefits, and Claim Timing

	All mean/sd	$\Delta WBA = 0$ mean/sd	$\Delta WBA > 0$ mean/sd	$\Delta WBA < 0$ mean/sd
In 5 completed qtrs pre-layoff:				
Avg. qtrly earnings	8143.82	10380.38	4352.75	4586.63
	(52804.79)	(66330.70)	(10440.70)	(2772.96)
SD qtrly earnings	2884.70	3365.21	2126.56	1972.25
	(115799.36)	(145814.89)	(21104.72)	(2389.71)
If claim filed in qtr of layoff:				
Max benefit amount (MBA)	6865.74	8044.36	4358.23	6331.74
	(3759.76)	(3539.60)	(3174.84)	(3049.48)
Wkly benefit amount (WBA)	274.36	317.69	180.57	258.98
	(139.51)	(129.81)	(121.09)	(114.81)
Benefit duration (PBD)	23.50	24.86	20.15	24.13
	(5.90)	(3.23)	(9.11)	(3.72)
Prior job:				
Quit	0.04	0.04	0.04	0.04
	(0.20)	(0.20)	(0.19)	(0.20)
Fired	0.12	0.13	0.11	0.11
	(0.33)	(0.33)	(0.32)	(0.31)
Recall expected	0.27	0.25	0.29	0.32
	(0.44)	(0.43)	(0.46)	(0.46)
# Employees	3784.67	3766.37	3837.36	3752.52
	(11615.44)	(11687.40)	(11485.86)	(11517.55)
# Establishments	52.06	53.58	49.06	50.85
	(233.02)	(241.77)	(217.04)	(219.12)
Avg. Quarterly Pay	9647.14	10981.86	7592.65	6910.06
	(22932.79)	(25625.14)	(18609.72)	(12866.52)
Claim-timing:				
Claimed in next qtr	0.28	0.27	0.33	0.28
	(0.45)	(0.44)	(0.47)	(0.45)
Claimed in 1st wk of next qtr	0.07	0.07	0.08	0.06
	(0.25)	(0.25)	(0.27)	(0.24)
Days btwn layoff and claim	35.21	33.30	38.65	37.72
	(67.84)	(66.94)	(68.79)	(70.23)
N	22,211,577	13,880,919	6,035,894	2,294,764

Notes: ΔWBA is defined as $WBA_2 - WBA_1$, where subscripts denote BP if claim is filed in quarter of layoff (1) or quarter after layoff (2).

Table 1.4: Claimant Prior Job Industry by Change in WBA

	All	$\Delta WBA = 0$	$\Delta WBA > 0$	$\Delta WBA < 0$
	mean	mean	mean	mean
Ag, Forestry, Fishing, Hunting	0.06	0.04	0.08	0.09
Mining, Quarrying, Oil/Gas Extr	0.00	0.00	0.00	0.00
Utilities	0.00	0.00	0.00	0.00
Construction	0.10	0.11	0.10	0.08
Manufacturing	0.10	0.11	0.08	0.09
Wholesale Trade	0.04	0.05	0.04	0.04
Retail Trade	0.11	0.10	0.13	0.13
Transportation, Warehousing	0.03	0.03	0.03	0.04
Information	0.04	0.05	0.03	0.02
Finance, Insurance	0.04	0.04	0.03	0.03
Real Estate, Rental/Leasing	0.02	0.02	0.02	0.02
Professional, Sci, Tech	0.07	0.09	0.06	0.05
Mgmt of Companies/Enterprises	0.01	0.01	0.01	0.01
Admin Support, Waste Mgmt, Remed	0.11	0.10	0.14	0.13
Educational Services	0.05	0.05	0.04	0.04
Health Care, Social Assistance	0.08	0.08	0.08	0.09
Arts, Entertainment, Recreation	0.02	0.02	0.02	0.02
Accomodation, Food Services	0.07	0.06	0.08	0.09
Other Services	0.04	0.04	0.04	0.04
Public Admin	0.01	0.01	0.01	0.01
N	20,627,613	12,904,569	5,707,664	2,015,380

Notes: ΔWBA is defined as $WBA_2 - WBA_1$, where subscripts denote BP if claim is filed in quarter of layoff (1) or quarter after layoff (2). Industry groups are equivalent to 2-digit NAICS codes of the separating employer.

Table 1.5: Quarter Change Option Take-up

	Offered mean/sd	Did not delay mean/sd	Delayed mean/sd
Female	0.47	0.47	0.47
	(0.50)	(0.50)	(0.50)
Age (on layoff date)	36.36	35.87	37.14
	(12.94)	(12.91)	(12.95)
White	0.32	0.32	0.32
	(0.47)	(0.47)	(0.47)
Completed Education:			
< HS	0.22	0.22	0.23
	(0.42)	(0.41)	(0.42)
HS	0.33	0.32	0.33
	(0.47)	(0.47)	(0.47)
> HS	0.41	0.41	0.40
	(0.49)	(0.49)	(0.49)
Tenure at separating employer:			
≤ 1 yr	0.49	0.48	0.52
	(0.50)	(0.50)	(0.50)
2-3 yrs	0.24	0.25	0.23
	(0.43)	(0.43)	(0.42)
4-5 yrs	0.08	0.09	0.08
	(0.28)	(0.28)	(0.27)
> 5 yrs	0.14	0.14	0.14
	(0.34)	(0.34)	(0.35)
WBA difference ($)	42.22	31.45	59.27
	(59.28)	(50.38)	(67.71)
RR difference	0.08	0.06	0.10
	(0.09)	(0.08)	(0.10)
N	398,546	244,265	154,281

Notes: Sample limited to claimants who file their claim in the last week of a BP, but would have seen their benefits increase (by any amount) if they had instead waited just one more week. These claimants are notified by the agency that their benefits would have been different if they had delayed their claim one more week, made aware of the exact change in benefits they would have been eligible for, and given the opportunity to revisit their decision (i.e. to delay their claim by one week after it had already been filed). The second column includes only those claimants who did not accept this option, the third column includes only those claimants who did not.

Appendix

1.A Additional Institutional Context

PBDs are determined by the following formula:

$$PBD = \begin{cases} 26 & \text{if } WBA \cdot 26 \leq BPW \cdot \frac{1}{2} \\ \frac{BPW \cdot \frac{1}{2}}{WBA} & \text{if } WBA \cdot 26 > BPW \cdot \frac{1}{2} \end{cases}$$

The PBD never exceeds 26 weeks, and the MBA never exceeds one half of the BPW. To further clarify the PBD formula, we can plug in the WBA formula *for the case where the WBA is less than the maximum*. This allows the PBD formula to be expressed as a function of only the HQW and the BPW:

$$PBD = \begin{cases} 26 & \text{if } 4 \cdot RR \leq \frac{BPW}{HQW} \\ \frac{BPW \cdot 13}{HQW \cdot 2 \cdot RR} & \text{if } 4 \cdot RR > \frac{BPW}{HQW} \end{cases}$$

Again, this appears complex, but it makes clear that the PBD is a function of the ratio $\frac{BPW}{HQW}$ and that (since this ratio is bounded below by 1) the PBD is bounded below by $\frac{13}{2 \cdot RR}$ (e.g. 13 weeks if the RR is 0.5). This formulation also makes clear that a PBD below 26 weeks is only assigned to claimants whose HQW accounts for too large a proportion of their total BP earnings.

1.B Generalized Difference-in-Differences and Event Study Specifications

Several factors could explain the differences between the results in section 1.5.4 and section 1.5.3. First, the sample is different—limited to the years around the policy changes in the diff-in-diff approach and spanning all 18 years of the full analysis sample in the selection on observables approach. Second, the differences may reflect remaining unobserved confounders that are biasing the estimates in Figure 1.9. Third, the discrepancies may reflect differences in the characteristics of the claimants who have large ΔWBA values. Regressions in Figure 1.9 rely on variation in ΔWBA driven by earnings volatility conditional on a series of controls. As shown in, e.g., Table 1.2 and Table 1.3, claimants with large ΔWBA values due to earnings volatility tend to be broadly less advantaged than those with ΔWBA values close to zero. On the other hand, as shown in Table 1.A1 (and explained further in the next subsection), many of the claimants with policy-driven variation in ΔWBA (which is the variation isolated by the diff-in-diff estimates in Figure 1.10) have little to no earnings volatility and are therefore likely a more advantaged group. These claimants may have an easier time responding to these incentives for a variety of reasons. Investigating this potential source of heterogeneity in the effects of ΔWBA on claim-timing decisions will be an important next step in this analysis.

The key assumption underlying specifications in section 1.5.4 is that outcomes in differentially exposed earnings groups would have evolved in parallel in the absence of the policy change. To provide suggestive evidence I test for pre-trends using event study specifications estimated separately for each of the four policy changes:

$$delay_{gq} = \sum_{q \in Q} \beta^q 1\{qtr = q\} \cdot \Delta WBA_g^Q + \gamma_g + \psi_q + \epsilon_{gq} \tag{1.8}$$

Where Q denotes a set of quarters around the policy change of interest, and ΔWBA_g^Q is the "treatment" received by group g in the relevant policy change quarter. As an example consider the policy change which occurs in 2001q4. The set of quarters Q included in the regression are 2000q1-2002q3 (the data begins in 2000q1, a second policy change occurs in 2002q4), the omitted quarter in the set of interactions with ΔWBA_g^Q is 2001q3, the policy change quarter of interest is 2001q4, and ΔWBA_g^Q is defined as the value of ΔWBA for the claimant's earnings group g in 2001q4 (e.g., if $HQW_1 = HQW_2 = \$10,000$, $\Delta WBA_g^Q = \$100$ as per Table 1.A1).

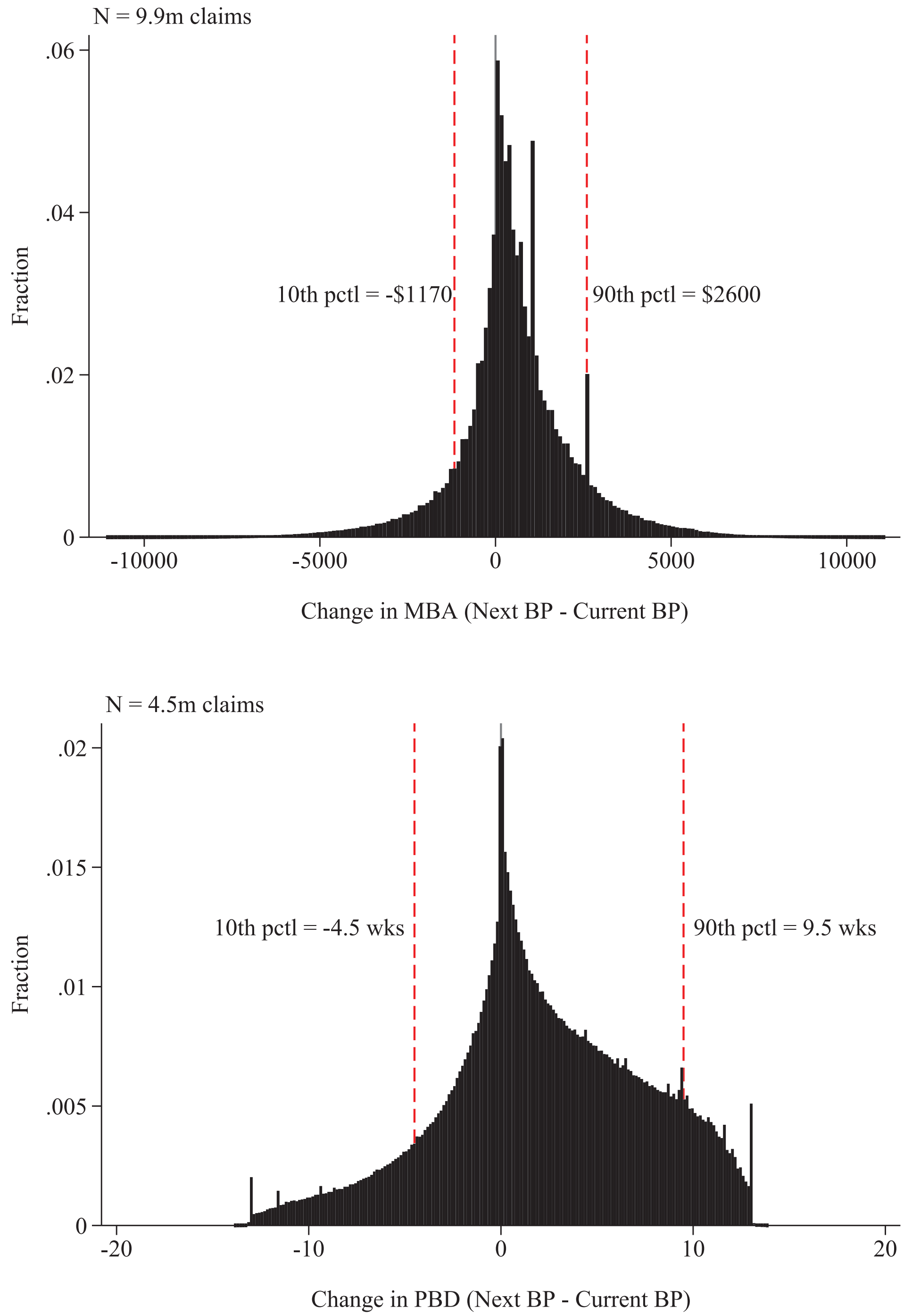
Several characteristics of this setting complicate these specifications. First, this is an event study setting with multiple treatments per treated unit. Second, the effects of the policy changes on claim-timing incentives are not permanent. Table 1.A1 demonstrates that for some claimants the effect of the policy change persists after the policy change quarter but at a different level (Claimants 2 and 3 in the table). This occurs for claimants who would have claim-timing incentives without the policy change. In the quarter of the policy change these incentives are altered because waiting to claim implies a new WBA schedule. In the quarter after the policy change these incentives are altered again because the new WBA schedule is applied to both potential BPs. For other claimants (like Claimant 1 in the table) the policy change is effective only within the policy change quarter and then immediately "turns off" in the quarter after. This is because these claimants have earnings histories such that they do not typically have an incentive to delay or speed up claiming. The incentive only exists in the policy change quarter because waiting for the next BP leads to a change in the effective WBA schedule.

These two different types of claimants need to be handled differently by the event study specifications. For claimants 2 and 3, it would not make sense to expect to observe no pre-trends in the event studies for the latter policy changes. This is because the "pre" period for the 2nd-4th policy changes are also the "post" period for the preceding policy change. Another complication for claimants 2 and 3 is that it does not make sense to interact ΔWBA_g^Q with "post" policy change quarters since a different policy-driven ΔWBA is effective for these claimants in those quarters. In the case of claimant 1, the uniqueness of the setting is actually helpful. Since the policy change immediately turns off in the quarter following the policy change, the existence of multiple treatments per treated unit does not complicate testing for pre-trends. Instead, for these claimants I can provide supporting evidence for the parallel trends assumption not only by testing for pre-trends but also by testing post-trends. For these reasons I limit the event-study estimates to claimants like claimant 1, that is, claimants with no "relevant" earnings volatility in the five quarters that span the two BPs of interest.

Results are shown in Figure 1.A8 and broadly consistent with a conclusion that there are no pre-trends, with a few exceptions. First, there are several quarters in the first pre-period where earnings groups who will later be more affected by the policy change are more likely to delay their claim. Second, this differential delay is also apparent in several quarters after the 2004q1 policy change and one quarter after the last (2005q1) policy change. This is somewhat concerning for the assumptions underlying the generalized difference-in-differences specifications presented in the prior subsection, as it suggests the possibility that parallel trends assumption does not hold in this setting. However, in each specification the coefficient in the policy change quarter is dramatically larger than the others implying that this may not be a major concern. Future work will need to investigate this further, perhaps via placebo tests in non-policy change quarters or by testing the robustness of the basic event study specifications (e.g. to adding controls).

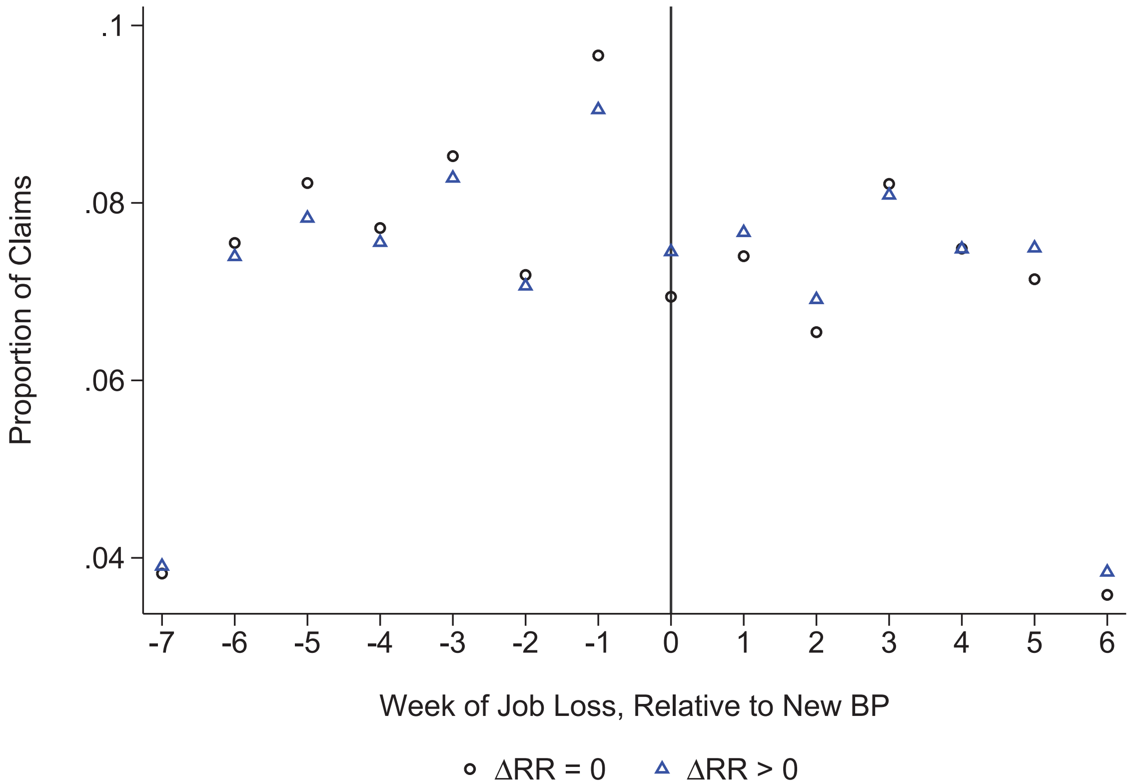
1.C Additional Figures and Tables

Figure 1.A1: Distribution of MBA and PBD Change With New Base Period



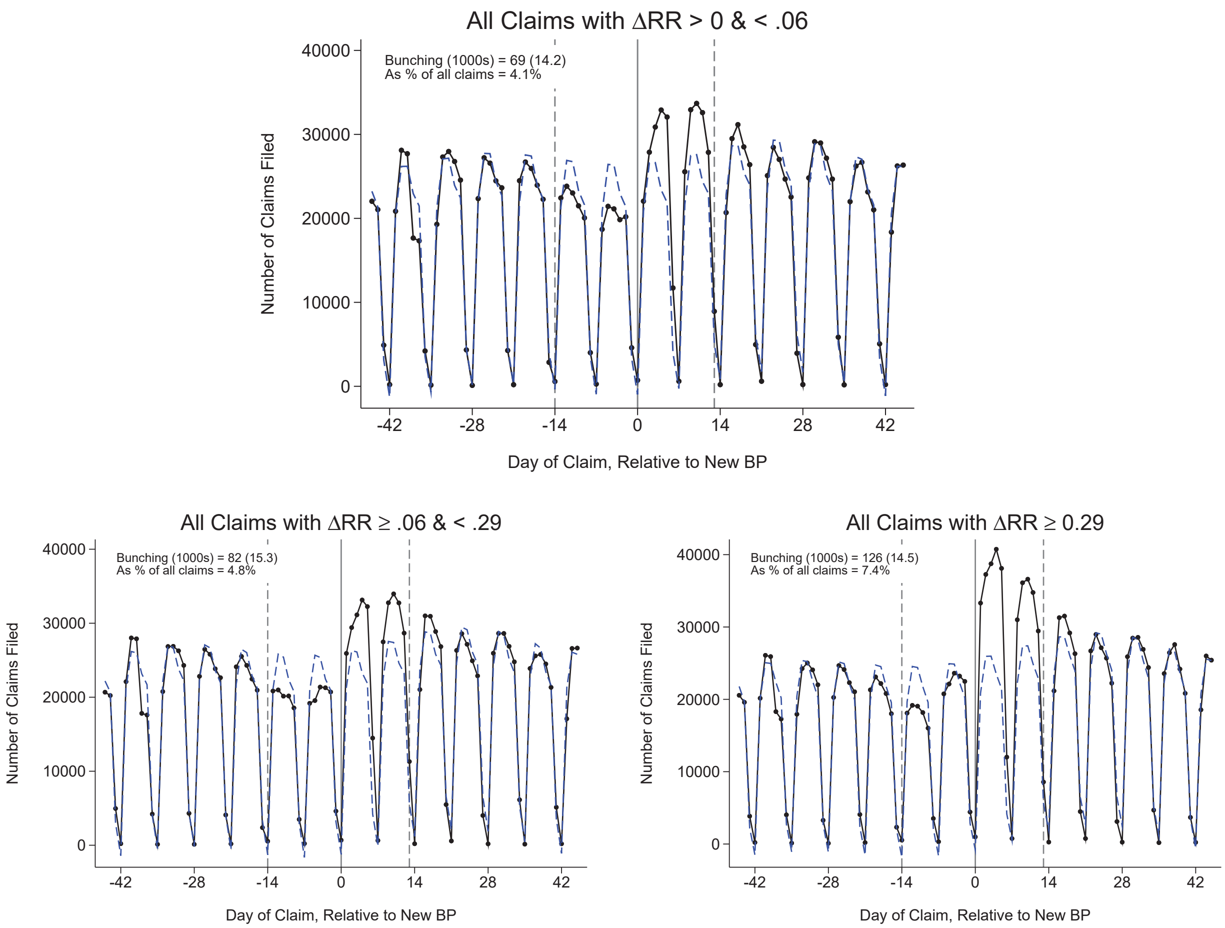
Notes: Histograms of $\Delta MBA = MBA_2 - MBA_1$ and $\Delta PBD = PBD_2 - PBD_1$. Where subscripts denote the BP if the claim is filed in the same quarter as the layoff (1) and the quarter after the layoff (2).

Figure 1.A2: (non) Bunching in layoff date distributions



Notes: Bins are layoff dates, relative to the closest BP change. For example, week -1 is the week before the BP changes (week ending on the Saturday before the first Sunday of a quarter).

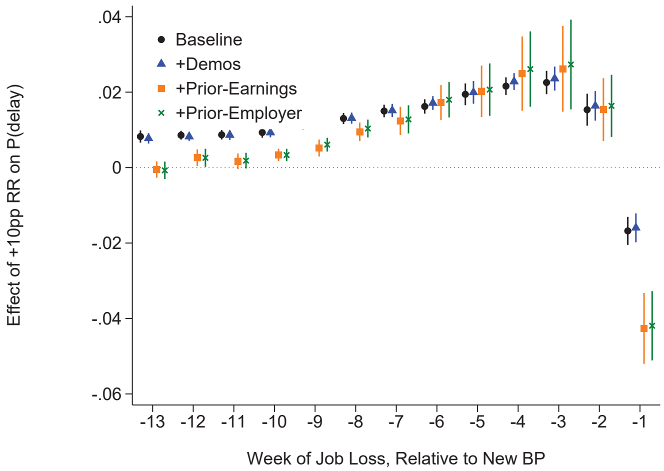
Figure 1.A3: Amount of bunching increases with incentives to bunch



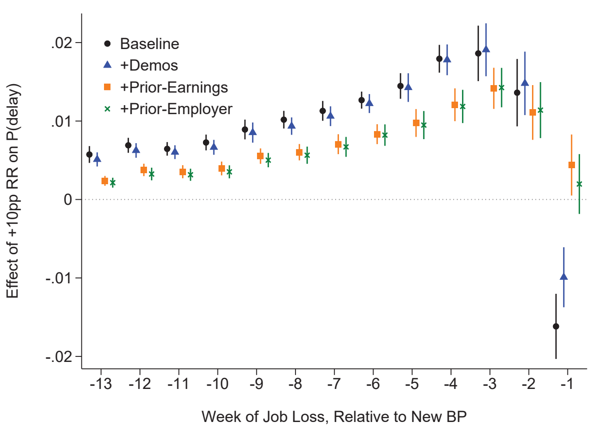
Notes: Panels show distributions of claim dates, centered at the first day of the "new" BP (first Sunday of the quarter after the claimant's last worked date). Black solid lines represent actual distributions, blue dashed lines represent counterfactual distributions which are estimated as described in section 1.5.2. Vertical dashed lines represent the excluded region used to estimate the counterfactual distribution. Estimates of the number (with bootstrapped standard errors) and percent of claims that are delayed are shown in the top left of the figure.

Figure 1.A4: Effects of ΔRR on P(claim filed in first week of next qtr): Alternate Samples

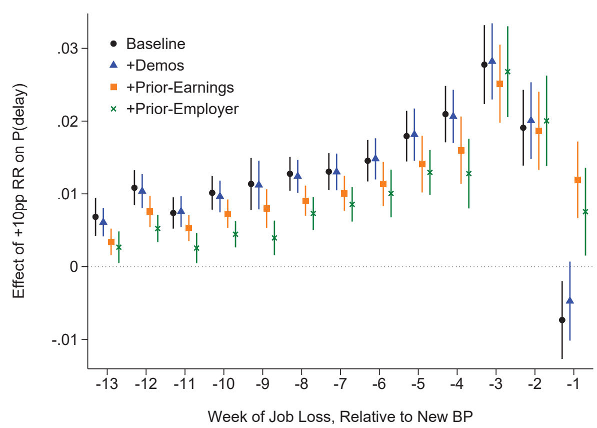
(a) ΔWBA $= 0$ Claims Excluded



(b) Post Policy Changes

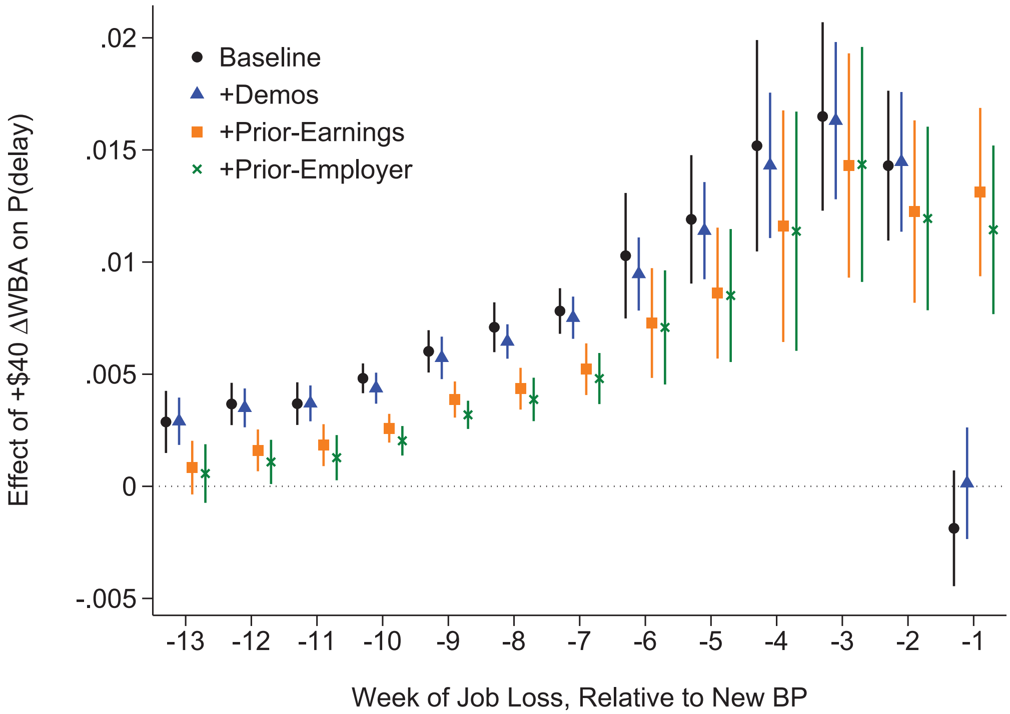


(c) Mass Layoff Sample



Notes: Estimates of the marginal effect of an extra 10pp in ΔRR on the probability of filing a claim in the first week of the quarter after the layoff. (i.e., $\hat{\beta}_2^{\tau} + \hat{\beta}_3$ from equation 1.5.) Colors denote separate models, estimated with sequentially more complete sets of controls. Each model is estimated via OLS with cluster-robust standard errors at the layoff-quarter level. Each model includes quarter and weekday of layoff FEs, week of layoff (relative to the BP change) dummies, and a control for ΔPBD. "Demos" includes completed education, gender, age, ethnicity, citizenship status, and 3 digit zip code. "Earnings" includes average quarterly earnings totals in the 5 calendar quarters that span the two possible BPs and a measure of "effective" earnings volatility in the same period as described in section 1.5.3. "Pre-separation Employer" refers to the separating employer and includes the reason for job loss, an indicator for whether a recall is expected, firm size, average quarterly earnings of coworkers during the layoff quarter, and sector (two-digit NAICS).

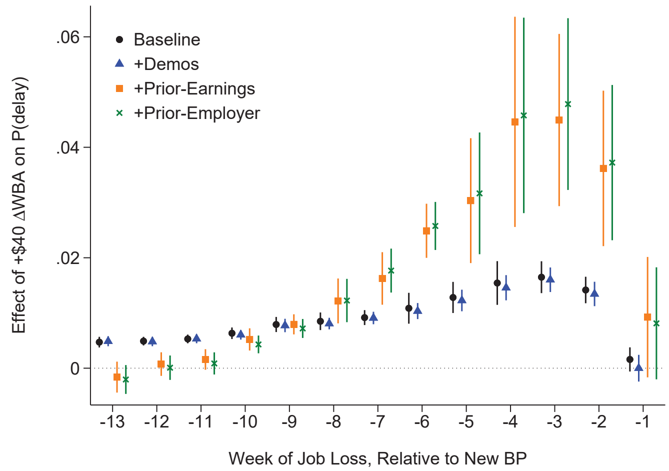
Figure 1.A5: Effects of ΔWBA on P(claim filed in first week of next qtr)



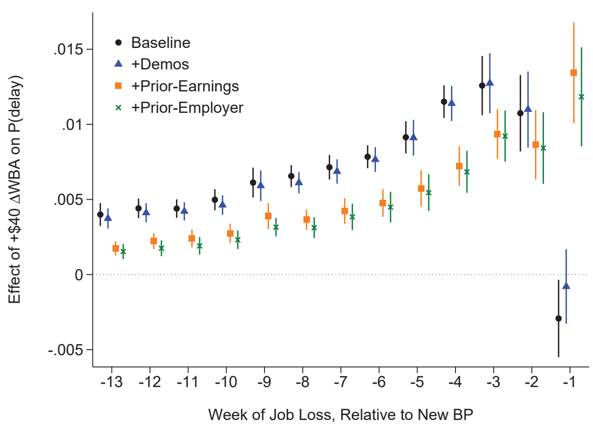
Notes: Estimates of the marginal effect of an extra \$40 in ΔWBA on the probability of filing a claim in the first week of the quarter after the layoff. (i.e., $\hat{\beta}_2^{\tau} + \hat{\beta}_3$ from equation 1.5.) Colors denote separate models, estimated with sequentially more complete sets of controls. Each model is estimated via OLS with cluster-robust standard errors at the layoff-quarter level. Each model includes quarter and weekday of layoff FEs, week of layoff (relative to the BP change) dummies, and a control for ΔPBD. "Demos" includes completed education, gender, age, ethnicity, citizenship status, and 3 digit zip code. "Earnings" includes average quarterly earnings totals in the 5 calendar quarters that span the two possible BPs and a measure of "effective" earnings volatility in the same period as described in section 1.5.3. "Pre-separation Employer" refers to the separating employer and includes the reason for job loss, an indicator for whether a recall is expected, firm size, average quarterly earnings of coworkers during the layoff quarter, and sector (two-digit NAICS).

Figure 1.A6: Effects of ΔWBA on P(claim filed in first week of next qtr): Alternate Samples

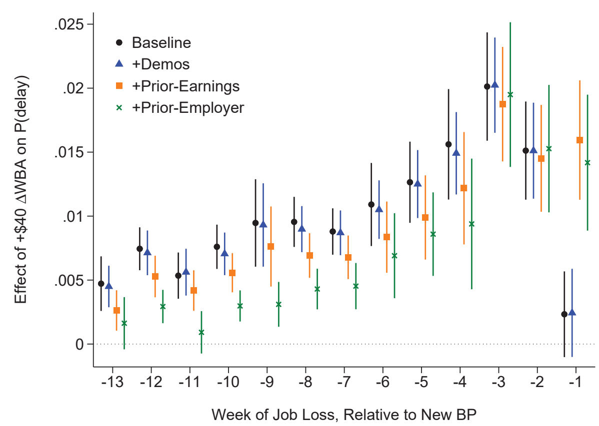
(a) ΔWBA = 0 Claims Excluded



(b) Post Policy Changes

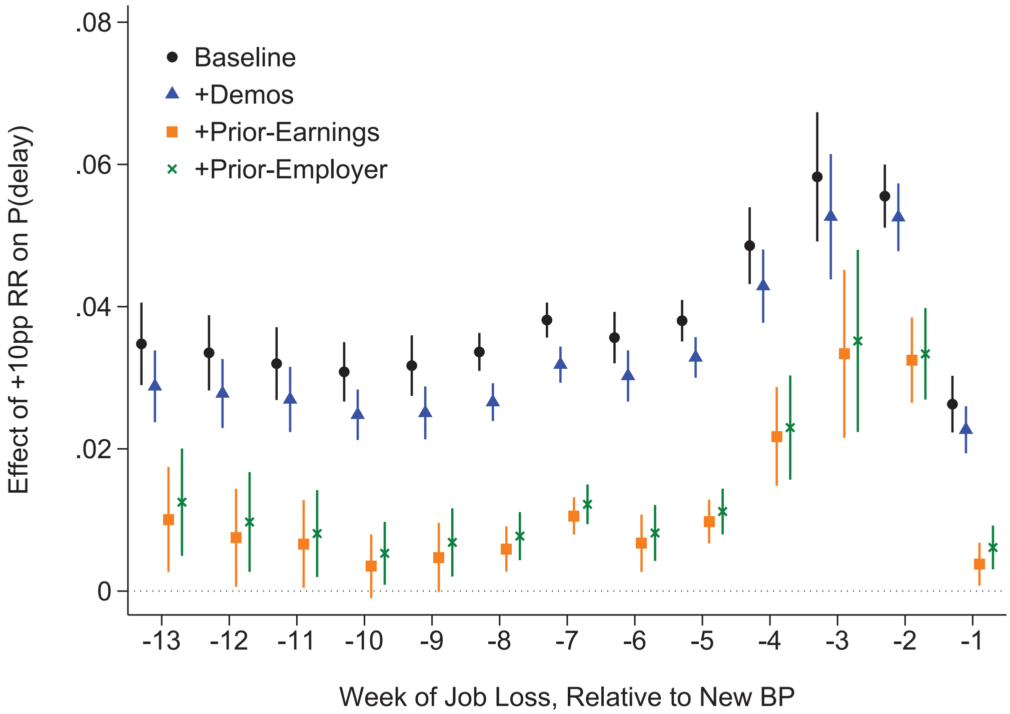


(c) Mass Layoff Sample



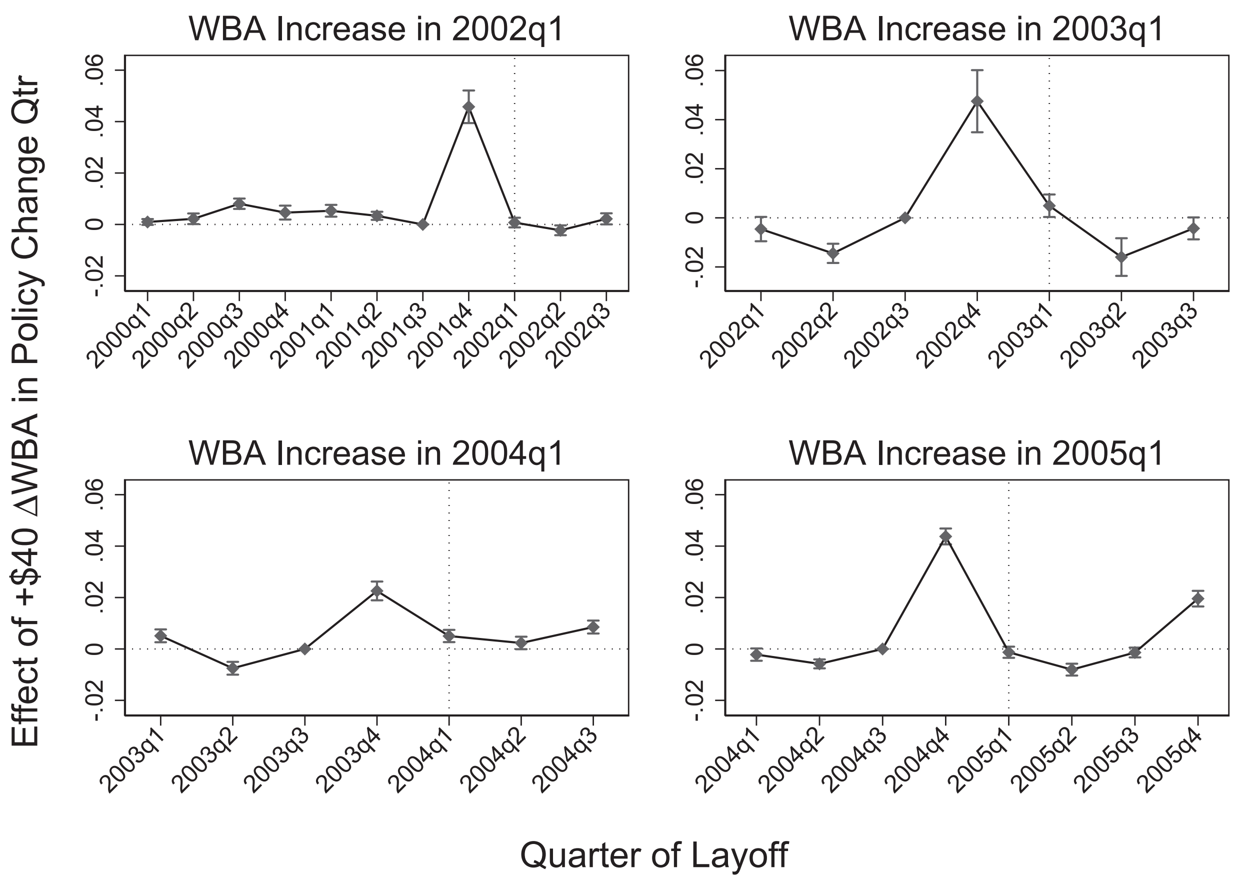
Notes: Estimates of the marginal effect of an extra \$40 in ΔWBA on the probability of filing a claim in the first week of the quarter after the layoff. (i.e., $\hat{\beta}_2^{\tau} + \hat{\beta}_3$ from equation 1.5.) Colors denote separate models, estimated with sequentially more complete sets of controls. Each model is estimated via OLS with cluster-robust standard errors at the layoff-quarter level. Each model includes quarter and weekday of layoff FEs, week of layoff (relative to the BP change) dummies, and a control for ΔPBD. "Demos" includes completed education, gender, age, ethnicity, citizenship status, and 3 digit zip code. "Earnings" includes average quarterly earnings totals in the 5 calendar quarters that span the two possible BPs and a measure of "effective" earnings volatility in the same period as described in section 1.5.3. "Pre-separation Employer" refers to the separating employer and includes the reason for job loss, an indicator for whether a recall is expected, firm size, average quarterly earnings of coworkers during the layoff quarter, and sector (two-digit NAICS).

Figure 1.A7: Effects of ΔWBA on P(claim filed 2+ weeks after layoff)



Notes: Estimates of the marginal effect of an extra \$40 in ΔWBA on the probability of filing a claim two or more weeks after the layoff. (i.e., $\hat{\beta}_2^{\tau} + \hat{\beta}_3$ from equation 1.5.) Colors denote separate models, estimated with sequentially more complete sets of controls. Each model is estimated via OLS with cluster-robust standard errors at the layoff-quarter level. Each model includes quarter and weekday of layoff FEs, week of layoff (relative to the BP change) dummies, and a control for ΔPBD. "Demos" includes completed education, gender, age, ethnicity, citizenship status, and 3 digit zip code. "Earnings" includes average quarterly earnings totals in the 5 calendar quarters that span the two possible BPs and a measure of "effective" earnings volatility in the same period as described in section 1.5.3. "Pre-separation Employer" refers to the separating employer and includes the reason for job loss, an indicator for whether a recall is expected, firm size, average quarterly earnings of coworkers during the layoff quarter, and sector (two-digit NAICS).

Figure 1.A8: Effects of ΔWBA on P(claim filed in first week of next qtr), event study specification



Notes: Figures plot $\hat{\beta}^q$ estimates from equation 1.8 around each of the four policy changes shown in Figure 1.2. Each panel represents a separate regression including all claims filed by claimants laid off in the quarters shown with $HQW_1 = HQW_2$ (i.e., claimants with no relevant earnings volatility). The "treatment" variable is defined as the average change within the earnings group in ΔWBA in the relevant policy change quarter. This treatment is interacted with indicators for layoff quarters in a regression with time (layoff quarter) and group (earnings group) FEs. The coefficients on the "treatment" $\times$ layoff quarter dummies and their 95% CIs are graphed. The sample in each regression is limited to claims filed by individuals laid off in the quarters shown. Standard errors are cluster robust at the earnings group level.

Table 1.A1: Some examples of policy change driven variation in benefit risk

Claimant	HQW_1	HQW_2	Layoff Quarter	WBA_1	WBA_2
1	\$10k	\$10k	2001q3	\$230	\$230
1	\$10k	\$10k	2001q4	\$230	\$330
1	\$10k	\$10k	2002q1	\$330	\$330
2	\$5k	\$10k	2001q3	\$130	\$230
2	\$5k	\$10k	2001q4	\$130	\$330
2	\$5k	\$10k	2002q1	\$160	\$330
3	\$10k	\$5k	2001q3	\$230	\$130
3	\$10k	\$5k	2001q4	\$230	\$160
3	\$10k	\$5k	2002q1	\$330	\$160

Notes: HQW_1 = HQW if claim is filed in layoff quarter, HQW_2 = HQW if claim is filed in quarter after layoff. Orange rows highlight effects of the 2001q4 policy change in the quarter of the policy change. Magenta rows highlight effects of the policy change in the quarter after the policy change.

Chapter 2

Unemployment Insurance as a Worker Indiscipline Device? Evidence from Scanner Data (with Lester Lusher and Rebecca L.C. Taylor)

2.1 Introduction

The Great Recession saw a dramatic increase in the duration of benefits available through the Unemployment Insurance (UI) program in the United States. Previously limited to between 26 and 30 weeks, by late 2009 eligible unemployed individuals in some states were able to receive benefits for up to 99 weeks. These dramatic expansions together with the decline in job availability led to a near 500% rise in the program's per-capita expenditures, making it the largest safety net program by per capita spending at that time (Bitler and Hoynes, 2016). Several studies have added to a large body of work on the relationship between UI generosity and unemployment duration—i.e., the ex post moral hazard effect—by exploiting these plausibly exogenous changes in UI potential benefit duration (PBD) (Rothstein, 2011; Farber, Rothstein and Valletta, 2015; Farber and Valletta, 2015; Marinescu, 2017; Johnston and Mas, 2018).[1] Since UI benefit changes alter the expected cost of

[1]Prior studies on the ex post moral hazard effect of UI include Katz and Meyer (1990); Meyer (2002); Card et al. (2015). In a recent review of this large literature, Schmieder and Von Wachter (2016) report a median elasticity of

job loss, theory also predicts an *ex ante* moral hazard response to UI benefit changes among the employed. In a simple model where effort is costly but protective against job loss, workers will respond to an increase in UI generosity by exerting less effort (shirking).[2] Because it is difficult for an employer to differentiate shirking from poor performance, this ex ante moral hazard effect should exist even if shirking would disqualify a worker from receiving UI benefits.[3]

To date, only a handful of studies have tested this theoretical prediction. Burda, Genadek and Hamermesh (2020) use the American Time Use Survey (ATUS) to study the effect of unemployment rates on shirking, which they measure as time spent not working while at work. They also demonstrate that maximum and average UI benefit levels are correlated with the intensive margin of shirking, but they do not attempt to identify a causal effect. The only other empirical[4] paper on this question focuses on self-employed workers in Denmark, which has a relatively unique, voluntary UI system (Ejrnæs and Hochguertel, 2013).[5] They exploit a policy change which differentially incentivized certain cohorts to enroll in UI and find that affected self-employed individuals were moderately more likely to become unemployed. No paper has yet identified the causal effects of UI generosity on worker effort in the context of a mandatory UI system such as exists in the US.[6]

This paper aims to fill this gap in the literature by matching plausibly exogenous changes in the

unemployment duration with respect to potential benefit duration of 0.13.

[2]Since we do not model optimal worker effort from the perspective of either the firm or a social planner, we do not explicitly define shirking and we use the terms "a decrease in effort" and "an increase in shirking" interchangeably.

[3]We provide a theoretical basis for the predicted effects of UI on worker effort in Online Appendix 1. Goerke (2000) shows that this result relies on the assumption that a worker fired for shirking is relatively unlikely to have their benefits reduced. Later in the paper, we provide empirical evidence for this assumption through a series of supplemental analyses with administrative UI claims data. In brief, we show that a meaningful proportion of UI claimants come from workers who were fired, and that such claims are likely to be accepted. We also discuss how specific UI policies make it unlikely that a worker fired for shirking would be denied benefits.

[4]Theoretical work is more common. For example, ex ante moral hazard has been shown to be a source of market failure in (hypothetical) private UI markets (Chiu and Karni, 1998).

[5]Related empirical work shows that inflows into unemployment spike when UI eligibility is obtained (Christofides and McKenna, 1995, 1996; Green and Sargent, 1998; Rebollo-Sanz, 2012) and when benefit levels increase (Winter-Ebmer, 2003; Jäger, Schoefer and Zweimüller, 2018). Since shirking does not necessarily result in job loss, the importance of a shirking ex ante moral hazard effect includes, but is not limited to, an explanation of these results. Since employer responses to these benefit changes could potentially explain these spikes in inflows, these results do not necessarily imply an ex ante moral hazard effect.

[6]According to Schmieder and Von Wachter (2016), among OECD countries only Denmark and Finland have voluntary UI systems.

PBD of UI benefits in the United States during the Great Recession with task-level productivity measures from a large sample of individual supermarket cashiers. The productivity measures are derived from high-frequency scanner data covering over 500,000 transactions conducted by nearly 2,000 cashiers spanning 39 grocery stores that are part of a national supermarket chain. The stores in our sample are located within a roughly 25 mile radius in the Washington D.C. metropolitan area, including eight stores in the District of Columbia, 17 stores in Maryland, and 14 stores in Virginia. During the sample time period (December 2008 to February 2011), changes in the parameters of the Extended Benefits (EB) and Emergency Unemployment Compensation (EUC) programs led to a series of large discrete increases in UI PBD. These extensions were designed as a response to the economic downturn. They were available to all UI eligible individuals and they differentially[7] affected the jurisdictions in our sample. Following recent studies on the effect of these extensions on job search (e.g. Farber and Valletta, 2015; Marinescu, 2017), we utilize this quasi-experimental cross-state variation in the size and timing of these expansions for identification.[8] Following earlier papers using nearly identical data (Mas and Moretti, 2009; Taylor, 2020), we measure productivity as the time-length of checkout transactions processed by cashiers.[9] These data and variation grant us the ability to estimate models with both cashier-register and date fixed effects while including transaction-level controls (e.g., number of items scanned by product category).

We provide several pieces of evidence that these PBD extensions were salient to workers similar to those in our sample. First, using Google Trends search data and nationally representative polls, we demonstrate that individuals in the US were likely to be aware of these extensions. Notably,

[7]Extensions varied across states in timing, magnitude, or both, depending on the specific extension.

[8]An important point detailed further in section 2.3.1 is that while the extensions were at times directly related to changes in state unemployment rates, the parameters of these programs also changed on several occasions during the Great Recession. We primarily rely on PBD changes that occur as a result of these federal and state policy changes which altered program parameters, as opposed to PBD changes that occur as a result of changes in unemployment rates.

[9]Investigating the impacts of discrimination in the workplace, Glover, Pallais and Pariente (2017) also observe worker productivity as measured by length of supermarket cashier transactions.

we show that Google search frequencies for terms related to UI spiked dramatically on key PBD extension dates. Second, we demonstrate that the vast majority of a different sample of the retailer's cashiers had earnings histories that would make them UI eligible in our state-years.[10] Third, we point out that the relevant eligibility criteria for UI are unlikely to disqualify a claimant fired for low productivity. We also demonstrate that it is common for fired workers to receive UI benefits in the US and that such workers do not typically see their claims rejected. We therefore argue that the typical cashier in our sample is very likely to have been UI eligible in the event of a job loss resulting from shirking.

In our main results, we demonstrate a modest but statistically significant negative relationship between the PBD extensions and worker productivity. Specifically, we show that cashiers who experience increased PBD levels take longer to complete customer transactions. The effect is stronger for cashiers who work more shifts during the sample period, i.e. cashiers who are more likely to be UI eligible. The effect is also more prevalent for less productive cashiers, who are likely closer to the margin of being terminated for poor performance. In our preferred specification, we predict an average increase of 2.4 seconds in transaction length for cashiers who experienced an 18-week increase in PBD.[11] With a mean transaction length of just under two minutes, this is roughly equivalent to a two percent decrease in worker productivity.[12] Over time these effect sizes can accumulate into rather large losses. Back-of-the-envelope calculations similar to those in Mas and Moretti (2009) and Taylor (2020) suggest that stores would need to staff 144 additional hours per year to offset the loss in productivity associated with an 18 week increase in PBD levels.[13]

[10]These data are from Mas and Moretti (2009) and include cashiers working in a different state several years prior to our sample. Since these data include all transactions occurring at several stores, we can estimate hours worked and UI eligibility at the cashier-shift level. This is not possible in our data.

[11]The average change in PBD across states per change in our time frame was 18.6 weeks, with each state experiencing two separate extensions. From the beginning to the end of our sample, Washington D.C. and Virginia experienced a total increase of 40 weeks in PBD, while Maryland experienced a 27 week increase.

[12]With a sample standard deviation just over 100 seconds, this effect is also roughly equivalent to a 2.3 percent standard deviation decrease in worker productivity.

[13]Additional details on these calculations are provided in the results section of the paper (section 2.4.2).

Assuming a $10 median hourly wage for cashiers in the US, this would cost each store in our sample $1,440 per year in additional wages, for an estimated total cost of $4 million per year across our sample states.

We are able to rule out several potential confounds. First, by estimating models with cashier fixed effects, we rely strictly on cashiers who experienced varying levels of PBD; this addresses concerns regarding changes in cashier composition in response to increases in PBD levels.[14] Similarly, register fixed effects account for potential shifts in the use of different registers (e.g., express registers). We also examine whether changes in consumer behavior related to PBD expansions drive the results. We find little evidence that consumers' purchases respond to PBD levels, with no statistically significant changes in expenditures and items per transaction or in the use of price discounts and coupons. We bring in additional data to examine aggregate store-level sales and find that sales significantly decrease with the unemployment rate but remain the same, or slightly increase, with PBD expansions. Finally, we find no statistically significant relationship between PBD levels and local unemployment rates, which highlights the discrete nature of the PBD changes that occurred during our period. Importantly, we estimate specifications of our main model controlling for the aforementioned potential confounds and find that they do not change our results.

To investigate whether our results may be generalizable to other sectors and regions of the US, we combine our identification strategy with a shirking measure used by Burda, Genadek and Hamermesh (2020). Specifically, we use the ATUS to test for an effect of increased PBD levels on the self-reported percentage of time spent on non-work activities while at work. This is made possible by the unique detail of the ATUS—a repeated cross-sectional survey which measures time spent doing precise activities (e.g., eating, childcare, socializing) and the location where those activities are performed. Utilizing PBD changes across the US from 2003 to 2014, and estimating

[14] We also find cashier experience is uncorrelated with PBD levels, suggesting managers did not significantly change the employment of their cashiers through our period.

models with state and month-year fixed effects, we estimate a positive effect of the PBD of UI on shirking in the ATUS sample. For our fully specified model, off a mean of 6.68%, we estimate a 0.35 percentage point increase in time spent at work not working in response to an 18-week increase in PBD. This analysis suggests that shirking responses to these benefit changes were not limited to our sample of cashiers.

Our results offer several important contributions to the limited empirical literature on UI's ex ante moral hazard effect. They constitute, to our knowledge, the first quasi-experimental evidence of such an effect *either* among the non-self-employed or within a mandatory UI system such as those utilized by nearly all developed countries. These results have several important implications. First, they quantify an understudied margin through which changes to UI benefit generosity affect social welfare (Baily, 1978; Chetty, 2006). Second, they contribute to the relatively small base of empirical evidence on ex ante moral hazard effects in *any* type of insurance.[15] Third, we contribute to the wider literature on the determinants of worker effort (e.g., the efficiency wage literature) by providing new estimates for two important theoretical predictions (the effects of the unemployment rate and unemployment benefits on effort) (Shapiro and Stiglitz, 1984; Lazear, Shaw and Stanton, 2016). Our results provide evidence that worker effort varies over both the business cycle and corresponding policy response, rising with the unemployment rate[16] and falling with UI generosity.

The remaining sections of this paper are as follows. Section 2.2 presents the data. Section 2.3 describes the quasi-experimental variation in UI benefits that we exploit and demonstrates the relevance of these benefit extensions for the workers in our samples. Section 2.4 presents our empirical specification and main results. Section 2.5 discusses potential threats to our identification

[15] Clear empirical evidence exists for automobile and worker's compensation insurance (Fortin and Lanoie, 2000; Cohen and Dehejia, 2004). Mixed evidence exists in the case of health insurance (Newhouse and Group, 1993; Decker, 2005; Dave and Kaestner, 2009). Hansen, Nguyen and Waddell (2017) document increases in injury length and subsequent take-up of workers compensation in response to increased benefits.

[16] We jointly estimate the impact of both local (county or ward) and state unemployment rates and find that cashier effort rises with both of these.

strategy and conducts several robustness checks and placebo tests. Finally, section 2.6 replicates the analysis with the ATUS data and section 2.7 concludes.

2.2 Data Sources and Summary Statistics

Quantifying a change in cashier productivity requires a detailed dataset on the speed of checkout transactions linked to cashiers. To this end, we obtained access to proprietary scanner data from a large supermarket chain for 39 stores in the District of Columbia (DC) metropolitan area (a roughly 25 mile radius around DC), including 8 stores in DC, 17 stores in Maryland, and 14 stores in Virginia.[17,18] These data—which span from December 2008 until February 2011—were originally obtained by Taylor (2020) to study how the 2010 disposable carryout bag tax in DC affected the transaction time of supermarket checkout.[19] These data are ideal for our research question for two reasons. First, during the time period of these data, a series of discrete changes to the PBD of UI benefits occurred which varied across DC, Maryland, and Virginia. Second, the richness of the data allows us to construct measures of cashier productivity, following other studies using the same data source for different purposes (e.g., Mas and Moretti, 2009; Taylor, 2020). Specifically, for each checkout transaction in our sample, we have information on when and where a checkout transaction occurred (e.g., register 4 in store G and state S on Saturday, June 19, 2010 at 5:37pm), what was purchased (e.g., a gallon of milk costing \$3.06), which cashier processed the transaction, and importantly, how much time the transaction took to complete. Using these identifiers, we are able to track stores and cashiers over time, before and after the changes in the PBD of UI benefits.

Our main outcome variable is *transaction time*—the duration of each checkout transaction measured in seconds, from the start of a transaction until the start of the next transaction in

[17] We do not include stores that were remodeled to add self-checkout registers during the sample, as self-checkout adoption could confound our productivity results.

[18] Online Appendix Figure A1 presents a stylized map of the Washington DC Metropolitan area.

[19] The **(alias?)** archives and documents existing datasets from this supermarket chain.

line. We are able to construct this variable using the transaction time-stamp, which includes the day, hour, and minute each transaction was completed. The sample includes all transactions at cashier-operated registers in the 39 stores between 5:00pm and 6:00pm for every Saturday during the roughly two year period. This weekend hour was chosen because the retailer cited it as a peak shopping time in their stores.[20] Since there is only one time-stamp per transaction, having peak hours enables us to calculate transaction time under the assumption that transactions in the scanner data occur back-to-back, with little or no downtime in between. Taylor (2020) verifies this assumption using observational data collected in-store during peak hours, where transaction length is timed with a stopwatch by enumerators stationed near checkout.[21] The advantage of using one time-stamp—and thus the full duration from one transaction to the next—is that all actions a cashier takes before swiping the first item (e.g., starting the conveyor belt) and all the actions after finalizing the purchase (e.g., printing the receipt and handing it to the customer) are included in our productivity measure. Downtime, on the other hand, refers to when there are no customers at the registers or in line, which is unlikely to occur during peak hours.[22]

2.2.1 Summary Statistics for Scanner Data

Table 2.1 presents the average transaction, cashier, and store characteristics, by state (columns 1–3) and for the full sample (column 4). Starting in panel A, there are 515,636 transactions in the Saturday 5:00-6:00pm sample. The average transaction has an approximate duration of 120 seconds, is comprised of 12 items, and costs \$35.[23] Average transactions in DC stores take slightly

[20] We drop transactions occurring at self-checkout registers because only seven stores have self-checkout during the sample period, and these registers are not manned by an individual cashier.

[21] To additionally ensure that transactions occur back-to-back, we also drop all transactions that are more than three standard deviations longer than the average transaction of its size (in terms of number of items scanned) and all transactions that are longer than 20 minutes.

[22] Saturday from 5:00–6:00pm is not the only peak foot-traffic hour in a week; however, due to size constraints in obtaining data from the retailer at the transaction level, the original data request was limited to one hour per week for the sample of stores.

[23] Transaction expenditure is created by summing up the individual amounts paid per item in a transaction. This variable does not include sales tax.

longer to complete yet have roughly the same size and cost as stores in Maryland and Virginia. The average transaction saves roughly $8 in price discounts. Price discounts occur when stores put items on sale, or when customers use their rewards card or coupons.[24]

Panel B of Table 2.1 presents average characteristics at the cashier-level. There are 1,984 unique cashiers in our sample. The average cashier works 13.5 of the 113 Saturdays in our sample. The average span of days from when we first see a cashier until we last see them is 230 days, suggesting cashier is a position with high turnover. The average cashier works 40 minutes per hour and the median cashier works 48 minutes per hour—this being less than 60 minutes reflects cashiers entering/exiting their shift within the hour. Panel C presents average store-level characteristics. Stores in DC have more registers and more cashiers than stores in Maryland and Virginia. In the full sample, the average store has 7 registers and 51 cashiers.

Figure 2.1 plots the relationship between the number of items in a transaction and average transaction time. This figure reveals that the fixed time cost of processing a transaction is a minute, with the average one-item transaction taking 60 seconds to complete. Furthermore, the variable time cost is linear in transaction size, with each additional item adding roughly 3 seconds to the baseline transaction time.

2.2.2 Additional Data - American Time Use Survey (ATUS)

In additional analyses, we utilize the American Time Use Survey (ATUS) (Hofferth, Flood and Sobek, 2017). The primary benefit of the ATUS is that we can measure on-the-job effort for workers across different sectors and for the entire US. The ATUS is a repeated cross sectional survey of former Current Population Survey (CPS) respondents which elicits time diaries of individuals. The time diaries collect detailed information on the nature of activities, the duration of activities (in

[24] Price discounts are measured as the difference between the shelf price of an item and the price customers pay, summed across all items in a transaction.

minutes), and the location of the activities (e.g., at the workplace). A total of 159,937 surveys were conducted between 2003 and 2014.

Our measure of effort in the ATUS closely follows that of Burda, Genadek and Hamermesh (2020), who investigate the relationship between shirking and state unemployment rates in the ATUS. Among activities conducted "at the workplace" by a sample of adult workers in the ATUS,[25] we identify work-related activities using ATUS activity codes 50000 to 50299. We then reclassify "socializing, relaxing, leisure, eating, drinking, sports, exercise as part of the job" as non-work (codes 50201-50203), "travel related to work" as work (codes 180501, 180502, 180599), and "work and work-related activities not elsewhere classified (n.e.c.)" (code 59999) as work. For our primary dependent variable, we calculate each individual's percentage of time at the workplace that they engaged in non-work activities.

Online Appendix Table A1 presents summary statistics for our ATUS sample of 30,094 workers.[26] The average worker in our sample is 40 years old. Approximately 46% of workers are female, 83% are white, 92% are born in the US, 83% work in the private sector, 12% work part time, and 45% are paid hourly. The three most popular occupational sectors are management (11%), sales (10%), and office and administrative support (15%). Respondents report working an average of 42 hours per week, with weekly earnings of $900. For the day that the worker was surveyed, respondents spent an average of 479 minutes (approximately eight hours) working, and over 31 minutes not working while at the workplace.

[25]Specifically, we limit the sample to US citizens with a single job, who reported "usually" working at least 20 hours per week, are 18-65 years old, and are not self-employed (self-employed workers are ineligible for UI).

[26]These statistics use the probability weights provided.

2.3 UI Benefit Extensions, Awareness, and Eligibility

In this section, we first detail the expansions in Unemployment Insurance (UI) potential benefit duration (PBD) during the period of our study. We then provide evidence for two key prerequisites for the effect of interest. First, we show that awareness of the PBD extensions was widespread. Second, we show that workers similar to those in our data were likely to be eligible for UI benefits.

2.3.1 Unemployment Insurance Benefit Extensions

In the US, the PBD of UI benefits is set by individual states.[27] Typically limited to 26 weeks,[28] PBDs are regularly extended during economic downturns. During the period of our study these extensions were driven by three separate programs—the Extended Benefits (EB) program (1970-current), the Emergency Unemployment Compensation (EUC) program (7/2008-12/2013), and the Temporary Extension of Unemployment Compensation (TEUC) program (3/2002-12/2003).[29] The exact number of additional weeks available to UI claimants in each state by the EB and EUC programs are made available online at the weekly level by the US Department of Labor (DOL), **(alias?)**.

Table 2.2 presents the changes in PBD due to these programs for the state-weeks in our scanner data sample and Figure 2.2 depicts similar information for the state-months in our ATUS analysis.[30] The PBDs shown are those available to new claims filed on a given date. Since the timing and magnitude of these extensions varied by state, PBD variation resulting from the EB, EUC, and TEUC programs has been utilized as identifying variation in a series of recent studies on the effects of UI benefit generosity (Rothstein, 2011; Farber, Rothstein and Valletta, 2015; Farber and Valletta,

[27] PBD is determined by the state of employment, not residence.

[28] With the exception of Montana and Massachusetts, which provide 28 and 30 weeks, respectively.

[29] The EB and EUC programs are relevant for both the scanner data and ATUS analyses, whereas the TEUC program is relevant only for the ATUS analysis.

[30] Data for the ATUS sample extensions were obtained from a replication file for Farber, Rothstein and Valletta (2015). Note that, although the analyses in Farber, Rothstein and Valletta (2015) are limited to the years 2008-2014, the replication file includes extension on/off dates starting in January 2000.

2015; Marinescu, 2017; Chodorow-Reich, Coglianese and Karabarbounis, 2019; Boone et al., 2021). More detailed descriptions of these programs and related legislation can be found in these studies and in Online Appendix 2.

Here we emphasize the key point that the extensions which we exploit for identifying variation occur for one of three reasons: (1) a state's unemployment rate (specifically its average Insured Unemployment Rate (IUR) over the past 13-weeks or Total Unemployment Rate (TUR) over the past 3-months) crosses a threshold or "trigger" value currently in place, (2) the relevant authority (state government for EB, federal for EUC or TEUC) changes the trigger value to a level below the state's current 13-week IUR or 3-month TUR, or (3) the federal government alters the (EUC or TEUC) program by changing the number of weeks available or allowing the program to expire (either temporarily or permanently). Notably, only 2 of 7 separate PBD changes exploited in our scanner data sample (shown in Table 2.2) occurred as a direct result of changes to the relevant state's unemployment rate.[31] The remaining extensions occurred due to policy changes implemented at the state or federal level. This is important to note because extensions which are a direct result of changes in state unemployment rates may be problematic for our design, as further discussed in section 2.5.1.

Between 4/2011 and 8/2014 the states of Arkansas, Florida, Georgia, Illinois, Michigan, Missouri, North Carolina, and South Carolina each passed legislation reducing their regular PBDs below 26 weeks (Isaacs, 2019). This variation is not relevant for our scanner data sample but is utilized in our ATUS analyses. The specific dates and magnitudes of these reductions are also explained in more detail in Online Appendix 2.

[31] This ignores three EUC program lapses shown in Table 2.2 which also do not result from changes in unemployment rates.

2.3.2 Awareness and Relevance of UI Benefit Extensions

To investigate the possibility that workers were unaware of the extensions that we study (which would preclude our effect of interest), we turn to Google Trends to look at search frequency of the terms "Unemployment benefits" and "Emergency Unemployment Compensation" across the US on Google's search bar. Google Trends reports the relative search frequency of particular items on Google Search within a queried geography (e.g., the US) and time period (e.g., January 2008 - December 2009), indexed to a range of 0 and 100.

Figure 2.3 plots these trends. For the search item "Unemployment benefits," there are three jumps in search frequency corresponding to enactment (June 30, 2008) and subsequent adjustment (November 21, 2008 and November 6, 2009) of the Emergency Unemployment Compensation (EUC) program; note, however, the search frequency for "Unemployment benefits" was highest, within this period, during the time when the American Recovery and Reinvestment Act (ARRA) was implemented. In the second panel of Figure 2.3, we report the trends for the search item "Emergency Unemployment Compensation." Though noisier, we find that across the sample of 104 weeks, search frequency was at its highest during the weeks after the EUC program enactment and two subsequent alterations. In fact, the two weeks of the EUC alterations produced the two highest search volumes for EUC within our sample, while the week after the enactment of the EUC program carried the fourth largest search volume. Though these results do not reflect *absolute* search volumes, the relative spikes in search volume reflect, among Google Search users, an awareness of the EUC program enactment and expansions.

The online appendices provide further evidence of both UI benefit awareness and the relevance of PBD extensions to workers during the Great Recession. In Online Appendix 3, we use Google Trends to show that "Unemployment benefits" was also a popular search relative to several other common search terms during this time period—such as "Social security," "Disneyland," and "Wall-

E." This appendix additionally provides evidence on UI awareness from national polls conducted during the Great Recession. In Online Appendix 4, we use the CPS (Flood et al., 2017) to provide evidence that unemployment spells often extended into and beyond periods covered by the PBD expansions and thus are sufficiently long to be relevant to the workers in our data.

2.3.3 UI Eligibility

UI eligibility is determined by two sets of criteria. Monetary eligibility requires a minimum amount of earnings in a certain time period pre-claim, and non-monetary eligibility conditions UI benefits on the reason for the job loss. In this section we provide supporting evidence that our workers are likely to meet these criteria in the event of a shirking-induced job loss.

Mas and Moretti (2009) utilize a different sample of our retailer's cashiers, covering a two-year period for six stores between 2003 and 2006.[19] Since their data includes all transactions occurring at these stores, we can use it to estimate monetary eligibility. To do so, we assume that the retailer is each cashier's only employer and that each cashier was paid the relevant minimum wage. We apply the UI eligibility rules for each state-year in our sample to each cashier-shift in theirs and find an average eligibility rate of 74%. Online Appendix Table A2 presents the results from this analysis.

The key non-monetary criteria in all states is that the worker lost their job through no fault of their own. A concern in our setting is that employees who are fired for shirking may be "at fault." However, in practice a discharged claimant is only at fault if they have committed misconduct. Legal definitions of misconduct vary by state, but according to the DOL's Employment and Training Administration, many states rely on the definition established in the 1941 Wisconsin Supreme Court Case, Boynton Cab Co. v. Neubeck (see chapter 5 pg. 10 in Employment & Training Administration (2019):

> "Misconduct...is limited to conduct evincing such willful or wanton disregard of standards of behavior which the employer has the right to expect of his employee, or in carelessness or negligence of such degree as to manifest an equal culpability, wrongful intent or evil design, or to show an intentional and substantial disregard of the employer's interest or of the employee's duties and obligations to his employer."

Importantly, the burden of proof for an allegation of misconduct is on the employer. As Hagedorn et al. (2013) discuss in detail, attempting to prove misconduct is costly for employers and the probability of success is low. Especially relevant for our setting, Hagedorn et al. (2013) point out that an employer must prove "willfulness" on the part of the claimant in order to demonstrate misconduct. So long as worker effort is imperfectly observable by employers,[32] this standard seems unlikely to be met in the case of a worker laid off for decreasing their on the job effort in response to an improvement in their outside option.[33] Various public resources provided by employment law specialists reach similar conclusions. For example, **(alias?)** (a leading provider of do-it-yourself legal advice) states that "An employee who is fired for being a poor fit for the job, lacking the necessary skills for the position, or failing to perform up to expected standards will likely be able to collect unemployment."[34]

We support these arguments empirically in three ways. First, using administrative claim-level data from the State of California's Employment Development Department (EDD) , we find that approximately 10% of claimants in California were fired from their previous job. This share rises to 20-25% among cashiers and/or food store workers. Second, using the DOL's Benefit Accuracy Measurement (BAM) data we show that the proportion of UI claimants who were fired during the time period of our ATUS analysis was around 20% nationally. Third, we utilize the DOL Employment and Training Administration's quarterly report number 207 to document variation across states in

[32]This is a common assumption in efficiency wage models (e.g., Shapiro and Stiglitz, 1984).

[33]We direct the interested reader to Appendix VI in Hagedorn et al. (2013) for detailed descriptions of specific related cases in the context of California's UI program.

[34]Similarly, in their definition of misconduct, the National Employment Law Project states that "Neither workers fired simply for poor performance nor capriciously fired workers should be denied UI benefits" (National Employment Law Project, 2015).

the fraction of misconduct determinations resulting in a denial of benefits. Consistently from 2003 to 2014, in the median state, roughly 40% of such determinations result in denials; many of these denials only partially reduce benefits and many of these misconduct determinations are unrelated to shirking. Together these three results strongly suggest that a monetarily eligible worker who is fired for shirking is likely to receive UI benefits.[35]

Finally, performance need not be the direct reason for a worker's termination to still impact their probability of termination. Most obviously, a downsizing firm may first decide to terminate their lowest productivity workers. In our setting, it is feasible that the supermarkets first terminated their slowest cashiers in the event that they had to downside during the Great Recession. Later results will investigate differential shirking responses by cashier productivity to find that lower productivity cashiers (i.e. those most likely on the margin for termination) responded more strongly to changes in UI PBD.

2.4 Main Results

2.4.1 Econometric Specification

Our primary specification estimates the following equation:

$$TransactionLength_{tdcrs} = \beta\, UIPBD_{ds} + \lambda_d + \lambda_{crs} + \gamma X_{tdcrs} + u_{tdcrs} \tag{2.1}$$

where $TransactionLength_{tdcrs}$ is the length of transaction t performed on day d (e.g., February 12, 2010) by cashier-register cr (e.g., Cashier ID #456 working checkout line #5) in state s (e.g., Virginia), $UIPBD_{ds}$ is the maximum PBD of UI benefits available in state s on day d, λ_d and

[35] Plotted figures from these three datasets can be found in Online Appendix Figure A3. EDD is the state agency which administers UI in CA, and data were acquired through a partnership between the EDD and the California Policy Lab. The BAM program, run by the DOL's Employment and Training Administration, audits a random sample of UI benefit recipients in every state-week, and their data are publicly available.

λ_{crs} are date and cashier-register fixed effects, and X_{tdcrs} is a vector of transaction-level controls. The coefficient β can be interpreted as the predicted increase in transaction length (in seconds) in response to a one-week increase in the UI benefit duration. A positive β would imply that cashiers shirk in response to more generous benefits.

Cashier-register fixed effects denoted by λ_{crs} control for all unobserved factors that vary at the cashier-register level and affect transaction time. Importantly, since cashier-register fixed effects strictly rely on variation within cashiers, our identification strategy accounts for the possibility that the composition of employed cashiers changed with PBD. For example, without cashier fixed effects, our estimates would be biased away from a shirking effect if workers employed during higher levels of PBD were more productive. Register fixed effects account for any differences in processing speeds across registers (e.g., express lane vs. regular). By estimating cashier-register fixed effects, we account for any potential cashier-specific differences across registers; for instance, some cashiers may work more quickly on express lanes compared to other cashiers. Transaction-level controls X_{tdcrs} include: the total expenditures paid in dollars, the total number of items purchased, price discounts and coupons in dollars, indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, the cashier's experience as measured by total number of transactions completed up to that point in the sample, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, the cashier's length of shift measured in both number of transactions and in minutes, prior month local unemployment rates—at the county level for MD and VA (from the BLS[36]) and at the ward level for DC (from the DC Department of Employment Services)—and state unemployment rates (TURs, from the BLS).

[36]Since we observe the mailing address of stores in our sample we use a zip-to-county crosswalk from the Department of Housing and Urban Development.

2.4.2 Results

Table 2.3 presents results from four different versions of our baseline model (equation 2.1), all estimated via OLS with cluster-robust standard errors at the state-by-date level. Across all specifications, we estimate a statistically significant positive effect of PBD on transaction duration. Notably, the inclusion of the transaction-level control variables X_{tdcrs} does not greatly alter our coefficient of interest. In our preferred specification (column 4), we estimate a 0.130 second increase in transaction length for a one-week increase in PBD.

In column 5, we replicate column 4 with the inclusion of customer-card fixed effects. At this supermarket chain, customers may opt to use a customer rewards card. If a customer uses their rewards card we can follow them across multiple transactions. Approximately 70% of transaction are carried out by rewards card members. Since this approach relies on rewards card members who shopped on at least two separate Saturdays from 5-6pm with differing PBDs, we do not prefer it to column 4. Still, it is reassuring that our estimate for β in this model (0.136) is very close to that from our fully-specified model (0.130).

The average PBD extension in our sample is just over 18 weeks. With an 18-week extension, these estimates correspond to an increase in transaction duration between 2.4 and 2.7 seconds, or roughly 2% of the overall sample mean. Using a back-of-the-envelope calculation, stores would need to staff 144 additional hours of work to maintain the same level of store productivity as before the PBD extension.[37] With a $10 median hourly wage for cashiers in the US (BLS, 2017*b*), this would cost the stores in our sample $1,440 higher wage bills per year. Aggregating further, if

[37] On average, stores in our sample process 130 transactions per Saturday 5:00pm hour before the PBD extension. To maintain this level of production when the average transaction is 2.4 seconds slower, stores would need 312 seconds more work per hour (i.e., open up an additional register for 312 seconds). This is only for one hour per week. To aggregate this to the annual level, we use an industry white paper which finds that half of grocery shopping transactions in the US occur during 32 peak hours in a week (Goodman, 2008), where a peak hour is defined as a time wherein more than 3 million people shop during that hour of the week. This translates to stores needing to staff 144 additional hours per year than before the PBD extension (i.e, 312 seconds $\times$ 32 peak hours per week $\times$ 52 weeks in a year $\div$ 3600 seconds per hour).

the 39 supermarkets in our sample are representative of the 2,891 supermarkets and other grocery stores in DC, Maryland, and Virginia,[38] the PBD extension would cost supermarkets in these states $4 million per year collectively. The estimated coefficients on the size of the transaction provide additional context. A typical PBD extension in our sample increases the transaction durations of affected cashiers by a magnitude roughly equivalent to increasing the size of the transaction by 0.8 items (7% of the sample mean transaction size).

Coefficients on the state and local unemployment rate are negative and are only statistically significant, at the 10% level, in our customer fixed effects model. This result suggests that a cashier's direct response to the increasing unemployment rates is to boost effort and productivity (i.e. it works against our estimated effect of PBD). Coefficients on the local unemployment rate (county-level in Maryland and Virginia, ward-level in D.C.) are much smaller than those for the state.

2.4.3 Subsamples

To understand heterogeneity in the effect of PBD on transaction duration, in Table 2.4 we estimate equation (2.1) for several subgroups. These subgroups are defined based on cashier or register characteristics which are expected to alter the effects of PBD on worker effort.

The first split is defined by a measure of cashier experience: the number of shifts the cashier worked in our sample before the first PBD change (i.e., before April 5, 2009). As described in section 2.3.3, more experienced cashiers are more likely to be UI eligible. Results in Table 2.4 suggest a slightly stronger treatment effect for such cashiers. In a subsample of cashiers who worked more than the top quartile of shifts in our sample, the predicted effect of PBD on transaction duration is 25% larger relative to the full sample while the treatment effect for the lower quartiles is in line

[38] Store counts come from the **(alias?)**, 2017 Retail Trade Summary Statistics.

with the full sample. Estimates remain statistically significant for both of these subsamples.[39]

Next, we consider heterogeneity by cashier productivity. Since productivity is imperfectly observable by managers and unproductive workers are likely closer to the margin of being terminated, the effort decisions of less productive workers are expected to be more responsive to changes in PBD levels. Conversely, it is relatively unlikely that highly productive workers would be terminated for a drop in performance, and thus, these workers would be less responsive. To test this, we first estimate each individual cashier's fixed effect in the pre-policy period (pre-April 5, 2009), conditioning on transaction-level controls and date fixed effects.[40] We then estimate our preferred specification separately among productive ($\geq$75% percentile of productivity) and less productive cashiers. Our primary cashier productivity split is presented across the fourth and fifth columns of Table 2.4. We find virtually no treatment effect for high-productivity cashiers, while less productive cashiers display a statistically significant increase in transaction length during periods with higher PBD levels.[41]

Lastly, we consider a set of subgroups defined by the type of register that was used to conduct the transaction—express vs. regular. With smaller transactions and more time sensitive customers, one may expect cashiers working at express registers to have little opportunity to shirk. Conversely, larger transactions conducted on regular registers plausibly present more of an opportunity for shirking. The results from the last two columns of Table 2.4 are consistent with this hypothesis. For

[39] Online Appendix Figure A5a demonstrates this pattern of results visually. As the minimum number of shifts worked grows beyond 8 shifts (i.e., the top quartile of shifts worked before the first PBD change), point estimates increase. However, even though most point estimates are statistically significantly different from zero, we cannot conclude that the difference in treatment effects between high and low experience cashiers is statistically significant.

[40] Transaction level controls include the number and types of items scanned and the register worked.

[41] Note that the number of observations across these two subgroups do not sum to the full sample since new cashiers enter the sample after the first PBD change on April 5, 2009. Approximately 38% of the full sample of transactions were conducted by cashiers that did not work before the first PBD change. Online Appendix Figure A5b tests the sensitivity of this split by plotting estimated treatment effects across an array of subsamples by cashiers' ranked productivity, starting with the full sample on the left and the most productive cashiers on the right (culminating with a subsample limited to cashiers ranked 600 and above). Once again, we observe no treatment effects when we focus strictly on the most productive workers. More generally, point estimates decreases as we move from the full sample on the left to the most productive workers on the right.

transactions conducted on express registers, the estimated effect is nearly zero. At regular registers, the treatment effect is larger than in the full sample, with an 18-week extension translating to a statistically significant 4.6 second increase in transaction length.[42]

2.5 Threats to Identification, Robustness Checks, and Placebo Tests

2.5.1 Threats to Identification

In our two-way fixed effect specification, potential sources of endogeneity bias in estimates of β are from omitted time-varying factors correlated with both PBD and transaction length. Below, we address three key examples: Changes in consumer purchases/composition, time-varying changes to local labor market or cashier characteristics, and other policy changes that may influence transaction length.

Changes in consumer purchases

Since higher levels of PBD are (partly) triggered by higher unemployment rates, they may be associated with changes to consumer purchasing behavior. This is a concern in our setting if such changes also affect transaction length. During downturns, consumers might make fewer trips to the grocery store while purchasing more goods during each visit. Consumers may also increase their "price-consciousness" during higher PBD periods, seeking out coupons and price discounts.[43] More generally, consumers might buy more of certain types of goods during high PBD periods which take longer to scan. Such responses are not driven by cashier behavior but could lead to

[42] Ignoring register type, Online Appendix Figure A6 demonstrates a consistent pattern of increasing treatment effects across larger transactions. For the full sample, we estimate a treatment effect of 0.13, and as we reduce the sample to include only larger transactions, this estimate slowly increases, culminating in a treatment effect over 0.40 seconds for transactions with at least 20 items.

[43] Nevo and Wong (2019) find that households purchased more on sale and used more coupons during the Great Recession.

increases in transaction length. In light of these concerns, it is important to note that our preferred specification controls for the number, price, and types of items purchased by consumers. While we find that these controls do not greatly change our coefficient of interest (as shown in Table 2.3), we provide supplementary evidence against this potential bias in this section.

We begin by analyzing how total supermarket sales vary with UI benefit generosity. While studies have shown that food expenditures fell during the Great Recession (Kumcu and Kaufman, 2011; Griffith, O'Connell and Smith, 2016) and that food consumption falls at UI benefit exhaustion (Ganong and Noel, 2019) (suggesting that PBD increases would increase food consumption among benefit exhaustees), no study that we are aware of has examined how total supermarket sales vary with PBD extensions. To address this open question, we use retail scanner data collected by Nielsen© and made available through the Kilts Center at The University of Chicago Booth School of Business. The retail scanner data consist of weekly price and quantity information generated by point-of-sale systems for nearly 40,000 participating grocery, drug, and mass merchandiser stores across the US. Nielsen's sample of stores cover more than half the total sales volume of US grocery and drug stores and more than 30 percent of all US mass merchandiser sales volume. We use the data spanning from January 2009 through December 2014.

We aggregate the raw micro data—i.e., store j sold x units of product z in week w, where a product is represented by a universal product code (UPC)—to the store-month level for three variables: (1) total sales in dollars, (2) total sales in the number of items sold, and (3) average expenditures per item.[44] The average store in 2009 had monthly sales equal to 183,808 items and $541,605 and average expenditures per item equal to $3.21. In Table 2.5, we test how store sales correlate with state PBD levels and lagged unemployment rates. In the simplest specifications with only PBD level, store fixed effects, and month-year fixed effects, we find no statistically significant

[44] The average expenditures per item is calculated as total sales in dollars divided by total sales in the number of items sold.

relationship between PBD and total store sales (columns 1, 3, and 5). When we add lagged state UE rates, we find a positive and significant relationship between PBD level and sales measured in dollars (column 2) and a positive but not significant relationship with sales measure in items (column 4). These effect sizes are quite small, equal to 0.03% and 0.01%, respectively. Moreover, similar to the prior literature, we find a negative relationship between the unemployment rate and store sales. This effect size is much larger, with a 1 percentage point increase in the prior month's unemployment rate corresponding to a 1% decrease in sales (both in dollars and in items sold). Lastly, we find a tiny but statistically significant relationship between the amount spent per item and PBD, with an 18-week increase in PBD duration corresponding to 0.03% increase in the amount spent per item (column 6). Overall, these results suggest store sales decrease in economic downturns and remain the same, or slightly increase, when UI benefit generosity increases.

Returning to our primary transaction-level scanner data, in Table 2.6, we test for each of the considerations involving changes in consumer behavior by collapsing our data to the store-date level and regressing a series of characteristics on PBD, conditional on date and store fixed effects. From the first two cells in panel A, we immediately see that consumers are not buying more per visit during periods with higher PBD levels, both in terms of total dollars and number of items. The coefficients on expenditures per transaction and items scanned per transaction are both statistically insignificant. Moreover, we find no statistically significant changes in the usage of price discounts during higher PBD periods. This is reassuring because we might expect transactions using price discounts, such as coupons and reward cards, to take longer to process. Therefore, while customer coupon- and sale-use may have been higher during the Great Recession (Nevo and Wong, 2019), we do not find that these behaviors responded to changes in PBD during the Great Recession.

We also do not find statistically significant changes in the number of registers opened during higher PBD periods. This is reassuring because, with only one time-stamp per transaction, we assume that the time between customer transactions is unchanging. If PBD extensions led to fewer customers, stores might reduce the number of lanes open to avoid register idle time. However, we do not find a reduction in the number of registers open, which we would expect if stores were experiencing fewer customers. Unreported in this manuscript, we also estimate a hazard model to test whether higher PBD periods decrease the probability that a customer returns to a store in subsequent weeks, given they have yet to have returned to the store. We find no statistically significant relationship between PBD extensions and the likelihood customers return, suggesting once again that PBD extensions do not lead to lower customer volume.

Finally, when we investigate by product category (Alcohol & Tobacco, Bakery & Deli, Dairy, Floral, Frozen Items, Meat & Seafood, Produce), we do not find statistically significant changes in the purchases of a particular type of products, with the exception of alcohol and tobacco which is positively correlated with higher PBD periods. Given alcohol and tobacco purchases are associated with slower checkout speeds—as cashiers must check the identification cards of the purchasers—absent the controls we include for purchases of alcohol and tobacco, our estimates for β would be biased in the direction of a shirking effect.

There are also important consumer behaviors that we cannot measure in our data and may be a concern for our identification strategy. First, we are unable to measure payment method, which could be correlated with both PBD and transaction length. Hurd and Rohwedder (2010) find that the ownership of credit cards declined by 2.8% of households during the Great Recession. Polasik et al. (2012) show that cash is a significantly faster payment method than traditional payment card, but that contactless cards and mobile payments have similar time efficiency to cash. Thus, while we cannot measure changes in payment methods directly, the literature suggests slower forms of payment were used less often during the Great Recession.

Changes in time-varying labor market conditions and cashier characteristics

Since our identification strategy utilizes variation within cashiers (and across days) with cashier-register fixed effects, our estimates will only be biased in response to cashiers if there are any time-varying cashier characteristics that are associated with PBD and transaction length. A key concern is that cashier effort responds directly to state unemployment rates. We first note that theory clearly suggests that this response would move in the opposite direction of our effect of interest. A weaker labor market implies that the costs of job loss are higher and workers are expected to respond by increasing their effort on-the-job.

We show in Table 2.6 that lagged state unemployment rates do not display a statistically significant relationship with PBD levels. This is perhaps not surprising given that (as discussed in section 2.3.1) most PBD extensions occurring in our scanner data sample did not result from changes in state unemployment rates. Further, as described in Online Appendix 2, when state unemployment rates are relevant, the specific rates that matter are both measured over longer time frames (e.g., 13 weeks) and in more complicated ways (e.g., benchmarking relative to similar rates in prior calendar years via "lookback" provisions). Finally, we further addressed concerns related to the correlation between PBD levels and unemployment rates by controlling for both state and local unemployment rates[45] in our preferred specification in Table 2.3.

A separate possibility is that cashier shift length changes with PBD levels. This would bias our estimates away from zero if cashiers work longer shifts during higher PBD levels and longer shifts correspond to reduced productivity. (e.g., if less productive cashiers were laid off, while more productive cashiers were given longer shifts.) Cashier fixed effects control for the change in the composition of cashiers, but they do not account for the possibility of increased shift length within

[45] It is reasonable to assume that it is the strength of the local economy, and not the economy of the entire state, that is affecting worker effort decisions. Throughout our sample we observe substantial within-state variation in unemployment rates across these different local areas.

cashiers. Since our sample contains all transactions during a single hour on Saturdays, we cannot directly test for changes in shift length. However, we proxy for this using the average number of open registers, the average experience of employed cashiers, and the number of employed cashiers. From Table 2.6, we find no statistically significant relationship between PBD levels and these three covariates, suggesting there was little response from labor supply to differing PBD levels.[46]

Other policy changes

A final class of relevant omitted factors are other policy changes that are correlated with PBD extensions and that may have influenced transaction length. For instance, more generous food stamp policies (officially known as Supplemental Nutrition Assistance Program or SNAP) may have been adopted during periods with higher PBDs. Returning to Table 2.6, in panel C, we regress the number of state-month SNAP participating households on PBD levels and find no statistically significant relationship.[47] Further, if SNAP had been correlated with PBD levels, we find it unlikely that food stamp usage would influence transaction length since SNAP benefits are paid in the form of Electronic Benefit Transfer (EBT) cards that are swiped at checkout in the same manner as debit cards with a PIN (Bartfeld et al., 2015). EBT cards are specifically designed to look and act like debit cards in order to reduce the potential stigma of participating in SNAP.

Another policy of relevance is the adoption of plastic bag taxes. During the period of our study, the only jurisdiction to adopt a bag tax was DC, but this adoption still leads to a statistically significant correlation (at the 1% level) between the bag tax policy and PBD levels (see Table 2.6). This generates an obvious concern for a bias in the same direction as a shirking channel, since plastic bag taxes have been shown to have a significant negative impact on worker productivity (Taylor, 2020). To account for this in our preferred specification (in Table 2.3), we simply control

[46]This finding is further supported by Mas and Moretti (2009), who suggest managers have relatively little say in cashier shift timings and length.

[47]SNAP participation data come from the US Department of Agriculture, **(alias?)**, SNAP Data Tables.

for the adoption of the bag tax policy.

2.5.2 Robustness Checks

In the online appendices, we present various robustness checks to our main estimates. First, we consider an analysis where we collapse our data to the cashier-register-day level and calculate the (log of the) average of the cashier's transaction length for that day. Results from these additional specifications are in Online Appendix Table A3 and are consistent with our main results. Next, in Table A4, we show that our main results remain unchanged after dropping all cashier(-shift)-level controls. One motivation for dropping these controls is their potential endogeneity with PBD levels. In columns 3–6 of Table A4, we replicate our preferred specifications after (a) ignoring periods where EUC benefits temporarily dropped to zero (and coding these periods with the pre-change PBD level) and (b) dropping all weeks where EUC benefits temporarily dropped to zero. Results remain largely the same across these specifications. We also consider the sensitivity of our standard errors to various clusters in Table A5. Results for our two primary specifications remain statistically significant after clustering by (a) state-date, (b) state-month, (c) cashier, (d) cashier-register, and (e) store.

2.5.3 Placebo Tests

Permutation test

In order to further test the robustness of our main results, we estimate our fully specified model across a variety of placebo treatments. We adopt the permutation test outlined by Bertrand, Duflo and Mullainathan (2004) and utilized in several studies including Chetty, Looney and Kroft (2009) and Ebenstein and Stange (2010). To perform the test, we estimate the preferred model after reassigning treatment status, and use the distribution of these "placebo" estimates for inference.

A benefit of such an approach is that no assumption is made on plausible serial correlation of the error term; instead, the “true” estimate is compared against many placebo estimates generated from reassignment of treatment. Since treatment patterns are assigned across three states, we simply consider reassignment of state-treatment statuses across our three states, and juxtapose the true estimate against the remaining five combinations of placebo estimates. These results are in Table 2.7. Of the six plausible combinations of state-treatment assignments to states, the true estimate of 0.130 is the largest.

Lack of PBD effect at self-checkout registers

In this section, we replicate the main analysis for a separate sample of registers. Before, we excluded transactions completed at self-checkout registers because cashiers do not process these transactions, and thus, these transactions do not provide a measure of cashier productivity. However, the data from self-checkout registers can be used as another type of placebo test. Specifically, we should expect no effect of PBD extensions on transaction length at self-checkout registers under the hypothesis that PBD strictly affects the cashiers (conditional on transaction-level controls).

In Table 2.8, we estimate a variant of equation (2.1) without the cashier c index and using scanner data from self-checkout registers. Of the 39 stores in the sample, only 7 have self-checkout registers—2 in DC, 2 in MD, and 3 in VA. The specification in column 1 includes date and register-store fixed effects. Column 2 adds the set of controls excluding those related to cashiers (such as cashier experience). Though this analysis suffers from considerably reduced statistical power, we find no evidence of a statistically significant relationship between PBD extensions and transaction length. Thus, reassuringly, we do not find effects where there are no cashiers, adding internal validity to our results above.

2.6 Shirking by Workers in the American Time Use Survey (ATUS)

In order to determine whether our results extend to other state-years, industries, and occupations, we use the American Time Use Survey (ATUS) to test for shirking responses to PBD extensions that occurred between 2003 and 2014. Exploiting the same variation described in section 2.3.1, we estimate models with various combinations of state fixed effects, month-year fixed effects, and a vector of controls. These controls include state unemployment rate, the maximum weekly benefit amount available to UI recipients in the state[48], the worker's age, "usual" amount of hours worked per week, weekly earnings, and dummies for family income, gender, race, type of US citizenship, class of worker (e.g., federal government vs. state government vs. private for profit), and general occupational category (e.g., "sales and related occupations" vs. "healthcare support occupation").

In Table 2.9, similar to Table 2.6, we investigate whether average worker characteristics are associated with state-month-year PBD levels. We find no statistically significant relationship across eight worker characteristics considered, including earnings and number of hours worked in the week prior. This helps alleviate concerns that PBD changes may lead to changes in worker composition, which are especially important in this analysis given that the ATUS is a cross-section.

Our main results with the ATUS sample are in Table 2.10. In three of the four specifications we find statistically significant (at the 10% level) increases in the percentage of time at work spent not working in response to more generous PBD levels. In our fully-specified model, we estimate a 0.35 percentage point increase (off a sample mean of 6.68%) in time spent not working in response to an 18-week increase in PBD level.[49]

We perform several additional analyses to test the sensitivity of these results. First, in Fig-

[48]Retrieved from the replication package for Hsu, Matsa and Melzer (2018).

[49]Since the outcome variable is bounded between zero and one, and the mean is relatively small (0.0668), we rescaled our PBD measure to be counts of 18-weeks (instead of counts of single weeks).

ure 2.4, we consider the permutation test outlined in Bertrand, Duflo and Mullainathan (2004), plotting the empirical distribution of estimated placebo treatment effects from 3,000 randomizations. For each randomization, workers are randomly assigned a state treatment pattern at the state level (without replacement). Results from this test suggest statistical significance at the 10% level for our fully specified model. Second, in Online Appendix Table A6, we find no statistically significant response in minutes spent at the workplace in response to higher PBD levels, suggesting that increased shift length cannot be driving these findings. Overall, these results suggest that the ex ante moral hazard effect observed in our cashier data is potentially pertinent across the US and in other sectors as well. Third, in Figure 2.5, we test for impacts of future and past PBD levels on current worker behavior in the ATUS data. Lagged effects suggest a type of delayed response by workers, while future PBD levels serve as a falsification test for our results. In panel A, we show that same week PBD and seven of the eight PBD lags correspond to an increase in transaction length that is statistically significant at the 10% significance level. This suggests that PBD in current and past periods affects current productivity, which we would expect if responses to PBD changes are delayed for some individuals. On the other hand, nearly all of the lead coefficients are statistically insignificant. Only the closest of the eight leads reports positive effects.[50,51]

To get a sense of the magnitude of the ATUS effects, we perform the following back-of-the-envelope calculations. Given the average worker in our sample spends 510.45 minutes at their workplace on the days they work, of which 31.83 minutes are spent not working,[52] a 0.35 percentage point increase in time spent not working at work translates to a 1.60 minutes increase in shirking per workday. For workers that work five days a week, this aggregates to an additional 6.9 hours of shirking per year. Further aggregating across all full-time workers in the US, an 18-week increase in

[50]We cannot be certain why this lead is significant; however, we think the size and significance of it could reflect current productivity responding to news about eminent changes in PBD levels.

[51]In panel B, we run a similar specification in the scanner data which produces analogous results.

[52]This estimate is similar to what the BLS (2015) reports for full-time workers in the US using the 2014 ATUS data.

the PBD level would lead to 773 million additional hours of shirking per year in the US—equivalent to $14 billion at the median hourly wage.[53]

We can compare the ATUS estimates with the cashier estimates in several ways. First, given the units between the two outcomes are different, we can compare our coefficients to the standard deviations of their respective outcomes and roughly interpret in standard deviation units. For an 18-week increase in PBD in the scanner data, our full model estimates a 2.34 percent of a standard deviation increase in transaction length. For ATUS, an 18-week increase in PBD is associated with a 2.84 percent of a standard deviation increase in percentage of time spent at work doing non-work activities. These estimates are markedly similar to each other.

We can also compare the estimates in terms of time lost. From the scanner data, the average transaction takes approximately two minutes to complete, and a cashier takes 2.4 seconds longer to complete a transaction in response to an 18-week increase in PBD. Assuming cashiers work a full 8-hour shift, the cashier would "lose" approximately 576 seconds (i.e., 240 possible transactions $\times$ 2.4 seconds per transaction), or 9.6 minutes, in response to an 18-week PBD increase. ATUS workers on the other hand are spending 1.6 minutes more doing non-work activities while at work. The "true" gap between these time estimates is likely to be smaller, however. Given the scanner data only look at peak processing times, it is likely that cashiers are not spending all of their time processing transactions during non-peak hours. Furthermore, our measure for ATUS workers involves non-work activities—it may be that these workers are additionally experiencing decreases in productivity for actual-work activities and thus "costing" additional time.

[53] There were 111,487,000 full-time workers 18 years or older in the US working 35 hours or more per week in 2017, and the US median hourly wage across all occupations was $18 in 2017 (BLS, 2017*a*; 2017*b*).

2.7 Conclusions

Numerous studies have investigated the ex post moral hazard effect of more generous UI benefits on unemployment duration. Despite strong theoretical evidence for its importance, empirical evidence of an *ex ante* moral hazard effect of UI (in which workers reduce on-the-job effort in response to increases in UI benefit generosity) remains scant.

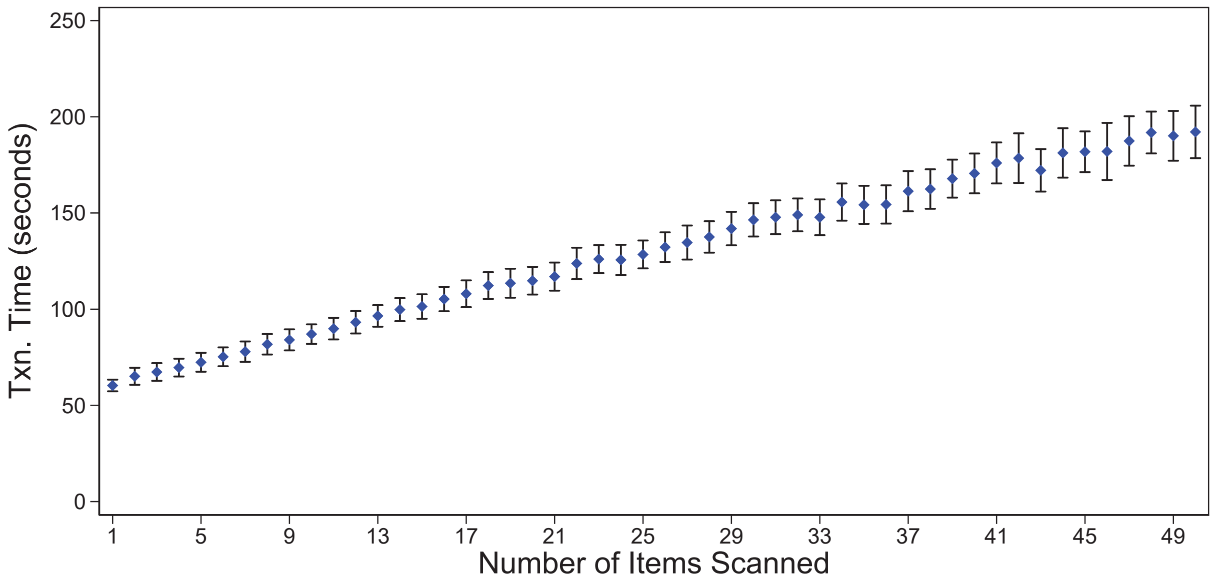
In this paper, we exploit state variation in the timing and size of potential benefit duration (PBD) of the UI program in the United States, occurring during the Great Recession, to provide estimates of the ex ante moral hazard effect of UI on worker productivity. Our scanner data consist of roughly 500,000 transactions which occurred at 39 locations of a large national supermarket chain in Maryland, Virginia, and Washington D.C. between November 2008 and February 2011. We estimate statistically significant negative effects of UI benefit duration on worker effort, where effort is measured by observing the length of time (in seconds) a cashier takes to complete a transaction. Our primary specifications utilize cashier-register and date fixed effects, as well as a series of transaction-level controls, to account for an array of potential confounding factors.

Preferred specifications suggest the average 18-week increase in PBD observed in our sample increases transaction time by roughly 2% of the sample mean. Though point estimates are modest, back-of-the-envelope calculations suggest non-trivial losses in time. In order to make up these productivity losses, each affected store would need to acquire over 144 additional hours of cashier labor per year. Our results are driven by the cashiers who are more likely to be terminated due to shirking (lower productivity cashiers) and by the cashiers more likely to be eligible for UI benefits (cashiers who worked more days during the sample period). Shirking is significantly attenuated by transactions on express registers, or those transactions of which there is likely less opportunity for cashiers to shirk. Results using a national cross-sectional survey of workers from the American Time Use Survey (ATUS) further corroborate this ex ante moral hazard effect.

Given the size and ubiquity of unemployment insurance programs, the potential policy implications for these results are substantial. Unemployment insurance programs exist in all OECD countries and are very large—in the United States, per capita expenditures on the UI program have exceeded those for all other safety net programs during each of the last four recessions (Bitler and Hoynes, 2016; Schmieder and Von Wachter, 2016). Our results suggest that, when evaluating the merits of benefit extensions, policymakers should consider the behavioral costs that are likely to occur not only among unemployed recipients of UI, but also among the employed who are potential future recipients.

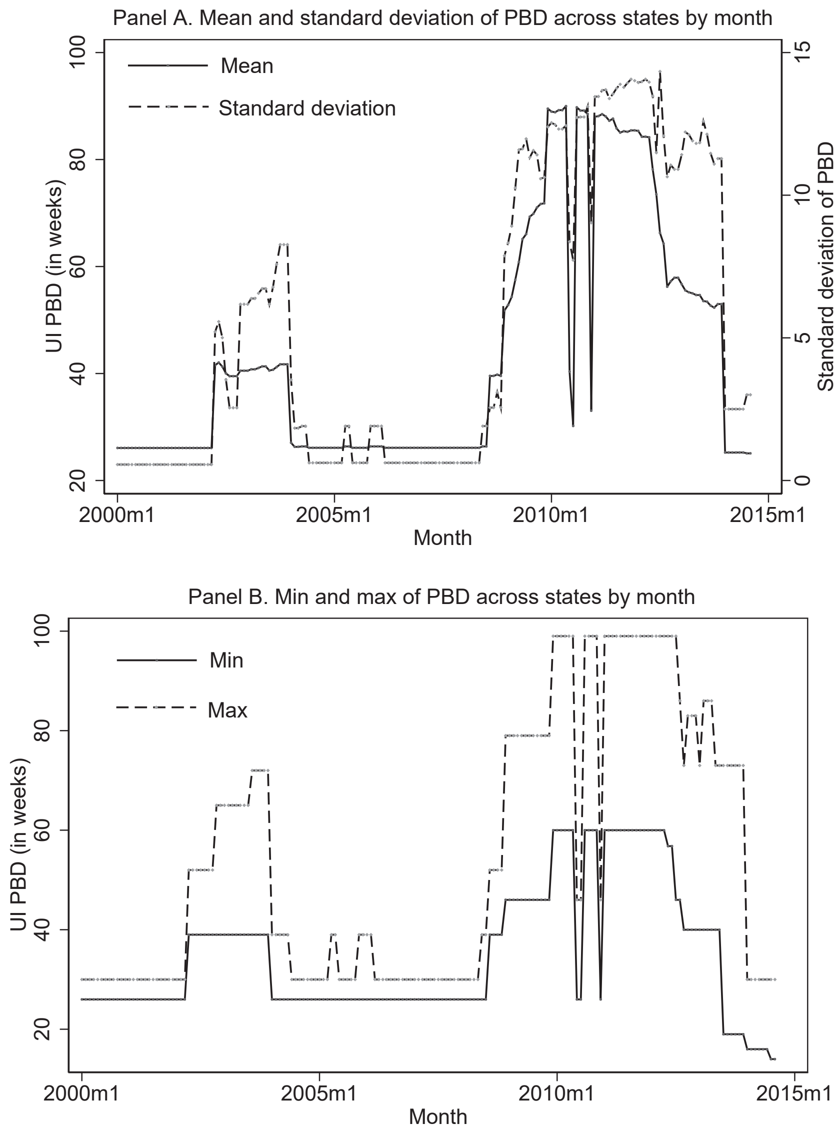
2.8 Figures and Tables

Figure 2.1: Relationship between the number of items scanned and average transaction time



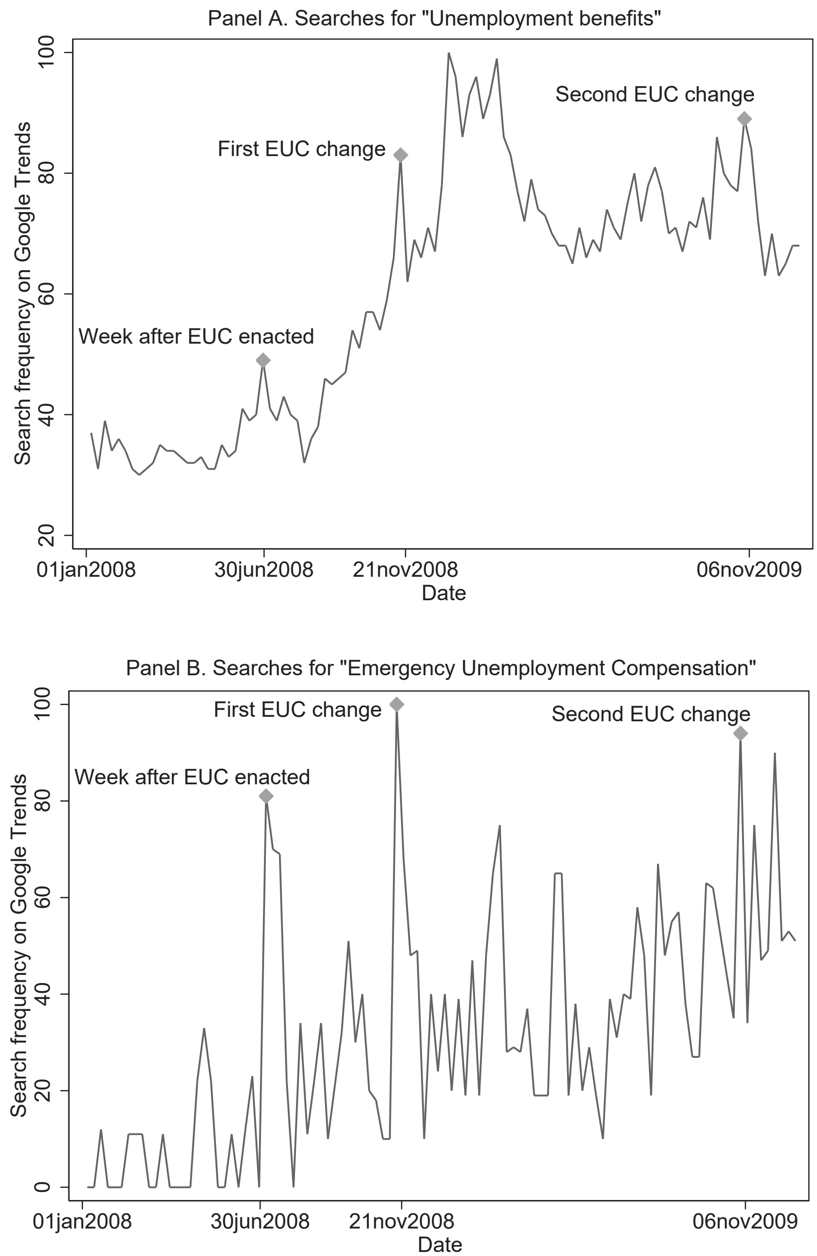
Notes: Estimates come from regressing transaction time on dummy variables for the number of items scanned per transaction, also controlling for the price and types of items purchased, as well as store, cashier, register and date fixed effects. The first 50 dummy variables are plotted, adjusted to include the intercept coefficient. Thus, the figure shows the average length of a transaction by transaction size. Upper and lower 95% confidence intervals are depicted, estimated using two-way cluster robust standard errors on store and date.

Figure 2.2: Trends in UI potential benefit duration (PBD) for ATUS sample



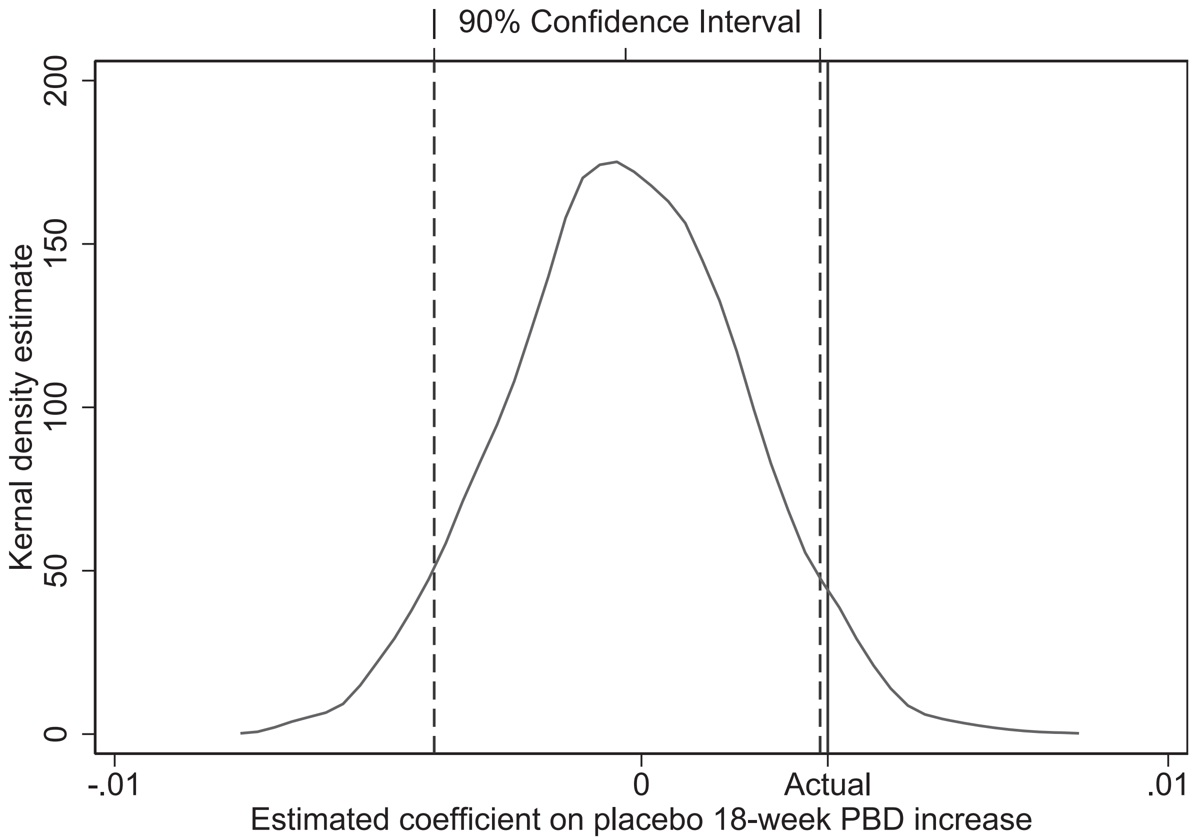
Notes: Data were obtained from a replication file for Farber, Rothstein and Valletta (2015).

Figure 2.3: Searches on Google via Google Trends



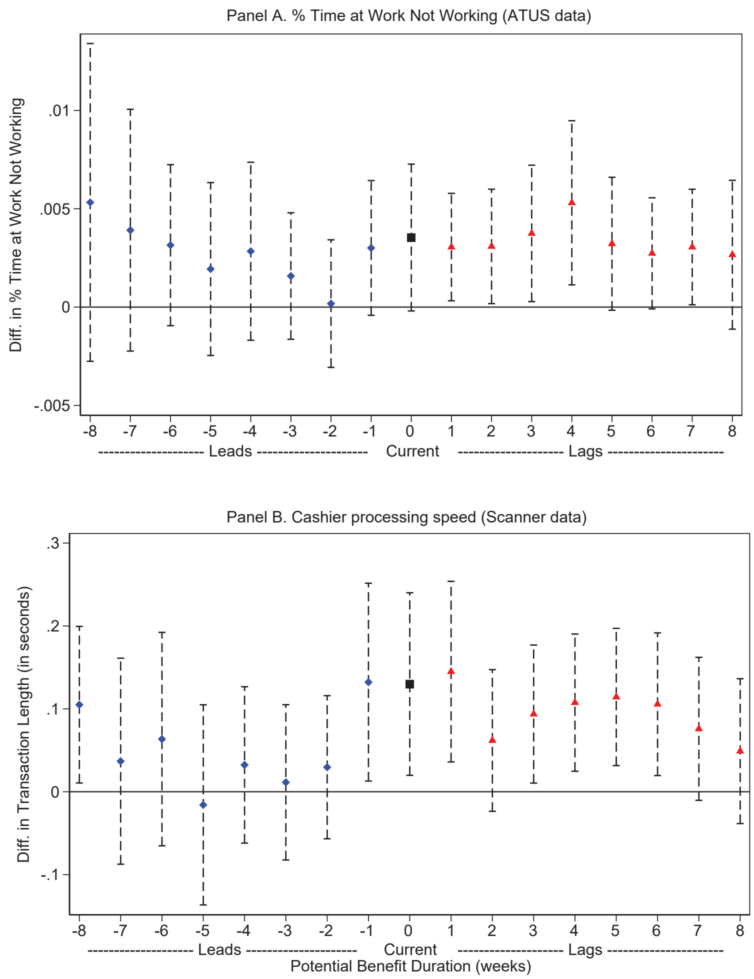
Notes: Google Trends data retrieved from Google Inc. Search frequency, indexed to a 0 to 100 scale, shows how often a particular search-item on Google Search (i.e. "Unemployment benefits" and "Emergency Unemployment Compensation") is entered relative to the total search-volume for the search-item across the queried time period (January 2008 - December 2009) within the United States. An index of 100 reveals the week(s) with the highest search frequency of that item within the queried time period.

Figure 2.4: Permutation test of inference with ATUS - % time at work not working



Notes: Figure plots the smoothed empirical distribution of estimated placebo treatment effects from 3,000 randomizations, where workers, by state, were randomly assigned state treatment patterns (without replacement). Dashed lines report the 90% confidence interval (the 5 and 95 percentiles of the distribution), while the solid line reports the actual point estimate. All estimates come from our fully specified model. Permutation test outlined in Bertrand, Duflo and Mullainathan (2004).

Figure 2.5: Impact of potential benefit duration lags and leads



Notes: Each point presents the β coefficient and its 95% confidence interval for separate regressions of an increase in PBD levels occurring w weeks ago on (a) the percentage of time the worker spent at work doing non-work activities and (b) transaction length. Lag coefficients (red triangles) represent delayed responses to PBD changes whereas lead coefficients (blue diamonds) serve as a falsification test. For example, the coefficient for w=1 corresponds to the impact of a PBD increase occurring in the week prior.

Table 2.1: Summary statistics from scanner data

	DC	Maryland	Virginia	Full Sample
Panel A. Sample characteristics, transaction level				
Transaction Time (in seconds)	128.06	122.11	111.38	119.62
	(107.68)	(102.77)	(90.79)	(100.03)
Total # of Items Scanned per Transaction	12.06	11.25	12.13	11.77
	(13.32)	(12.68)	(13.26)	(13.06)
Price Discounts/Coupons per Transaction	6.99	7.52	8.40	7.71
	(9.89)	(10.73)	(11.60)	(10.88)
Total Expenditure per Transaction	33.98	31.97	38.75	34.97
	(39.62)	(37.39)	(42.80)	(40.12)
# Items Returned per Transaction	0.00	0.00	0.00	0.00
	(0.07)	(0.08)	(0.07)	(0.07)
Observations	127,445	197,949	190,242	515,636
Panel B. Sample characteristics, cashier level				
Cashier Experience (in # of transactions in sample)	223.98	253.78	299.59	259.90
	(294.91)	(334.45)	(465.77)	(372.96)
Cashier Experience (span of days in sample)	210.28	245.39	227.03	229.44
	(241.15)	(257.03)	(258.31)	(253.28)
Cashier Experience (# of Saturday shifts in sample)	12.13	13.50	14.69	13.49
	(14.41)	(15.90)	(19.49)	(16.77)
Minutes Worked per Hour Shift	40.85	39.80	40.10	40.18
	(17.63)	(18.57)	(18.65)	(18.36)
Observations	569	780	635	1,984
Panel C. Sample characteristics, store level				
# of Registers per Store	8.13	6.47	7.00	7.00
	(2.64)	(1.28)	(1.96)	(1.92)
# of Unique Cashiers per Store	71.13	45.88	45.36	50.87
	(23.06)	(12.35)	(16.91)	(19.20)
Observations	8	17	14	39

Notes: Authors' calculations from scanner data. Cashier experience is measured as the total number of transactions observed in the sample for the given cashier prior to the current transaction. Cashier fatigue is measured as the total number of transactions during the current shift for the given cashier prior to the current transaction.

Table 2.2: Potential benefit duration (PBD) changes during sample period

	Extended Benefits	Emergency Unemployment Compensation	Total
Washington D.C.			
12/1/2008	0	33	59
4/5/2009	0	33	59
4/12/2009	20	33	79
5/3/2009	20	33	79
11/8/2009	20	53	99
4/5/2010	20	0	46
4/15/2010	20	53	99
6/2/2010	20	0	46
7/22/2010	20	53	99
11/30/2010	20	0	46
12/17/2010	20	53	99
Maryland			
12/1/2008	0	20	46
4/5/2009	0	20	46
4/12/2009	0	33	59
5/3/2009	0	33	59
11/8/2009	0	47	73
4/5/2010	0	0	26
4/15/2010	0	47	73
6/2/2010	0	0	26
7/22/2010	0	47	73
11/30/2010	0	0	26
12/17/2010	0	47	73
Virginia			
12/1/2008	0	20	46
4/5/2009	0	20	46
4/12/2009	0	20	46
5/3/2009	13	33	72
11/8/2009	13	47	86
4/5/2010	13	0	39
4/15/2010	13	47	86
6/2/2010	13	0	39
7/22/2010	13	47	86
11/30/2010	13	0	39
12/17/2010	13	47	86

Notes: Numbers represent maximum duration (in weeks) of UI benefits available during the time period beginning on the date in the first column. Total weeks are calculated as the sum of any EB extensions, EUC extensions, and the standard pre-extension PBD for all states (26 weeks).

Table 2.3: Main results from scanner data

	Transaction Length (in seconds)				
	(1)	(2)	(3)	(4)	(5)
Potential benefit duration	0.139	0.150	0.133	0.130	0.136
	(0.061)	(0.065)	(0.047)	(0.056)	(0.079)
Total expenditures			0.134	0.134	0.230
			(0.013)	(0.013)	(0.018)
Total items scanned			2.821	2.825	2.591
			(0.052)	(0.052)	(0.067)
Price discounts/coupons			0.720	0.726	0.506
			(0.024)	(0.024)	(0.031)
Local UE rate (prior month)			-0.678	-0.661	-0.977
			(0.429)	(0.463)	(0.576)
State UE rate (prior month)			-1.537	-1.354	-3.245
			(1.196)	(1.268)	(1.860)
Observations	515,636	515,618	515,618	515,433	354,281
Controls			X	X	X
Date FE	X	X	X	X	X
Register X Store FE	X	X	X		
Cashier X Store FE		X	X		
Cashier X Register X Store FE				X	X
Customer FE					X

Notes: Potential benefit duration is measured in weeks. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, the total number of registers open during the transaction, the cashier's experience as measured by total number of career transactions completed, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, and the cashier's length of shift measured in both number of transactions and in minutes. "Date" refers to exact date (e.g., August 3, 2017). Standard errors, shown in parentheses, are clustered at state by date level.

Table 2.4: Main results from scanner data by subsample

	Full	# of shifts		Productivity		Register type	
	Sample	High	Low	High	Low	Express	Regular
Outcome: Transaction Length							
Potential benefit duration	0.130	0.163	0.132	-0.084	0.173	-0.003	0.251
	(0.056)	(0.065)	(0.078)	(0.099)	(0.067)	(0.058)	(0.082)
Total expenditures	0.134	0.157	0.130	0.144	0.141	0.234	0.108
	(0.013)	(0.020)	(0.023)	(0.030)	(0.018)	(0.020)	(0.015)
Total items scanned	2.825	2.719	2.756	2.498	2.813	2.668	2.901
	(0.052)	(0.087)	(0.088)	(0.121)	(0.071)	(0.076)	(0.063)
Price discounts/coupons	0.726	0.707	0.778	0.567	0.789	0.644	0.760
	(0.024)	(0.034)	(0.043)	(0.050)	(0.033)	(0.033)	(0.029)
Local UE rate (prior month)	-0.661	0.110	-0.562	-0.584	-0.285	-0.839	-0.532
	(0.463)	(0.669)	(0.596)	(0.615)	(0.627)	(0.466)	(0.729)
State UE rate (prior month)	-1.354	1.450	-4.222	-1.387	-1.766	0.021	-2.704
	(1.268)	(1.941)	(2.010)	(2.831)	(1.722)	(1.349)	(1.944)
Observations	515,433	163,806	156,865	72,093	248,114	281,291	234,098
Controls	X	X	X	X	X	X	X
Date FE	X	X	X	X	X	X	X
Cashier X Register X Store FE	X	X	X	X	X	X	X

Notes: Potential benefit duration is measured in weeks. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, the total number of registers open during the transaction, the cashier's experience as measured by total number of career transactions completed, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, and the cashier's length of shift measured in both number of transactions and in minutes. In columns 2 and 3, shift subsamples are defined by cashiers working above the 75th percentile of shifts worked (high shifts) or below the 75th percentile (low shifts) before the first PBD change. In columns 4 and 5, productivity subsamples are defined by estimating cashier fixed effects from a regression of transaction length for the pre-policy period and separating by whether the cashier was above or below the 75th percentile fixed effect. "Date" refers to exact date (e.g., August 3, 2017). Standard errors, shown in parentheses, are clustered at state by date level.

Table 2.5: Is UI potential benefit duration correlated with total supermarket sales?

	Sales ($)	Sales ($)	Sales (# Items)	Sales (# Items)	$ Spent Per Item	$ Spent Per Item
18-week PBD increase	-14.734	184.010	-39.442	18.571	0.000	0.001
	(82.492)	(71.742)	(30.233)	(30.698)	(0.000)	(0.000)
State UE rate (prior month)		-7,011.488		-2,046.647		-0.011
		(1548.421)		(539.266)		(0.004)
Observations	2,381,922	2,381,922	2,381,922	2,381,922	2,381,922	2,381,922
Mean of Y	563,245	563,245	184,385	184,385	3.33	3.33
Store FE	X	X	X	X	X	X
Month-Year FE	X	X	X	X	X	X

Notes: Nielsen Scanner data aggregated to the store-month level from 2009–2014 for 39,622 grocery, drug, and mass merchandiser stores in the US. Outcome variables are at the store-month level: total sales in dollars (columns 1–2), total sales in items sold (columns 3–4), and average expenditures per item (columns 5–6). Models include store and month-year fixed effects. Standard errors, shown in parentheses, clustered at the state level.

Table 2.6: Is UI potential benefit duration correlated with other covariates?

Panel A	Expend per Txn.	Items Scanned per Txn	Price Discount & Coupons	# Open Registers	# Items Returned	Alcohol & Tobacco
PBD	-0.005	0.003	0.014	-0.006	-0.000	0.001
	(0.025)	(0.008)	(0.010)	(0.008)	(0.000)	(0.000)
Observations	4,407	4,407	4,407	4,407	4,407	4,407
Y mean	33.904	11.488	7.466	5.228	0.002	0.133
Date FE	X	X	X	X	X	X
Store FE	X	X	X	X	X	X
Panel B	Bakery & Deli	Dairy	Floral	Frozen	Meat & Seafood	Produce
PBD	-0.000	0.000	-0.000	-0.001	-0.000	-0.000
	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.002)
Observations	4,407	4,407	4,407	4,407	4,407	4,407
Y mean	0.371	1.150	0.048	0.612	0.712	1.638
Date FE	X	X	X	X	X	X
Store FE	X	X	X	X	X	X
Panel C	Lag Local UE Rate	Lag State UE Rate	# Employed Cashiers	Cashier Experience	Food Stamps	Bag Tax
PBD	-0.003	-0.001	-0.007	-0.561	0.624	0.017
	(0.012)	(0.001)	(0.010)	(1.131)	(0.400)	(0.005)
Observations	4,407	4,407	4,407	4,407	4,407	4,407
Y mean	6.074	7.442	6.073	492.275	249.411	0.105
Date FE	X	X	X	X	X	X
Store FE	X	X	X	X	X	X

Notes: Each cell reports a coefficient from a single regression of potential benefit duration on an outcome, collapsed to the store-date level. Potential benefit duration is measured in weeks. "Expend per Txn." is the average amount spent per transaction, measured in dollars. "Items Scanned per Txn." is the average number of items scanned per transaction. "Price Discount & Coupons" is the average dollar amount of price discounts per transaction (from sale items, coupons, and reward cards). "# Open Registers" is the average number of open registers per store-date. "# Items Returned" is the number of items scanned and then un-scanned (i.e., returned) per transaction. The seven department categories are the average number of products bought per transaction by category. The "Lag Local and State UE Rates" are the one month lagged county-month and state-month level unemployment rates. "# Employed Cashiers" is the number of cashiers working per store-date. "Cashier experience" is the average length of time the cashier appears in the sample. "Food Stamps" is the number of households participating in SNAP per state-month, measured in thousands. "Bag Tax" is an indicator for whether a bag tax was in place per store-date. Standard errors clustered at store level are shown in parentheses.

Table 2.7: Placebo tests - Reassignment of treatments across states

	Actual	Placebos (1)	(2)	(3)	(4)	(5)
Outcome: Transaction Length						
Potential benefit duration	0.130	-0.009	0.075	-0.052	-0.012	-0.153
	(0.056)	(0.066)	(0.063)	(0.057)	(0.065)	(0.068)
Total expenditures	0.134	0.134	0.134	0.134	0.134	0.134
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Total items scanned	2.825	2.825	2.825	2.825	2.825	2.825
	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)	(0.052)
Price discounts/coupons	0.726	0.726	0.727	0.726	0.726	0.726
	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)	(0.024)
Local UE rate (prior month)	-0.661	-0.756	-0.741	-0.726	-0.749	-0.703
	(0.463)	(0.463)	(0.462)	(0.465)	(0.462)	(0.463)
State UE rate (prior month)	-1.354	-0.744	-0.770	-1.006	-0.862	-0.270
	(1.268)	(1.211)	(1.222)	(1.233)	(1.287)	(1.195)
Observations	515,433	515,433	515,433	515,433	515,433	515,433
# of Treatment Swaps	0	1	1	1	2	2
Controls	X	X	X	X	X	X
Date FE	X	X	X	X	X	X
Cashier X Register X Store FE	X	X	X	X	X	X

Notes: Potential benefit duration is measured in weeks. Columns (1) through (5) consider all five remaining permutations of swaps of PBD levels by state. In (1), Washington D.C. and Virginia are swapped. In (2), Washington D.C. and Maryland are swapped. In (3), Virginia and Maryland are swapped. In (4), Virginia is assigned Maryland PBD levels, Maryland to D.C. levels, and D.C. to Virginia levels. In (5), Virginia is assigned D.C. PBD levels, Maryland to Virginia levels, and D.C. to Maryland levels. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, the total number of registers open during the transaction, the cashier's experience as measured by total number of career transactions completed, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, and the cashier's length of shift measured in both number of transactions and in minutes. "Date" refers to exact date (e.g., August 3, 2017). Standard errors, shown in parentheses, are clustered at state by date level.

Table 2.8: Placebo test - Results from self-checkout scanner data

	Transaction Length (in seconds)	
	(1)	(2)
Potential benefit duration	0.386	-0.001
	(0.492)	(0.491)
Total expenditures		-0.143
		(0.085)
Total items scanned		6.985
		(0.389)
Price discounts/coupons		1.516
		(0.146)
Observations	44,646	44,646
Controls		X
Date FE	X	X
Register X Store FE	X	X

Notes: This table uses scanner data only from self-checkout registers. Seven of the 39 stores have self-checkout registers. Potential benefit duration is measured in weeks. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, lagged county-month level unemployment rate, lagged state-month level unemployment rate, and the total number of registers open during the transaction. "Date" refers to exact date (e.g., August 3, 2017). Standard errors, shown in parentheses, are clustered at state by date level.

Table 2.9: Is UI potential benefit duration correlated with other ATUS covariates?

Panel A	Age	Female	White	Weekly Earnings	Usual # Work Hours
PBD	0.050	0.008	-0.001	-2.727	0.032
	(0.238)	(0.013)	(0.009)	(13.894)	(0.215)
Observations	6,041	6,041	6,041	6,041	6,041
Mean of Y	41.74	0.50	0.82	880.50	41.94
State FE	X	X	X	X	X
Month-Year FE	X	X	X	X	X
Panel B	Gov't Sector	Private Sector	Max UI Benefit	Lagged UE Rate	Work Hours Prior Week
PBD	0.007	-0.007	-0.016	1.022	-0.335
	(0.009)	(0.009)	(0.033)	(0.118)	(0.279)
Observations	6,041	6,041	6,041	6,041	5,963
Mean of Y	0.19	0.81	4.06	6.40	40.33
State FE	X	X	X	X	X
Month-Year FE	X	X	X	X	X

Notes: Each cell reports a coefficient from a single regression of potential benefit duration on an outcome, collapsed to the state-month-year level. Potential benefit duration is measured in weeks. "Age" is the average age of workers in our ATUS sample. "Female" is the fraction of workers who were female. "Usual # Work Hours" is the average number of self-reported weekly work hours. "Weekly Earnings" is the average worker weekly earnings in dollars. "White" is the fraction of workers who were White. "Gov't Sector" and "Private Sector" are the fraction of workers in the government vs. the private sector, respectively. "Max UI Benefit" and "Lagged UE Rate" are state-month-year maximum UI benefits and prior month unemployment rates, respectively. "Work Hours Prior Week" is the average number of work hours from the worker's week prior to completing the CPS. Standard errors, shown in parentheses, are two-way clustered at the state and month-year level.

Table 2.10: Results from American Time Use Survey (ATUS)

	% Time At Work Not Working			
	(1)	(2)	(3)	(4)
18-week PBD increase	0.0022	0.0024	0.0025	0.0035
	(0.0013)	(0.0017)	(0.0015)	(0.0019)
State UE rate (prior month)			-0.0001	-0.0006
			(0.0010)	(0.0010)
Maximum WBA (100s)			0.0046	0.0042
			(0.0029)	(0.0030)
Observations	30,094	30,094	30,094	30,094
Mean of Y	0.0668	0.0668	0.0668	0.0668
State FE	X	X	X	X
Month FE	X		X	
Year FE	X		X	
Month-Year FE		X		X
Controls			X	X

Notes: Controls include state unemployment rate and maximum UI benefits (in dollars), the individual's age, "usual" amount of hours worked per week, weekly earnings, hourly wage, and dummies for family income, gender, race, US citizenship, whether the individual had multiple jobs, class of worker (e.g., federal government vs. state government vs. private for profit), and general occupational category (e.g., "sales and related occupations" vs. "healthcare support occupations"). Observations weighted according to ATUS probability weights. Standard errors, shown in parentheses, clustered at state level.

Appendix

2.A Theoretical Model

Although the comparative static of interest is straightforward, and has been previously established in the literature (e.g., Shapiro and Stiglitz, 1984), in this appendix we lay out a simple theoretical model for a worker's choice of effort while on the job. The model makes clear the key assumptions required for an ex ante moral hazard effect of UI to exist, and helps to suggest the types of workers who are expected to respond ex ante to changes in UI benefits.

Consider a worker who chooses effort, e, to maximize expected utility:

$$E(U) = (1 - p(e))U(C_e) + p(e)U(C_u) \tag{2.2}$$

where $p(e)$ is the probability that worker is fired (decreasing in e), C_e is consumption while employed, C_u is consumption while unemployed and $U(\cdot)$ is increasing and concave. We make the following additional assumptions:

1. $p''(e) > 0$
2. $C_e = w - e$, where w is the wage
3. $\frac{\partial C_u}{\partial b} > 0$ & $\frac{\partial C_u}{\partial d} > 0$, where b is UI benefit level and d is UI benefit duration
4. $C_e > C_u$

The first order condition is:

$$(1 - p(e))U'(C_e) = -p'(e)(U(C_e) - U(C_u)) \tag{2.3}$$

where the left-hand side is the marginal cost of an increase in effort and the right-hand side is the marginal benefit of an increase in effort. The second order condition is:

$$p''(e)(U(C_u) - U(C_e)) + 2p'(e)U'(C_e) + (1 - p(e))U''(C_e) \equiv S(\cdot) \tag{2.4}$$

Applying the implicit function theorem to the FOC and denoting e^* the optimal effort:

$$\frac{\partial e^*}{\partial C_u} = -\frac{p'(e)U'(C_u)}{S(\cdot)} \tag{2.5}$$

The assumptions ensure that $S(\cdot)$ and $\frac{\partial e^*}{\partial C_u}$ are negative so that an increase in UI benefits or duration will decrease effort.

An implicit assumption, which clearly holds in the context of supermarket cashiers, is that employers partially observe effort (in order for $\frac{\partial p(e)}{\partial e} < 0$ to hold). Cases in which $p'(e)$ violates the above assumptions can provide some intuition for expected heterogeneity in equation 4. Consider a worker who cannot be fired. This worker has $p(e) = 0$ $(\forall e)$ and does not change e^* in response to ΔC_u. A worker with slightly less strong employment protection will have very small $|p'(e)|$ and a weak, but still negative, relationship between C_u and e. Although the workers in our setting are unionized, past work with data from this supermarket chain has observed that these workers can be fired if they are perceived as under-performing (see Mas and Moretti, 2009). Assumption (1) implies that there are "decreasing returns" to effort. This seems reasonable in most cases and is necessary for $\frac{\partial e^*}{\partial C_u} < 0$ to *always* hold. $\frac{\partial e^*}{\partial C_u} < 0$ will still often hold with concave $p(e)$, depending on the relative magnitude of the terms in the SOC.

We do not model the optimal e^* from the employer's or social planner's perspective. Therefore, we do not explicitly define shirking and we use the terms "a decrease in effort" and "an increase in shirking" interchangeably. A general equilibrium approach would model the employer's choice of wage offers and it is worth considering whether or not such employer responses affect the partial equilibrium relationships that we estimate. It is at least possible for both employers and customers to foresee changes in worker effort provision in response to UI benefit changes. We investigate these possibilities by looking for changes in cashier characteristics and transaction characteristics in response to PBD changes. Concerns about employer responses are also partially reduced by observations in past work with data from this supermarket chain which suggest that workers are primarily responsible for choosing their own shifts (Mas and Moretti, 2009).

2.B Unemployment Insurance Program Extensions in the US

2.B.1 The Extended Benefits Program

The EB program is state run and has existed since 1970. Under EB, a state's PBD is extended by either 13 or 20 weeks if the state's 13-week average Insured Unemployment Rate (IUR) or 3-month average Total Unemployment Rate (TUR) meet certain threshold, or "trigger," levels. The TUR is simply the ratio of the number of unemployed workers to the total number of workers in the state. The IUR is the ratio of UI claimants to the total number of workers in UI-eligible jobs in the state. All states are required to provide an additional 13 weeks of UI benefits if the IUR is at least 5.0% and at least 120% of the average of the state's IURs for the same 13 week period during the past 2 years. In addition to the this, states decide whether to follow one, both, or neither of the following optional triggers:

1. If the IUR is at least 6.0% (regardless of past IURs) then an additional 13 weeks of benefits are made available. This is known as the "IUR option."

2. If the TUR is at least 6.5% and at least 110% of the same TURs in either of the prior 2 years, then an additional 13 weeks of benefits are made available. Additionally, if the TUR is at least 8% and at least 110% of the same TURs in either of the prior 2 years, then an additional 20 weeks of benefits are made available (for 20 weeks total of EB, not 33). This is known as the "TUR option."

The EB program was originally financed 50% by states and 50% by the federal government. However, starting on February 17, 2009, the American Recovery and Reinvestment Act (ARRA) temporarily made the EB program fully federally financed. This additional federal financing remained in effect through the entirety of our sample. The 2-year "look-back" timeframe present in several of the threshold rules was temporarily changed to a 3-year period in December 2010, and this change also remained in effect throughout the remainder of our sample (Whittaker and Isaacs, 2013; Marinescu, 2017).

2.B.2 The Emergency Unemployment Compensation Program

The EUC program was enacted by the federal government as a response to the Great Recession and was federally run and funded throughout its existence. First established by the Emergency Unemployment Compensation Act on June 30, 2008, the EUC program originally provided 13 weeks of additional eligibility in all states. The design of the EUC program was changed twice during the Great Recession. On November 21, 2008 the EUC was given a two tier structure, 20 weeks of additional eligibility was provided for all states in tier 1 and an additional 13 weeks was provided for states with a TUR $\geq 6\%$ or a IUR $\geq 4\%$. On November 6, 2009 the second tier was increased to 14 weeks and given to all states regardless of TUR or IUR, a third tier was created providing 13 weeks to states with TUR $\geq 6\%$ or a IUR $\geq 4\%$, and a fourth tier was created providing 6 weeks to states with TUR $\geq 8.5\%$ or a IUR $\geq 6\%$ (Whittaker and Isaacs, 2014; Marinescu, 2017). The tiers in each of these iterations are cumulative, so that after November 6, 2009 in a state that selected the TUR option for the EB program, the maximum possible PBD available included the original 26 weeks, 20 weeks of EB, 20 weeks of EUCI, 14 weeks of EUCII, 13 weeks of EUCIII, and 6 weeks of EUCIV (for a total of 99 weeks).

As a temporary program EUC was originally given an expiration date of March 28, 2009. Congress extended the program multiple times so that it did not expire indefinitely until well after our sample ends. However, on four separate occasions during our sample (in March, April, June, and November of 2010) Congress failed to extend the program before its previous expiration date so that there were temporary lapses in EUC availability. The first two of these lapses were short (2 and 10 days respectively) while the latter two were relatively long (nearly 2 months).

2.B.3 The Temporary Extension of Unemployment Compensation Program

The TEUC program, also federally run and funded, was available to new claimants between March 2002 and December 2003.[54] Benefits continued to be available for existing but unexhausted TEUC claims into early 2004. The TEUC program extended UI benefits for either 13 or 26 weeks, with the additional 13 weeks (second tier) of benefits available in states with an IUR (13 week) of at least 4% and at least 120% higher than in the same time period during the prior two years (Valletta, 2014).

2.B.4 Additional Detail on Extensions in Scanner Data Sample

As described briefly in section 2.3.1, the PBD extensions we exploit for identifying variation occur for one of three reasons: (1) a state's unemployment rate crosses a threshold or "trigger" value currently in place (see first subsection of this appendix for specific unemployment rate and trigger values used), (2) the relevant authority (state government for EB, federal for EUC or TEUC) changes the trigger value to a level below the state's current unemployment rate, or (3) the federal government alters the (EUC or TEUC) program by changing the number of weeks available or allowing the program to expire (either temporarily or permanently). Here we provide additional narrative detail for each of the extensions occurring in our scanner data sample (ignoring EUC program lapses 4/5/2010-4/14/2010, 6/2/2010-7/21/2010, and 11/30/2010-12/16/2010). The dates and PBD levels for each of these changes are shown in Table 2. Sources for the information provided below are the EB and EUC trigger notices made available online by the US Department of Labor.[55]

1. Washington D.C., 4/12/2009, EB: Number of weeks available through the EB program in-

[54] Variation in PBD from the TEUC program is only used in our ATUS analyses, since the program does not overlap with our scanner data sample.

[55] See the Office of Unemployment Insurance website, Online, *accessed 14 Sep. 2018*.

creases from 0 to 20. This change resulted directly from D.C. adopting the TUR option. The TUR in D.C. exceeded both trigger values (13 week and 20 week) under the TUR option, but was below the IUR trigger value.

2. Maryland, 4/12/2009, EUC: Number of weeks available through the EUC program increases from 20 to 33. This change resulted from MD's TUR crossing the threshold value of 6%. (During this time period the second tier of EUC benefits provided an additional 13 weeks to states with TUR $\geq$ 6% (Isaacs and Whittaker, 2014).)

3. Virginia, 5/3/2009, EB: Number of weeks available through the EB program increases from 0 to 13. This change resulted directly from VA adopting the TUR option. The TUR in VA exceeded the 13 week trigger value under the TUR option, but was below the IUR trigger value and the 20 week TUR trigger value.

4. Virginia, 5/3/2009, EUC: Number of weeks available through the EUC program increases from 20 to 33. This change resulted from VA's TUR crossing the threshold value of 6%. (During this time period the second tier of EUC benefits provided an additional 13 weeks to states with TUR $\geq$ 6% (Isaacs and Whittaker, 2014).

5. Washington D.C., 11/8/2009, EUC: Number of weeks available through the EUC program increases from 33 to 53. This change resulted from a policy change at the federal level which restructured the EUC program, increasing the number of weeks available through the EUC's second tier to 14 (from 13), and creating third (13 weeks), and fourth (6 weeks) tiers. The TUR in D.C. exceeded the threshold value for the third and fourth tiers (Isaacs and Whittaker, 2014).

6. Maryland, 11/8/2009, EUC: Number of weeks available through the EUC program increases from 33 to 47. This change resulted from a policy change at the federal level which restruc-

tured the EUC program, increasing the number of weeks available through the EUC's second tier to 14 (from 13), and creating third (13 weeks), and fourth (6 weeks) tiers. The TUR in MD exceeded the threshold value for the third tier but not the fourth (Isaacs and Whittaker, 2014).

7. Virginia, 11/8/2009, EUC: Number of weeks available through the EUC program increases from 33 to 47. This change resulted from a policy change at the federal level which restructured the EUC program, increasing the number of weeks available through the EUC's second tier to 14 (from 13), and creating third (13 weeks), and fourth (6 weeks) tiers. The TUR in VA exceeded the threshold value for the third tier but not the fourth (Isaacs and Whittaker, 2014).

2.B.5 Changes to State Regular PBD During ATUS Sample

Between 4/2011 and 8/2014 the states of Arkansas, Florida, Georgia, Illinois, Michigan, Missouri, North Carolina, and South Carolina each passed legislation reducing their regular PBDs below 26 weeks (Isaacs, 2019). This variation is not relevant for our scanner data sample but is utilized in our ATUS analyses. Here we provide additional detail on each of these policy changes, listed in order of the month that the relevant PBD change is first recorded in our data. Unless otherwise noted sources are the Department of Labor's Reports on State UI Legislation.[56].

1. Arkansas, 4/2011: Arkansas Senate Bill 593 reduced AR's PBD to 25 weeks. (See 2011 report #5.)

2. Missouri, 5/2011: Missouri House Bill 163 reduced MO's PBD to 20 weeks (Johnston and Mas, 2018).

[56]See the Office of Unemployment Insurance website, Online, *accessed 12 Mar. 2020*

3. <u>South Carolina, 7/2011</u>: South Carolina House Bill 3672 reduced SC's PBD to 20 weeks. (See 2011 report #6.)

4. <u>Florida, 1/2012</u>: Florida House Bill 7005 reduced FL's PBD to between 12 and 23 weeks depending on the state's unemployment rate. Specifically, the PBD of UI in FL is updated up to once annually on January 1st based on the unemployment rate in the state during the third quarter of the previous year. During our ATUS sample, FL's PBD decreased from 26 to 23 weeks in 1/2012, to 19 weeks in 1/2013, and to 16 weeks in 1/2014. (See 2011 report #5.)

5. <u>Illinois, 1/2012</u>: Illinois House Bill 1030 reduced IL's PBD to 25 weeks. (See 2011 report #7.)

6. <u>Michigan, 2/2012</u>: Michigan House Bill 4408 reduced MI's PBD to 20 weeks. (See 2011 report #2.)

7. <u>Georgia, 7/2012</u>: Georgia House Bill 347 reduced GA's PBD to between 14 and 20 weeks depending on the state's unemployment rate. Specifically, the PBD of UI in GA is updated up to twice annually on January 1st and July 1st based on the unemployment rate in the state during the previous October and April, respectively. During our ATUS sample, GA's PBD decreased from 26 weeks to 19 weeks in 7/2012, to 18 weeks in 7/2013, and to 15 weeks in 7/2014. (See 2012 report #1.)

8. <u>North Carolina, 7/2013</u>: North Carolina House Bill 3672 reduced NC's PBD to between 12 and 20 weeks depending on the state's unemployment rate. Specifically, the PBD of UI in NC is updated up to twice annualy on January 1st and July 1st based on the unemployment rate in the state during the previous October and April, respectively. During our ATUS sample, NC's PBD decreased from 26 to 19 weeks in 7/2013, and to 14 weeks in 7/2014. (See 2013 report #10.)

2.C Additional Evidence of Awareness of UI Benefit Extensions

This appendix uses Google Trends and national polls to provide additional evidence of general awareness about UI benefit extensions during the Great Recession. Though Google Trends does not report raw search numbers, they do allow comparison of popularities across five search items per query.[57] By scaling across all five search items and 104 weeks, one can compare search indices within weeks to get a better sense of the absolute popularity of a particular search item. In Online Appendix Figure 2.A2, we conduct three separate five-item searches, juxtaposing "Unemployment benefits" against four other popular searches. In the first panel, we compare searches for "Unemployment benefits" against "Disability insurance," "Food stamps," "Pell," and "Recession." Across our time frame, "Unemployment benefits" was a more popular search term than each of these four items. People searched for "Unemployment benefits" roughly twice as often as "Food stamps." The term with the largest search volume was "Recession" in January 2008 (at the onset of the Great Recession), and yet the popularity of this search was only slightly greater than the popularity for searches for "Unemployment benefits" during the ARRA implementation. In latter 2009, people searched "Unemployment benefits" at nearly three times the rate of the term "Recession." In the second panel of Online Appendix Figure 2.A2, we compare "Unemployment benefits" to "Earned Income Tax Credit," "Social security," "Welfare," and "TANF." Again, "Unemployment benefits" was one of the more popular search items, with "Social security" being only slightly more popular on average.

Finally, in order to compare the absolute popularity of "Unemployment benefits" to non-economics terms, in the third panel, we include the search terms "Disneyland," "Eiffel Tower," "Wall-E," and "Summer camp." Once again, "Unemployment benefits" was one of the more pop-

[57]Though the Google Trend's scale cannot be mapped into total search volume on Google Search, estimates do exist on the popularity of Google Search overall. For instance, roughly 3.5 billion searches are made per day. From 2008 to 2009, there were nearly 1.4 trillion total searches made on Google Search. Source: WordStream, Online, *access 31 Jul. 2019.*

ular search items during this time period. Wall-E was one of the most popular movies in 2008; during the week of Wall-E's peak search-popularity in June of 2008, people still searched for "Unemployment benefits" at roughly 20% the frequency of "Wall-E" (i.e. for every five searches for "Wall-E," there was one search for "Unemployment benefits"). Searches for "Summer camp" are unsurprisingly cyclical, yet during the summer of 2009, these searches seldom exceeded searches for "Unemployment benefits." During the first EUC change and the ARRA period, search volume for "Unemployment benefits" is comparable to "Disneyland." Searches for "Unemployment benefits" roughly double the amount of searches for "Eiffel Tower," despite the Eiffel Tower being the fifth most searched item on Google Maps.

To further understand workers' awareness of UI benefits during the Great Recession, we also examine polls that were conducted during these years. Since 2001, Gallup has surveyed Americans about their top concerns (e.g., crime and violence, drug use, hunger and homelessness, the economy, unemployment).[58] In March 2008 (six months before Lehman Brothers went bankrupt), 36% of respondents answered that they worry a great deal about unemployment. By March 2010, this had increased to 59%. Those worrying a great deal remained above 50% for the next three years and then steadily declined to 23% in 2018. Thus, UI benefit extensions came during a time when Americans were highly concerned about unemployment. In a poll more closely related to UI extensions, YouGov/Huffington Post surveyed 1000 U.S. adults in April 2014 about unemployment benefits extensions.[59] When asked—"How much have you heard about Congress letting unemployment benefits expire for people who have been unemployed more than six months at the end of last year?"—23% responded that they had heard a lot, 45% had heard a little, and 32% had heard nothing at all. This poll provides suggestive evidence that a majority of Americans had some level of awareness about extended UI benefits.

[58]Source: Gallup, Online, *accessed 3 May 2018*.
[59]Source: *YouGov.com*, Poll Results: Unemployment, April 18–21, 2014, Online, *accessed 3 May 2018*.

2.D Length of Unemployment Spells

The PBD extensions that we exploit in our analysis will only directly affect unemployed workers who remain unemployed for longer than 46 weeks. Thus, we should only expect a shirking response to these extensions if workers believed that there was a meaningful chance of suffering an unemployment spell longer than 46 weeks.

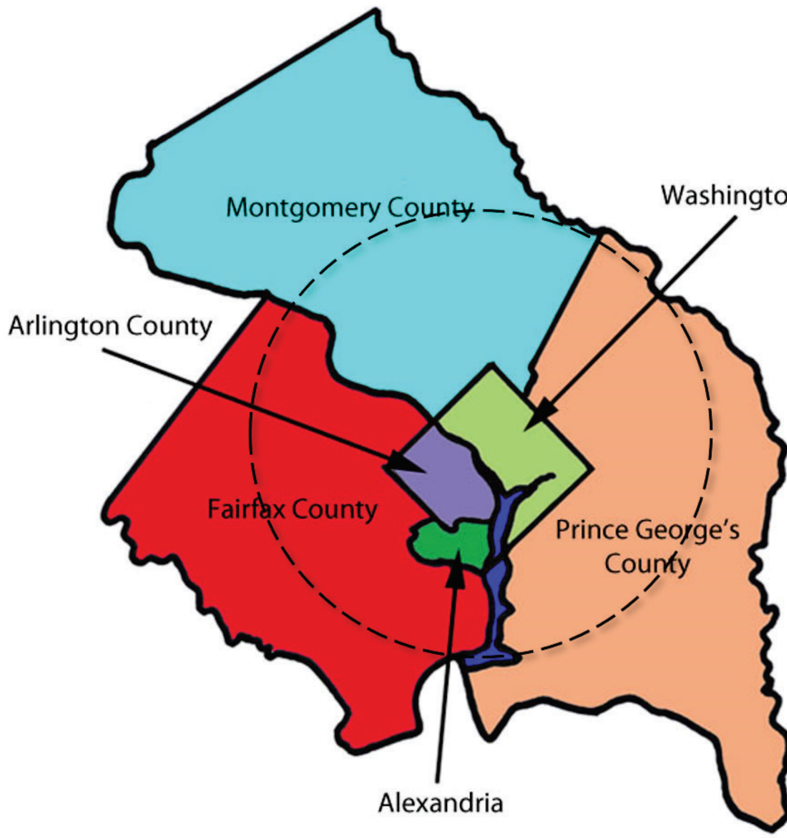
From the basic CPS monthly files for the months in our scanner data sample (December 2008 to February 2011), we extract a sample of 4,031 unemployed adult workers who resided in the Washington D.C. metropolitan area (Flood et al., 2017). The average duration of unemployment at the time of the survey was 29 weeks with a median of 18 weeks, while the 75th percentile of the distribution of unemployment duration was 43 weeks.[60] The lengths of unemployment spells are increasing drastically during this time (e.g., the overall mean increases from 24 weeks in the first half of our sample to 34 weeks in the second half) and this is consistent with what is seen nationally.[61] These estimates of unemployment durations are based on unadjusted samples of the stock of unemployed workers, and may be underestimating the true length of the unemployment spell due to right censoring (Kiefer, Lundberg and Neumann, 1985). Thus, it is reasonable to conclude that a low-skilled worker in the Washington D.C. metro area during the time period of our sample would have been concerned with the possibility of long term unemployment.

[60] Relevant statistics split by various subsamples are also plotted in Online Appendix Figure 2.A4.

[61] According to the Federal Reserve Economic Data (FRED), national mean unemployment durations nationally increased from 20 weeks to 37 weeks during our sample. (US Bureau of Labor Statistics, Average Weeks Unemployed [UEMPMEAN], retrieved from FRED, Federal Reserve Bank of St. Louis; Online, *accessed 15 Mar. 2020.*)

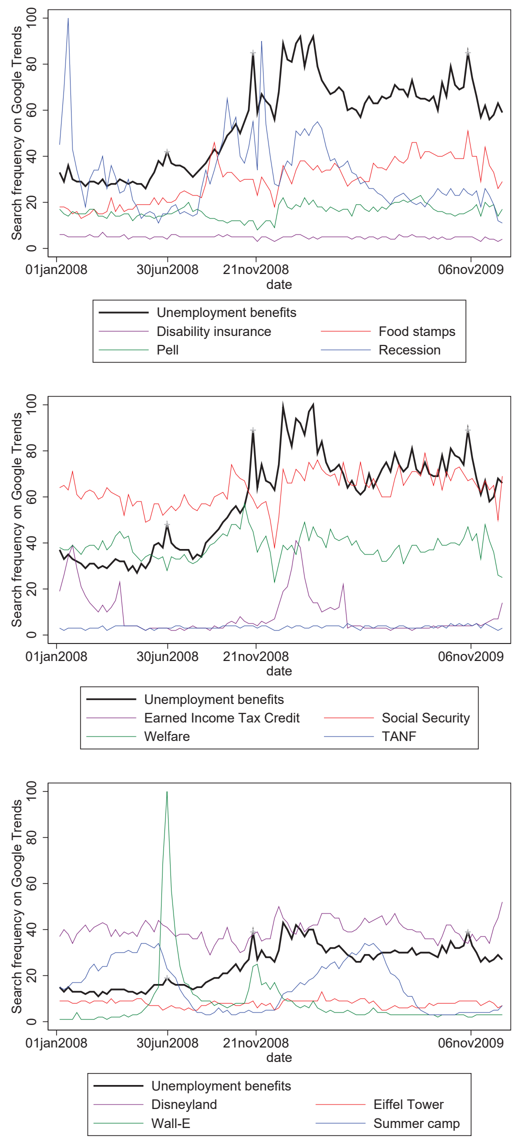
2.E Additional Figures and Tables

Figure 2.A1: Map of the Washington DC Metropolitan Area



Notes: This figure provides a stylized map of the Washington DC Metropolitan area. The circle represents the area in which the 39 stores in the scanner data sample are located. Montgomery & Prince George's Counties are in Maryland. Arlington County, Fairfax County, and the City of Alexandria are in Virginia.

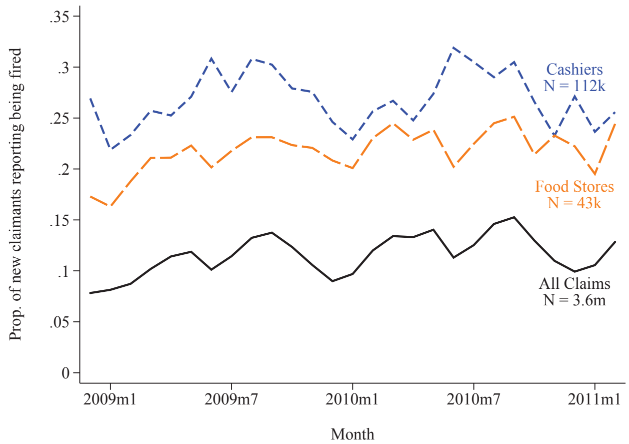
Figure 2.A2: Searches on Google via Google Trends - “Unemployment benefits” vs. other searches



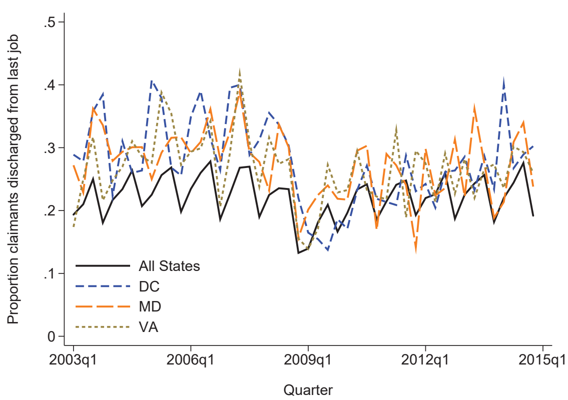
Notes: Google Trends data retrieved from Google Inc. Each panel reports a five-item query for “Unemployment benefits” (in black) vs. four other items for our time period (January 2008 - December 2009). For each term-week, Google Trends first calculates the ratio of the term’s search volume to the total number of searches (i.e. an absolute search measure for that term-week). Then, Google Trends proportionally scales all ratios across weeks and the five queried search terms to a [0,100] scale. So, within a given week, the ratio of two indices reveals the ratio of search frequency between two terms. For example, during the week of “Wall-E”’s peak search popularity, there were roughly five times the amount of searches for “Wall-E” than there were “Unemployment benefits.”

Figure 2.A3: Time series of UI claimants

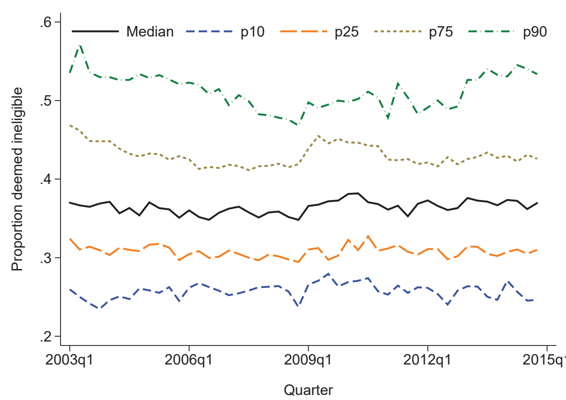
(a) Fired workers regularly file for UI (*Administrative UI claims data from California*)



(b) Over 20% of accepted UI claimants were fired (*Benefit Accuracy Measurement data*)

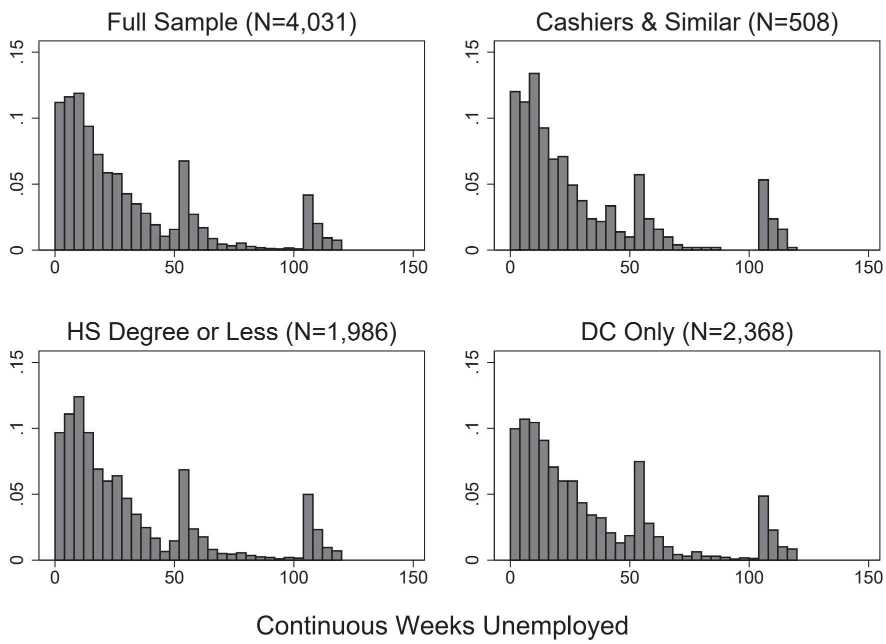


(c) Most claimants with misconduct determinations still receive UI (*US Department of Labor*)



Notes: Panel (a) utilizes administrative UI claims microdata acquired from the California Employment Development Department. The graph depicts the proportion of claims filed in CA during the time period of our scanner data sample, in which the claimant was fired from their last job. Panel (b) graphs the proportion of claimants who were fired from their last job in the Department of Labor's (DOL) Benefit Accuracy Measurement (BAM) data. The BAM program audits a randomly selected subsample of claimants receiving UI benefits in every state-week. Claims with missing separation reasons are excluded from panels (a) (39%) and (b) (0.4%). Panel (c) graphs variation across state-quarters in the proportion of misconduct determinations resulting in a denial of benefits from the DOL Employment and Training Adminstration's report 207. As described in section 2.3.3, all claims by discharged workers result in such a determination.

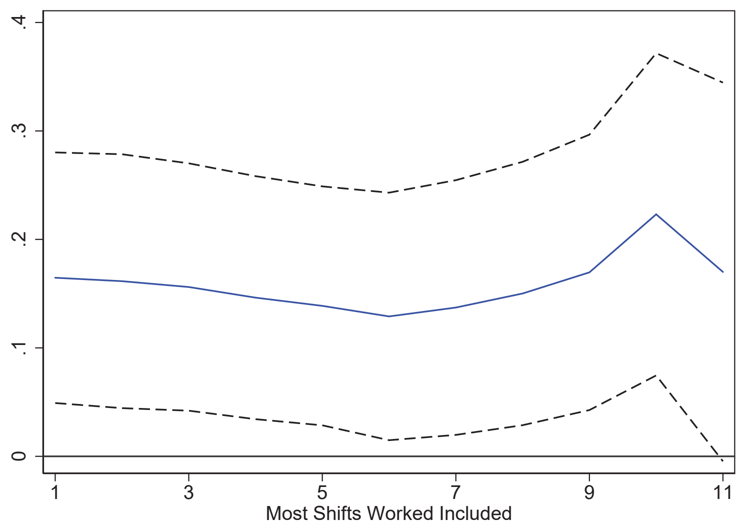
Figure 2.A4: Distribution of unemployment durations in CPS sample



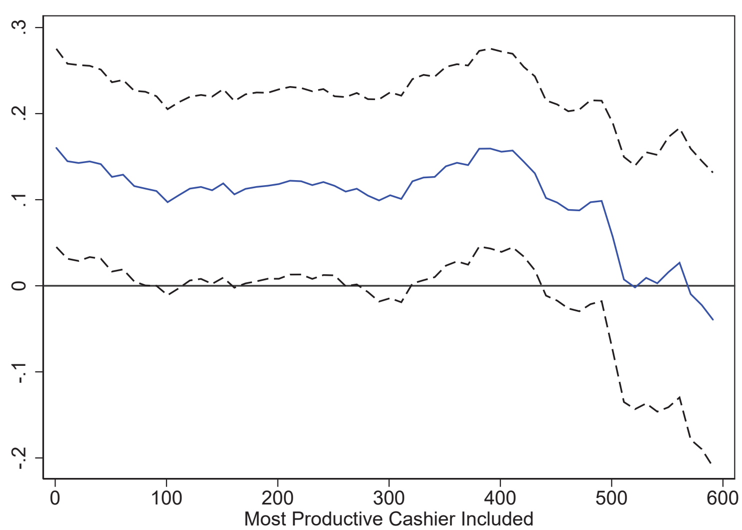
Notes: Each panel plots the distribution of unemployment spells for a cross-section of workers from the CPS monthly files for the months in our cashier sample (December 2008 to February 2011) who resided in the Washington D.C. metropolitan area. Jumps in distribution roughly correspond to (self-reported) unemployment spells of one year and two years.

Figure 2.A5: The effect of PBD on transaction duration by cashier subsamples

(a) Cashier experience - Number of shifts worked

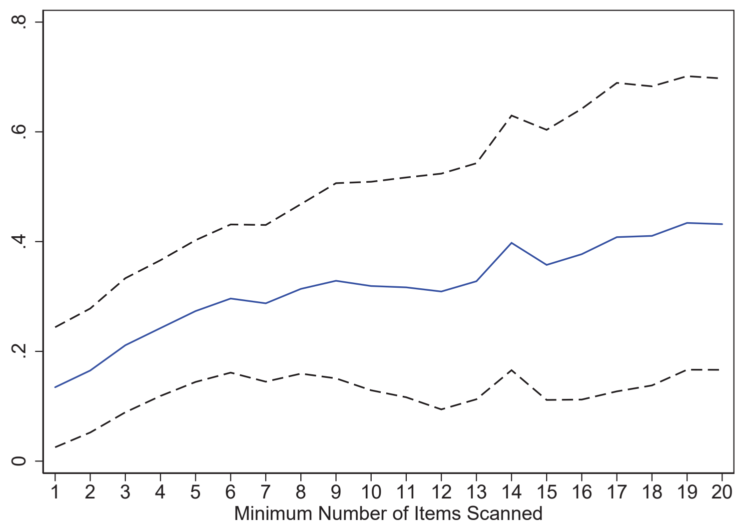


(b) Cashier productivity - Ranking of cashier pre-policy fixed effect



Notes: Point estimates (solid line) and 95% confidence intervals (dashed) for estimates of the effect of PBD on transaction duration from our fully specified model (cashier-register and day fixed effects, and controls) across numerous specifications. Each model is estimated in a different subgroup restricted to transactions completed by particular cashiers. In panel (a), starting with the full sample on the left (cashiers who worked at least 1 shift before the first policy change), estimates increase slightly as we focus on cashiers who worked, at a minimum, a higher number of shifts. In panel (b), starting with the full sample of cashiers on the left (where higher rankings correspond to higher productivity), estimates decrease as we focus on cashiers with higher rankings of pre-policy productivity.

Figure 2.A6: Does the effect of PBD on transaction duration vary with transaction size?



Notes: Point estimates (solid line) and 95% confidence intervals (dashed) for roughly 20 estimates of the effect of PBD on transaction duration from models with cashier-register fixed effects and controls. Each model is estimated in a different subgroup restricted to transactions that included more than a certain number of items.

Table 2.A1: Summary statistics from ATUS sample (N=30,094)

Worker-level variable	**Mean**	**Std. Dev.**
Age (in years)	40.352	12.395
Female	0.462	0.499
Race:		
- White	0.834	0.372
- Asian	0.029	0.169
- Black	0.115	0.319
Born in the US	0.918	0.275
Works in private sector	0.831	0.375
Occupation sector:		
- Management occupations	0.111	0.314
- Sales and related occupations	0.101	0.301
- Office and administrative support	0.151	0.358
Works part time	0.121	0.326
Usual number of weekly hours	41.732	9.173
Weekly earnings (in $)	900.487	1694.676
Paid hourly (not salary)	0.454	1.635
Number of minutes at the workplace:		
- Not working (shirking)	31.833	37.550
- Working	478.613	139.776

Notes: American Time Use Survey (ATUS) data initially collected at the respondent-activity level from the years 2003 to 2014, then collapsed to the respondent level. Observation weights provided by ATUS.

Table 2.A2: Estimated UI eligibility in Mas and Moretti (2009) sample

	mean	sd	min	max
Total hours to date	1347.01	903.61	0.0028	4414.14
Shift length (hours)	6.18	2.99	0.0008	16.08
Tenure to date (days)	384.43	189.91	0.0000	748.00
UI Eligibility Rate:				
- DC 2008	0.83	0.37	0.0000	1.00
- DC 2009	0.84	0.37	0.0000	1.00
- DC 2010	0.84	0.37	0.0000	1.00
- DC 2011	0.84	0.37	0.0000	1.00
- MD 2008	0.80	0.40	0.0000	1.00
- MD 2009	0.81	0.39	0.0000	1.00
- MD 2010	0.81	0.39	0.0000	1.00
- MD 2011	0.81	0.39	0.0000	1.00
- VA 2008	0.74	0.44	0.0000	1.00
- VA 2009	0.77	0.42	0.0000	1.00
- VA 2010	0.79	0.41	0.0000	1.00
- VA 2011	0.79	0.41	0.0000	1.00
N cashiers	412			
N cashier-shifts	55,205			

Notes: This information is based on a subset of the data used in Mas and Moretti (2009) which includes every transaction for 6 stores in the same metropolitan area of the Western Census region between (roughly) 2004 and 2006. After estimating cumulative hours worked at the cashier-shift level, we drop managers from the sample and estimate UI eligibility in our state-years for all cashier-shifts worked in a store that had been in the sample for 3 or more calendar quarters.

UI eligibility rules vary by state and are based on earnings histories in the location of employment, not residence. The UI eligibility rules in our sample are as follows (Source: Department of Labor, Online, *accessed 14 Sep. 2018*): In Maryland, \$900 in wages in the first four of the last five completed calendar quarters, with $\geq$\$576 in the highest earning of those quarters, and >\$0 in wages in two of those quarters; In Virginia, \$2,700 in wages in either the first four or the last four of the last five completed calendar quarters, with $\geq$\$2,700 in wages during the highest two earning of those quarters; In Washington D.C., \$1,950 in wages in either the first four or the last four of the last five completed calendar quarters, and either $\geq$\$1,300 in the highest earning of those quarters or $\geq$\$1,950 in the two highest earning of those quarters.

These estimates are likely to be conservative since a cashier's first day in the Mas and Moretti (2009) sample is likely not their first day at the retailer, a cashier may have relevant earnings from other employers, and a cashier may earn more than the minimum wage.

Table 2.A3: Results with data collapsed to cashier-register-day level

	Avg.(Transaction Length)		ln(Avg.(Transaction Length))	
	(1)	(2)	(3)	(4)
Potential benefit duration	0.187	0.184		
	(0.064)	(0.078)		
18-week PBD increase			0.019	0.017
			(0.009)	(0.011)
Total expenditures	0.082	0.060	0.000	-0.000
	(0.073)	(0.073)	(0.000)	(0.000)
Total items scanned	2.875	2.710	0.011	0.010
	(0.296)	(0.318)	(0.002)	(0.002)
Price discounts/coupons	0.559	0.679	0.005	0.005
	(0.135)	(0.146)	(0.001)	(0.001)
Local UE rate (prior month)	-0.699	-0.814	-0.010	-0.010
	(0.721)	(0.743)	(0.004)	(0.004)
State UE rate (prior month)	-1.403	-0.583	-0.002	0.002
	(2.081)	(2.236)	(0.013)	(0.013)
Observations	30,179	27,279	30,121	27,218
Controls	X	X	X	X
Date FE	X	X	X	X
Register X Store FE	X		X	
Cashier X Store FE	X		X	
Cashier X Register X Store FE		X		X

Notes: Potential benefit duration is measured in weeks. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, the total number of registers open during the transaction, the cashier's experience as measured by total number of career transactions completed, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, and the cashier's length of shift measured in both number of transactions and in minutes. "Date" refers to exact date (e.g., August 3, 2017). Standard errors, shown in parentheses, are clustered at state by date level.

Table 2.A4: Main results – Sensitivity to cashier controls and ignoring/dropping EUC=0 weeks

	Without Cashier Controls		Ignore EUC = 0		Drop EUC = 0	
	(1)	(2)	(3)	(4)	(5)	(6)
Potential benefit duration	0.174	0.175	0.133	0.130	0.133	0.146
	(0.058)	(0.068)	(0.047)	(0.056)	(0.048)	(0.058)
Total expenditures	0.133	0.134	0.134	0.134	0.135	0.134
	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)	(0.013)
Total items scanned	2.854	2.852	2.821	2.825	2.848	2.852
	(0.053)	(0.052)	(0.052)	(0.052)	(0.054)	(0.054)
Price discounts/coupons	0.730	0.735	0.720	0.726	0.717	0.723
	(0.024)	(0.024)	(0.024)	(0.024)	(0.025)	(0.025)
Local UE rate (prior month)	-0.687	-0.596	-0.678	-0.661	-0.516	-0.567
	(0.493)	(0.535)	(0.429)	(0.463)	(0.442)	(0.485)
State UE rate (prior month)	-2.542	-2.845	-1.537	-1.354	-1.568	-1.217
	(1.389)	(1.503)	(1.196)	(1.268)	(1.225)	(1.293)
Observations	515,618	515,433	515,618	515,433	471,826	471,647
Controls	X	X	X	X	X	X
Date FE	X	X	X	X	X	X
Register X Store FE	X		X		X	
Cashier X Store FE	X		X		X	
Cashier X Register X Store FE		X		X		X

Notes: Potential benefit duration is measured in weeks. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, the total number of registers open during the transaction, the cashier's experience as measured by total number of career transactions completed, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, and the cashier's length of shift measured in both number of transactions and in minutes. "Date" refers to exact date (e.g., August 3, 2017). Standard errors, shown in parentheses, are clustered at state by date level.

Table 2.A5: Main results by various standard error clusters

	State-Date		State-Month		Cashier		Cashier-Register		Store	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Potential benefit duration	0.133	0.130	0.133	0.130	0.133	0.130	0.133	0.130	0.133	0.130
	(0.047)	(0.056)	(0.043)	(0.046)	(0.055)	(0.059)	(0.048)	(0.053)	(0.071)	(0.076)
Total expenditures	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134	0.134
	(0.013)	(0.013)	(0.012)	(0.012)	(0.013)	(0.013)	(0.012)	(0.012)	(0.021)	(0.020)
Total items scanned	2.821	2.825	2.821	2.825	2.821	2.825	2.821	2.825	2.821	2.825
	(0.052)	(0.052)	(0.073)	(0.072)	(0.050)	(0.050)	(0.047)	(0.047)	(0.072)	(0.072)
Price discounts/coupons	0.720	0.726	0.720	0.726	0.720	0.726	0.720	0.726	0.720	0.726
	(0.024)	(0.024)	(0.035)	(0.034)	(0.021)	(0.021)	(0.020)	(0.020)	(0.033)	(0.034)
Local UE rate (prior month)	-0.678	-0.661	-0.678	-0.661	-0.678	-0.661	-0.678	-0.661	-0.678	-0.661
	(0.429)	(0.463)	(0.579)	(0.633)	(0.378)	(0.422)	(0.347)	(0.400)	(0.418)	(0.531)
State UE rate (prior month)	-1.537	-1.354	-1.537	-1.354	-1.537	-1.354	-1.537	-1.354	-1.537	-1.354
	(1.196)	(1.268)	(1.175)	(1.102)	(1.355)	(1.453)	(1.187)	(1.342)	(2.201)	(2.634)
Observations	515,618	515,433	515,618	515,433	515,618	515,433	515,618	515,433	515,618	515,433
Clusters	339	339	36	36	1,966	1,963	8,370	8,185	39	39
Controls	X	X	X	X	X	X	X	X	X	X
Date FE	X	X	X	X	X	X	X	X	X	X
Register X Store FE	X		X		X		X		X	
Cashier X Store FE	X		X		X		X		X	
Cashier X Register X Store FE		X		X		X		X		X

Notes: Potential benefit duration is measured in weeks. Transaction expenditure is measured in dollars. Controls for each regression include indicators for whether the transaction included items from particular departments (e.g., alcohol), an indicator for whether a plastic bag tax was in place at the store, number of households participating in SNAP per state-month, the total number of registers open during the transaction, the cashier's experience as measured by total number of career transactions completed, the cashier's "fatigue" as measured by the number of transactions the cashier had previously completed on that shift, and the cashier's length of shift measured in both number of transactions and in minutes. "Date" refers to exact date (e.g., August 3, 2017). Standard errors are shown in parentheses.

Table 2.A6: Results from American Time Use Survey (ATUS) - Minutes spent at workplace

	Minutes spent at workplace			
	(1)	(2)	(3)	(4)
18-week PBD increase	0.2209	1.0525	0.9007	1.1862
	(1.3582)	(2.9187)	(1.4834)	(2.5087)
State UE rate (prior month)			-0.6143	-0.6460
			(0.9657)	(0.9530)
Maximum WBA (100s)			-3.7401	-3.8411
			(2.9008)	(3.0414)
Observations	30,094	30,094	30,094	30,094
Mean of Y	510.4462	510.4462	510.4462	510.4462
State FE	X	X	X	X
Month FE	X		X	
Year FE	X		X	
Month-Year FE		X		X
Controls			X	X

Notes: Controls include state unemployment rate and maximum UI benefits (in dollars), the individual's age, "usual" amount of hours worked per week, weekly earnings, hourly wage, and dummies for family income, gender, race, US citizenship, whether the individual had multiple jobs, class of worker (e.g., federal government vs. state government vs. private for profit), and general occupational category (e.g., "sales and related occupations" vs. "healthcare support occupations"). Observations weighted according to ATUS probability weights. Standard errors, shown in parentheses, are clustered at state level.

Chapter 3

Am I My Brother's Barkeeper? Sibling Spillovers in Alcohol Consumption at the Minimum Legal Drinking Age (with Eunju Lee)

3.1 Introduction

Excessive alcohol consumption has been shown to harm young adults in a variety of dimensions including health, educational performance, and criminal activity (Carpenter and Dobkin, 2009; Carrell, Hoekstra and West, 2011; Lindo, Swensen and Waddell, 2013; Carpenter and Dobkin, 2015, 2017). Among the many potential determinants of adolescent alcohol use, peer effects have received substantial attention from academics and policymakers.[1]

Leading empirical work provides strong evidence of positive peer effects in alcohol consumption via random assignment to college roommates (Duncan et al., 2005; Eisenberg, Golberstein and Whitlock, 2014; Guo et al., 2015). However, this design exploits variation in exposure to *a peer*. These results do not tell us what peer behavior or characteristic explains the spillovers—they are

[1] In its Underage Drinking Fact Sheet, the National Institute of Alcohol Abuse and Alcoholism lists peer pressure as one of three key causes of alcohol consumption among adolescents (https://pubs.niaaa.nih.gov/publications/UnderageDrinking/UnderageFact.htm).

“contextual” as opposed to “endogenous” peer effects, where the latter is an effect driven specifically by peer alcohol consumption (Manski, 2000). Distinguishing between these two types of peer effects has key implications for policy. With endogenous peer effects, an intervention that influences the alcohol consumption of recipients also has an effect on the consumption of their untreated peers. This is not necessarily the case under contextual effects. Further, this design cannot be replicated in many other peer groups of interest, notably siblings—a peer group that has been the focus of a substantial and rapidly growing literature on a wide range of outcomes.[2]

This paper provides causally interpretable estimates of peer effects in alcohol consumption between siblings. We focus on siblings residing in the same household and exploit a discontinuous increase in older sibling alcohol consumption at the minimum legal drinking age (MLDA) using a regression discontinuity design (RDD).[3] Our results are, to our knowledge, the first quasi-experimental estimates of peer effects in alcohol consumption between siblings and our estimates are interpretable as the causal effect of sibling alcohol consumption. This extends the prior literature, which primarily estimates the causal effect of exposure to heavier drinking roommates (who may have other unobserved characteristics that influence peer alcohol consumption).

Attempts to estimate causal peer effects face three main difficulties. First, the “reflection” problem implies that a simple regression of peer A’s behavior on peer B’s cannot determine the direction of any observed effect Manski (1993). Second, peer groups are (typically) endogenous, which is a concern if individuals choose peers who have similar preferences related to the behavior of interest. Finally, individuals who share a peer group are likely to experience unobservable common shocks which are correlated with their outcomes.

[2] Including, for example, risky behaviors (Duncan, Boisjoly and Harris, 2001; Fagan and Najman, 2005; Harris and López-Valcárcel, 2008; Altonji, Cattan and Ware, 2016), health (Cawley et al., 2019; Daysal et al., 2019; Ho, 2017; Breining, 2014), fertility (Heissel, 2021), education (Goodman et al., 2015; Joensen and Nielsen, 2018; Karbownik and Özek, 2019; Altmejd et al., 2021), and military service (Bingley, Lundborg and Lyk-Jensen, 2021).

[3] An identification strategy first used by Carpenter and Dobkin (2009) to study the effect of alcohol consumption on mortality.

Our setting and identification strategy address each of these difficulties. By focusing on siblings, a peer group which is naturally occurring, we avoid the potential for selection into the peer group. By restricting the sample to siblings with different ages, we ensure that the variation in alcohol consumption that we utilize is both exogenous (avoiding the problem of common shocks) and specific to one of the two siblings (avoiding the reflection problem).

To estimate spillover effects in alcohol consumption we utilize data from the 1997 National Longitudinal Survey of Youth (NLSY97), which has three key features for our analyses. First, the data contains unique "household roster" information on all individuals living with the NLSY97 respondent. These rosters include the birth months of each household member and their relationship to the respondent, which we use to construct our running variable. Second, it contains several dimensions of alcohol consumption for survey respondents that are measured at relatively high frequency (consumption in the past month). Third, all NLSY97 respondents were between the ages of 12 and 17 during the first wave of the survey, which ensures that many survey respondents will have older siblings near the age 21 cutoff.

Our main results estimate the effect of the oldest sibling's legal access to alcohol on the younger (second oldest) sibling's alcohol consumption.[4] In the full sample of 4,278 sibling-pair-years, our estimates imply a *negative* relationship between the older sibling's legal access to alcohol and the younger sibling's alcohol consumption but are typically somewhat imprecise. In our preferred specification we estimate that the number of binge drinking days (5 or more drinks) in the past 30 days reported by younger siblings decreases by 0.34 days at the cutoff. This is a substantial effect given that the average younger sibling in the sample reports 1.14 binge drinking days in the past month. While this estimate is statistically significant (95% CI of [-0.650, -0.025]), similar

[4]In most cases we only observe alcohol consumption for one sibling in a household (the sibling that is a NLSY97 respondent). However, we do estimate the discontinuity in older sibling alcohol consumption in a separate sample of NLSY97 respondents who have younger (usually not NLSY-respondent) siblings. Results are similar to the existing literature utilizing this research design, typically implying increases in consumption of 20-30%.

specifications which differ in terms of outcomes or control variables are often less precise. However, we do consistently find a moderate negative effect across a wide range of specifications. Further, our preferred estimate is well into the left tail of a distribution of placebo discontinuities estimated at older sibling ages that are close, but not equal to, 21 years.

These estimates are almost always precise enough to rule out the relatively large positive effects reported in prior work. Notably, we compare our estimates to those from Eisenberg, Golberstein and Whitlock (2014) (henceforth EGW) since they utilize what is arguably the gold-standard research design (randomized college roommates) on this question, and their results are smaller than nearly all other estimates in the related literature.[5] EGW's key comparable result is that being assigned a roommate who binge drank in 30 days before move-in leads to a 19% increase in the probability of any binge drinking by the respondent in a second 30-day period roughly 8 months later. After scaling our preferred estimate for the same outcome by the control mean, we obtain a 95% CI of [-27.6%,+9.2%]. Further, among 64 total specifications estimated for this outcome, 59 confidence intervals (95%) exclude an effect of +19%. It is worth emphasizing again that we *do not* view these results as contradictory since EGW estimate contextual and not endogenous peer effects.

As a whole, we view our results as suggestive evidence that the causal relationship in alcohol consumption between siblings in this age range is negative. This is admittedly counterintuitive. However, a large literature has established that a series of negative outcomes spike at the MLDA (e.g., Carpenter and Dobkin, 2017, 2015; Lindo, Swensen and Waddell, 2013; Carrell, Hoekstra and West, 2011; Carpenter and Dobkin, 2009) and it is plausible that younger siblings update their beliefs about the costs of alcohol consumption after observing their older siblings experience these negative consequences.[6] Further, the wider literature on peer effects has often considered the poten-

[5]Notably Gaviria and Raphael (2001), Duncan et al. (2005), Lundborg (2006), Fletcher (2012), Guo et al. (2015), and Altonji, Cattan and Ware (2016) all report similar or larger peer effects of alcohol consumption in their preferred specifications.

[6]Some specific negative outcomes that have been studied in the literature are likely too rare to explain our spillover effects. However, despite the lack of quantitative evidence it is likely that other less-severe and more common negative

tial for spillovers to operate through similar pathways related to "expectations" about a behavior.[7] Unfortunately, our data does not allow us to directly test for evidence of this mechanism,[8] but we do find that our effects vary in several key dimensions in ways that are consistent with a non-zero effect.

First, our estimates are especially negative for measures of excessive alcohol consumption (binge drinking) on the intensive margin. Prior work suggests that negative effects of alcohol consumption experienced at the MLDA are driven by these types of consumption (Carpenter, Dobkin and Warman, 2016), and it is reasonable to expect that any spillover effects would be most apparent on this margin as well. Second, our estimates are more negative (and statistically significant) among siblings who are the same gender. It is reasonable to assume that these sibling pairs would be more exposed to one another, and that any true spillover effect would be stronger in this group. Third, our estimates are more negative when the older sibling is a male—this fits with prior literature which finds that males report larger increases in alcohol consumption, and especially excessive alcohol consumption, at the MLDA (Carpenter, Dobkin and Warman, 2016).

Three specific concerns which may threaten the interpretation of our results are worth highlighting. First, an older sibling's legal access may directly affect the access of the younger sibling (e.g., if the older sibling purchases alcohol for the younger). However, this effect would bias our negative estimates toward zero.[9] Second, parents may compensate for an expected spillover effect with a change in parenting style (e.g., by adopting a stricter parenting style) when the older sibling turns 21, which would likely negatively bias our results. The NLSY97 includes information

outcomes also spike at age 21.

[7]See, e.g., the categorization of peer effect mechanisms in Manski (2000).

[8]Both the NLSY97 data and several other related datasets, notably the National Health Interview Survey (NHIS) and the National Longitudinal Study of Adolescent Health (Add Health) do contain related info. However, the NLSY97 only asks these questions in a small number of survey waves (restricting our sample size), the NHIS has few respondents under age 21, and Add Health lacks the detailed household information necessary to construct our running variable.

[9]Further, if an access effect exists it would likely be driven by an increase in older sibling consumption. In this scenario the access effect would be part of our parameter of interest (the endogenous peer effect) and would not bias our results away from the true value of that parameter.

parenting style which allows us to provide some suggestive evidence against this concern.[10] Third, it is possible that younger siblings are not meaningfully exposed to alcohol consumption of older siblings who are near the MLDA. To further support our focus on this peer group we use the American Time Use Survey (ATUS)—which includes a minute-level breakdown of what activities were performed during a single day, where, and with whom—to demonstrate that siblings in the relevant age range spend substantial amounts of time together.

We contribute to three separate literatures. First, we add to a limited number of studies which provide quasi-experimental estimates of endogenous peer effects in alcohol consumption. The most similar work is Fletcher and Marksteiner (2017), who estimate endogenous peer effects in alcohol consumption with a different identification strategy (randomized assignment to an alcohol cessation program) in a very different population (adult spouses in which one spouse is a heavy drinker). Second, we provide what we believe to be the first quasi-experimental estimates of spillovers in alcohol consumption between siblings. The existing economics literature on this topic has focused on other outcomes (e.g., smoking as in Harris and López-Valcárcel (2008)) or has relied on structural assumptions (Altonji, Cattan and Ware, 2016), while a substantial literature outside of economics has documented the strong correlation in consumption between siblings (e.g., Duncan, Boisjoly and Harris, 2001; Fagan and Najman, 2005; Trim, Leuthe and Chassin, 2006; Van Der Vorst et al., 2007; Whiteman, Jensen and Maggs, 2013). Third, we contribute to a growing literature on sibling spillovers more generally, where a growing body of evidence has demonstrated the importance of sibling influences in a variety of important domains including health outcomes (Cawley et al., 2019; Daysal et al., 2019; Ho, 2017; Breining, 2014), fertility (Heissel, 2021), education (Goodman et al., 2015; Joensen and Nielsen, 2018; Karbownik and Özek, 2019; Altmejd et al., 2021), and military

[10] Even in the presence of these confounders our results are interpretable as the spillover effects of the older sibling's legal access to alcohol which is itself policy relevant given contemporary debates on lowering the MLDA (Carpenter and Dobkin, 2011; DeJong and Blanchette, 2014). See, e.g. https://www.usatoday.com/story/news/nation-now/2017/11/09/drinking-age-19-these-wisconsin-lawmakers-aim-change-law/849365001/ for details on a MLDA change recently under debate in the Wisconsin state legislature.

service (Bingley, Lundborg and Lyk-Jensen, 2021).

The rest of this article is structured as follows. Section 3.2 describes the data. Section 3.3 details the methods used in the analysis. Section 3.4 describes the results, and section 3.5 concludes.

3.2 Data

3.2.1 NLSY97

The NLSY97 is a longitudinal survey of 8,984 American youths who were between the ages of 12 and 17 at the time of the first wave of the survey in 1997.[11] Designed and administered by the Bureau of Labor Statistics the NLSY97 includes a wide variety of questions related to family processes, education, employment, health, and family formation among other topics. The NLSY97 has three key features which make it uniquely suitable for our analysis.

First, information on the alcohol consumption of respondents is obtained in all waves. Specifically, we use the following outcomes, all reported for the past 30 days prior to the survey: the number of days on which any drinks were consumed (drinking days), the number of days on which 5+ drinks were consumed (binge days), and the binary versions (any/none) of drinking days and binge days.[12] Second, the survey includes detailed information on the composition of each respondent's household, including the relationship of each household member to the respondent and the age of each household member (including the respondent) in months.[13] Third, each wave includes a detailed set of covariates at the respondent and household levels which are also potentially re-

[11] Waves are annual from 1997-2011, and biennial beginning in 2013.

[12] Additional measures of alcohol consumption used to define subgroups, but not used as outcomes include whether or not the respondent had ever consumed alcohol as of the previous year and whether or not the respondent had ever binge consumed alcohol as of the previous year.

[13] Specifically, the public use version of the NLSY97 includes the exact date of the interview along with the month of birth for the respondent and all members of the respondent's household. This allows us to estimate the age in months of the respondent and all siblings who ever co-resided with the respondent during the survey, regardless of whether those older siblings are themselves NLSY97 respondents. The NLSY97 also collects a very limited set of information on close relatives such as siblings who live outside of the respondent's household. This information is unfortunately too limited to make use of these siblings in our analyses.

lated to alcohol consumption and therefore will serve as control variables in our analysis. These include the race and gender of the respondent and siblings, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year.[14]

The NLSY97 sampling process first selected a large number of households and then identified all individuals in each household who were age-eligible for the survey. In this way the NLSY97 attempted to survey all age-eligible siblings in a household. For reasons that we will describe in the next section, we will primarily focus on 2,614 NLSY respondents who have only one older sibling in their household, and where that older sibling is between the ages of 19 and 23. Key subgroups of interest are defined based on parental educational attainment, whether the siblings report the same gender, the gender of the older sibling, the age difference within the sibling pair, and the younger siblings previous (as of prior survey) drinking experience. Summary statistics are shown in Table 3.1.

An alternative sample of siblings who are all NSLY97 respondents is available but will not be used in our main analysis. However, their reported alcohol consumption can be used to provide descriptive evidence of between sibling correlation in alcohol consumption in the NLSY97. Table 3.2 shows results from a regression of the younger sibling's past month alcohol consumption on the older sibling's past month alcohol consumption for multiple measures of consumption in a sample of sibling pairs that are both NLSY97 respondents.[15] These results are not causally interpretable and are not directly comparable to our sample of interest. However, they show that in this sample, much like in prior literature described previously, there is a very strong correlation in alcohol

[14]This is not an exhaustive list of potentially relevant control variables included in the NLSY97. A large number of relevant questions in the NLSY97 are either asked too intermittently or have too many missing values to be useful in our analysis.

[15]Households with 3 or more sibling NLSY97 respondents are excluded for simplicity and the sample is limited to individuals under the age of 23.

consumption between siblings.

Among a set of measures of past month alcohol consumption in the NLSY97, we focus on measures of the excessive alcohol consumption in our main analyses: the number of binge drinking days (5+ drinks) in the past month, and an indicator for any binge drinking days in the past month. This helps make our results more comparable to the existing literature, which typically focuses on measures of binge drinking, and more policy relevant, since excessive drinking is more likely to result in the various negative outcomes that are associated with alcohol consumption. The focus on a small subset of the available outcomes also helps to reduce problems of multiple testing. However, results for all outcomes are presented for certain analyses in order to demonstrate robustness.

3.2.2 ATUS

In a supplementary analysis, we use the ATUS to provide descriptive evidence that siblings in the relevant age range constitute an important peer group (Hofferth et al., 2020). The ATUS is a repeated cross-sectional survey, which regularly samples a subset of Current Population Survey (CPS) respondents and solicits a detailed time-diary for one day's worth of activities. These time diaries detail the minute-level activities of respondents, including what activity was being performed, where, and with whom. Since the ATUS sample is drawn from the CPS, the two surveys can be linked so that the household structure of each ATUS respondent is observable. We use this information to select all respondents from the 2003-2019 waves of the ATUS who lived with one older sibling 24 years old or younger. We then use the information on who activities were performed with, to characterize the strength of these sibling relationships and provide suggestive evidence regarding the potential for younger siblings to be exposed to changes in older sibling alcohol consumption. Summary statistics for our analysis sample of 3,336 ATUS respondents are shown in Table 3.3.

3.3 Methods

We implement a reduced form regression discontinuity design in which the the running variable is the age of the older sibling in months and the cutoff is at 21 years (252 months). The MLDA was first used as an exogenous source of variation in alcohol consumption in a RDD by Carpenter and Dobkin (2009) to study the effect of alcohol consumption on mortality and has since been used to study a wide range of other outcomes.[16] So long as no other factors related to the outcome are changing discontinuously at the cutoff, this approach will provide causally interpretable results. Many of the aforementioned studies have provided convincing evidence of both a strong first stage (large, discontinuous increases in consumption at age 21) and of the credibility of this research design (e.g., establishing that observable covariates do not change discontinuously at the cutoff).

One person's alcohol consumption may be influenced by the alcohol consumption of all their other siblings. As pointed out by Dahl, Løken and Mogstad (2014) this presents a complication when using a RDD to estimate peer effects within a naturally occurring peer group since it is unclear how to define the running and treatment variables. To avoid these complications, we define the peer group as siblings who currently reside in the same household and consider the effect of the oldest sibling in the peer group on the second oldest sibling. This ensures that we have a large sample of sibling pairs in which we observe the information necessary to implement the RDD (older sibling age in months and younger sibling alcohol consumption) and allows us to avoid the complications involved with defining the running variable in peer groups with more than two members.

[16] Including crime (Carpenter and Dobkin, 2015; Hansen and Waddell, 2018), morbidity (Carpenter and Dobkin, 2017), marijuana consumption (Yörük and Yörük, 2011; Crost and Guerrero, 2012; Crost and Rees, 2013; Yörük and Yörük, 2013), the consumption of other illegal drugs (Deza, 2015), risky sexual behavior Yörük and Yörük (2015), and mental health (Yörük and Yörük, 2012).

The reduced form RD is then implemented by estimating the following equation via OLS:

$$alc_{2,h,t} = \gamma_1 1\{age_{1,h,t} \geq 21\} + f(age_{1,h,t}) + X_{2,h,t}\gamma_2 + X_{1,h,t}\gamma_3 + W_{h,t}\gamma_4 + \theta_{2,h} + \mu_{2,h,t} \quad (3.1)$$

where the outcome is some measure of past month alcohol consumption for the younger sibling, the reduced form effect of older sibling's legal access to alcohol on the younger sibling's consumption is estimated by γ_1, $f(age_{1,h,t})$ is a flexible polynomial in the running variable (e.g., older sibling age fully interacted with the cutoff dummy $1\{age_{1,h,t} \geq 21\}$), $X_{2,h,t}$ is a vector of younger sibling covariates, $X_{1,g,t}$ is a vector of older sibling covariates, $W_{h,t}$ is a vector of household h level covariates, and $\theta_{2,g}$ is a younger sibling fixed effect.

We separately estimate the increase in older sibling drinking at age 21 in a similar equation (where the outcome is the alcohol consumption of the older sibling, $alc_{1,h,t}$, and younger sibling characteristics are removed the regression). The sibling pairs used in that analysis are not the same as those used to estimate the reduced form effect.[17]

We do not directly observe older sibling age in months. Instead, the running variable is estimated from the birth month of the older sibling and the month of the relevant interview (younger sibling's interview in the reduced form, older in the first stage). This rounding implies that the cutoff indicator will be mismeasured for some sibling pairs in which the older sibling is exactly 21 years (252 months) old. Following the recommendations in Dong (2015), we address this misclassification bias with a "donut RD" specification that excludes sibling pairs in which the older sibling is exactly 21 years (252 months) old. Following the existing literature on MLDA-based RDDs, the sample is limited to sibling pairs in which the older sibling is between the ages of 19 and 23—i.e., the

[17] Sibling pairs in the reduced form sample are those in which the second oldest sibling the household is a NLSY97 respondent. Sibling pairs in the first stage are those in which the oldest sibling is a NLSY97 respondent. Although a two-sample two-stage least squares is theoretically possible, sample sizes are likely to be too small for it to be informative.

youngest older sibling in our sample is 19 years and 0 months old (228 months of age) and the oldest is 23 years and 0 months old (276 months of age). Standard errors are cluster robust at the younger sibling level.

Additional models test the sensitivity of our results in several of the dimensions mentioned above (bandwidth, correction for rounding-induced bias, inclusion of covariates, inclusion of fixed effects, and order of the running variable polynomial). Notably, we implement the continuity-based RD framework developed by Cattaneo, Idrobo and Titiunik (2019), including mean-squared error optimal bandwidth selection and bias-corrected standard errors.

The main assumption required for a causal interpretation of γ_1 is that no unobserved confounding factors change discontinuously at the cutoff. We "test" this assumption to the extent possible by estimating models similar to equation 3.1, where outcomes are predicted values from separate regressions which predict the alcohol consumption measures using a range of covariates. A second required assumption in RDDs more generally is that the running variable is not manipulated. Although this is not technically a concern in our setting (since age is not manipulable), a related problem can arise if rates of nonresponse to alcohol consumption questions change discontinuously at the cutoff. This is primarily a concern in the first stage where an older sibling may become more willing to report any alcohol consumption once they reach age 21. We test this assumption by demonstrating visually that the density of older sibling age is smooth through the cutoff and by testing formally for a discontinuity in the density at the cutoff.

3.4 Results

3.4.1 Design

We begin by demonstrating that older sibling drinking changes discontinuously at age 21. Results using count and binary versions of drinking and binge days in the past month are shown in Table 3.4. Similar to results from previous work using the same identification strategy in different samples,[18] there is a large discontinuous increase in alcohol consumption at age 21. This effect is apparent in all models shown in Table 3.4. In preferred models (donut specification with a linear function of the running variable, controls,[19] and individual level fixed effects), past month binge drinking and drinking days increase by 0.45 days and 1.45 days at the cutoff. Both increases are statistically significant at the 5% level. Increases on the extensive margin of consumption for binge drinking and drinking (from the same models) are 5.7pp and at 8.3pp respectively, which are both statistically significant at the 5% level. Figure 3.1 plots mean values of our four outcomes among older siblings in each month-of-age bin along with fitted lines and regression coefficients from the corresponding first stage regressions. These graphs reinforce the results summarized from Table 3.4.

Figure 3.2 graphs the distribution of the older sibling's age in months, for the first stage sample. While age is not manipulable, it is possible for nonresponse to change discontinuously at the cutoff and such a result could potentially affect the interpretation of the first stage results presented in Table 3.4. Figure 3.2 suggests that this is not occurring, given that the distribution is relatively smooth through the cutoff at 21 years of age.

In Table 3.5 and Figure 3.3 we present results which test for discontinuities in sibling pair characteristics at the cutoff that could potentially confound the causal effect of interest. We use

[18] See the previous section for citations.

[19] Month and year of the survey, the race and gender of the respondent and siblings, the age of the respondent, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year.

a set of covariates[20] to predict each of our four alcohol consumption outcomes and then use a regression analogous to our reduced form specification to test for a discontinuity in each of those predictions at the cutoff (standard errors are calculated via bootstrap). All discontinuities are statistically indistinguishable from zero.

3.4.2 Reduced form

Table 3.6 shows results from equation 3.1, estimated in the full sample of sibling pairs who reside in the same household and are the oldest siblings in the household, for all four measures of alcohol consumption. Focusing on the preferred model (the first column) which includes the previously defined vector of controls and individual level fixed effects, all point estimates are negative and large (e.g., younger sibling past month binge days decrease by 0.337 or 26.9% of the mean just before the cutoff). Assuming that no unobserved confounder changes discontinuously when the older sibling turns 21, this implies that the older sibling's increase in alcohol consumption at the cutoff has a causal negative effect on the alcohol consumption of younger siblings on this margin.

Point estimates are statistically significant for only one of the four considered outcomes in our preferred specification, suggesting that these estimates should be interpreted with caution. However, the outcome for which our results are strongest (count of binge drinking) is also arguably the most important—as it represents a particularly risky and problematic type of consumption. Further, point estimates for the preferred models are negative and at least modestly sized for nearly all outcomes in the preferred specification, and nearly all estimates in Table 3.6. As a percentage of the mean of the outcome just before the cutoff, coefficients in the preferred specification imply that binge drinking decreases by 9.2% on the extensive margin, while drinking (any amount) decreases by 5.9% on the intensive margin and 6.7% on the extensive margin.

[20]Specifically: age, gender of both siblings, educational enrollment, highest completed education, work status, indicators for whether the household lives in an urban area, census region dummies, AFQT score, household size, and interview month/year. (All variables refer to the younger sibling's info unless otherwise noted.)

Corresponding binned scatter plots for all four outcomes are shown in Figure 3.4. These figures include coefficients from a regression of each outcome variable on the age of the older sibling in months, whether the older sibling is over 21, and their interaction. The figures show that there is no significant increase in the alcohol consumption of younger sibling when their older sibling turns to 21, suggesting (as also shown in Table 3.6) that the size and precision of our negative estimates are somewhat sensitive to the inclusion of additional controls.

Why might an increase in older sibling alcohol consumption at the MLDA have negative effects on younger sibling consumption? One natural interpretation is that younger siblings learn about the costs of alcohol consumption by observing the experiences of their older siblings. While the NLSY97 data does not allow us to test for this mechanism directly (e.g., by using some measure of knowledge or beliefs about the effects of alcohol consumption as an outcome in our RDD), we argue that a true non-zero effect is likely to be stronger in certain subgroups than others, discussing this point and relevant subgroup estimates in section 3.4.3. Further, heterogeneity across different outcomes (described above) is also consistent with this proposed mechanism. Our results imply that younger siblings reduce consumption on all margins that we test. However, these responses are strongest for the margins (binge drinking, especially on the intensive margin) that are most risky and that have been found in the prior literature to drive negative outcomes at the MLDA (Carpenter, Dobkin and Warman, 2016).

To further strengthen our results, we perform a series of placebo tests estimating discontinuities at older sibling age values near but not equal to 21. Specifically, for each monthly age bin up to 12 months on either side of the cutoff, we re-estimate equation 3.1. Results are plotted in Figure 3.7, with the estimate at the actual cutoff value represented by a red vertical line. While this test cannot be used to conduct formal inference, the results are reassuring. The true estimate is in the far left tail of both placebo distributions.

3.4.3 Heterogeneity across subgroups

To determine which subgroups may be driving the relationships seen in the full sample, we re-estimate our reduced form specification in several subgroups of interest. Specifically: parental education (all parents completed high school vs. one or more did not complete high school), whether the two siblings report the same or different genders, the gender of the older sibling, the age difference between the siblings ($\geq$ or $<$ the median of 30 months), and the younger sibling's experience of drinking or binge drinking as of the prior year's survey (any or none).

We show our main results from this analysis in Table 3.7. Our specifications align with the preferred model from Table 3.6 (a linear donut RD specification with controls and FEs included) and the outcome is the number of binge drinking days in the past month. In Table 3.8 we report the same results from similar regressions with an indicator for any binge drinking as the outcome. Several of these subgroups are defined based on covariates which could reasonably be expected to mediate the sibling spillover effect of interest—either because the groups identify siblings who are likely more exposed to each other (gender match, age), proxy for heterogeneity in the first stage (older sibling gender (Carpenter, Dobkin and Warman, 2016)), or identify younger siblings with different levels of prior (pre-cutoff) knowledge about alcohol.

Point estimates are typically negative and large but not statistically significantly different from zero. One key exception is the subgroup where both siblings report the same gender. In that subgroup, younger sibling consumption decreases by 0.616 binge drinking days when the older sibling turns 21. Intuitively, a more pronounced effect in this subgroup makes sense if we expect that siblings of the same gender are likely to spend more time together. We also find that effects are strong when the older sibling is a male. In this subgroup, younger siblings report 0.662 less binge drinking days in the past month at the cutoff. This result is reassuring given estimates in the prior literature which suggest that the MLDA has a larger effect on the drinking behavior of males. We estimate similarly heterogeneous effects in these subgroups with the binary outcome in Table 3.8, although all estimates there are less precise.

Patterns for subgroups defined by prior drinking experience, and age difference, are less clear. One possible explanation for this is that these characteristics may be strongly correlated with each other, or with other key factors which predict alcohol consumption. If the age difference between siblings is large, then the younger sibling is likely to have less prior alcohol consumption simply because they are younger. This complicates the interpretation of these subgroup estimates. In Table 3.7, point estimates are noticeably more negative among siblings with larger age differences. Does this reflect the age difference? Or the fact that younger siblings in such pairs have less drinking experience? Further, less prior drinking experience for one sibling likely implies the same for the other (due to the strong positive correlation in consumption between siblings), creating a similar complication. If we observe smaller discontinuities among younger siblings with less prior drinking experience, is that evidence against an expectations pathway (since such younger siblings should have "more to learn")? Or does that estimate simply reflect a smaller discontinuity in older sibling alcohol consumption?

3.4.4 Comparison to prior literature

While we believe that the negative effects we estimate are plausible, our sample sizes are relatively small and the estimates are not always statistically significantly different from zero. However, given the uniformly positive effects estimated in the related literature (e.g., Gaviria and Raphael, 2001; Duncan et al., 2005; Clark and Lohéac, 2007; Lundborg, 2006; Fletcher, 2012; Eisenberg, Golberstein and Whitlock, 2014; Guo et al., 2015; Altonji, Cattan and Ware, 2016; Fletcher and Marksteiner, 2017) a null effect is interesting in its own right.

We simplify our comparison to the prior literature by focusing on Eisenberg, Golberstein and

Whitlock (2014) (EGW). We choose this study for two reasons. First, their estimates are smaller than nearly all other estimates in the related literature. Therefore, if our null results are sufficiently precise to rule out the EGW point estimate, the same is true for a series of other leading estimates in the literature.[21] Second, EGW utilize what we see as the best existing identification strategy on this question: randomized assignment to college roommates.

EGW's key comparable result is that being assigned a roommate who binge drank in 30 days prior to a baseline survey fielded in August of 2009 (just before the students moved in to their shared rooms) leads to a 19% increase in the probability of any binge drinking by the respondent in a second 30-day period roughly 8 months later. In Figure 3.5 we scale our preferred estimate for the same outcome by the control (immediately pre-cutoff) mean, obtaining a 95% CI of [-27.6%,9.2%] in our preferred specification and ruling out the EGW point estimate (red line) in all four specifications. In the right-hand panel, we repeat this comparison for a series of our subgroup estimates in our preferred specification. Here 95% CIs exclude the EGW point estimate in most subgroups.

Since we estimate the direct effect of the peer's alcohol consumption, while EGW estimate the effect of exposure to a peer who binge drank in the past, we do not view these results as contradictory. Instead, if college roommates and co-resident older siblings at similar ages are similar peer influences, these estimates suggest that the effects estimated by EGW result at least in part from roommate characteristics other than drinking behavior. As mentioned previously, this is critically important for policy. Since the presence of endogenous peer effects that result directly from the peer's alcohol consumption indicate that the costs of alcohol consumption, and the benefits of consumption-reducing interventions, are multiplicative.

Although the EGW treatment (a roommate who did not binge drank vs. one who did) may

[21] Notably Gaviria and Raphael (2001), Duncan et al. (2005), Lundborg (2006), Fletcher (2012), Guo et al. (2015), and Altonji, Cattan and Ware (2016) all report similar or larger peer effects of alcohol consumption in their preferred specifications.

seem larger than the effects of the MLDA on older-sibling alcohol consumption, timing implies that the difference is not as stark as it appears. The EGW treatment is a 100% increase in exposure peer binge drinking measured roughly 8 months in the past, while we measure a contemporaneous spillover effect. Especially given the trajectory of alcohol consumption in this age range (as demonstrated by various figures in this paper), it is likely that many of the "control" peers in EGW (no prior binge drinking) would have binge drank by the time the outcomes were measured 8 months later. Therefore, the contemporaneous treatment vs. control difference in peer binge drinking in the EGW sample is likely much smaller than 100%.

3.4.5 Evidence against offsetting parental responses

Parental responses are one key pathway through which the older sibling's legal access to alcohol could directly (i.e., not via an endogenous peer effect) affect the younger sibling's consumption of alcohol. The NLSY97 includes two separate questions which can provide some suggestive evidence on the role that parental responses may or may not play in explaining the results described above. First all respondents are asked to score the degree to which they are monitored by their parents (including, if applicable, parents they live with and parents they do not live with) on a scale from 0-16, with higher scores indicating closer monitoring. Second, respondents are asked to classify their parents' parenting styles (again, including if applicable parents who do and do not live with the respondent) as either uninvolved, permissive, authoritarian, or authoritative.

In the first panel of Table 3.11 we present results for models similar to equation 3.1 with the parental monitoring score for the parents that the respondent lives with. If the respondent lives with two parents, the outcome is their average score. In the second panel of Table 3.11, similar results are presented with an indicator for whether or not at least one parent was reported to be either authoritarian or authoritative.

The survey questions underlying these outcomes are asked only in survey rounds between 1997 and 2000. Luckily, a substantial portion of the sibling pair years that meet the sample restriction requirements are in this time period. However, some are not, and sample sizes are therefore smaller than the previously described full sample results in Table 3.6. Models in Table 3.11 are the same as those in Table 3.6 in order to demonstrate the robustness of the results.

In each of the 5 models for each of the three outcomes, point estimates for a discontinuity in parental behavior at the cutoff are either negative (suggesting parents become less strict or engage in less monitoring), not statistically significant, or both. In the preferred model (1) neither point estimate is economically significant. We interpret this as suggestive evidence that parental responses are not driving the observed decrease in younger sibling alcohol consumption at the cutoff.

3.4.6 Time use of siblings near the MLDA

A concern in our setting is that the peer group we focus on may not be relevant for the behavior we are interested in. Our emphasis on "null" results magnifies the importance of this concern for two reasons. First, prior literature focuses on different peer groups. If those peer groups are more relevant for alcohol related behavior of adolescents, it is not surprising that we would find smaller effects in our setting. Second, if the siblings in our sample do not spend meaningful amounts of time together, there would be no opportunity for younger siblings to observe, and be affected by, the behavior of their older sibling.

While a small literature on the shared time use of siblings does exist (e.g., Dunifon, Fomby and Musick, 2017; Wikle and Hoagland, 2020), there does not appear to be any work on the age ranges relevant for our analysis. We therefore address this concern with a simple descriptive analysis which demonstrates that siblings at ages near the MLDA spend substantial amounts of time together.

Using the ATUS sample described previously (3,336 ATUS respondents who live with an older sibling under age 24), we calculate average time spent per day with an older sibling. For comparison, we also calculate average time spent per day with friends.[22] Finally, we present these means in several different subsets of activities. First, we move from all activities reported in the time-diary to activities performed while not at work or school—under the assumption that such "discretionary" time is more relevant for the behavior we are interested in. Second, we further restrict our focus to time spent not at work or school and without any parents present for the same reason. Finally, we break down this last category of discretionary and unsupervised time into the main activity groups defined by ATUS. Results are shown in Figure 3.8.

The average respondent reports spending roughly 1 hour and 45 minutes with an older sibling during their diary-day, just under one hour of which is discretionary and unsupervised time, when alcohol consumption would be most likely to occur. Since a substantial portion of the average diary-day in our sample is spent either at work, at school, or with parents (see Table 3.3), this is a meaningful amount of time. Moreover, the activities that siblings engage in together during this discretionary and unsupervised time (right panel of the figure) are broadly similar to the activities engaged in with friends. While it is clear that respondents spend more time with friends (especially within discretionary and supervised time), we feel that this is strong evidence that the sibling pairs we study in our main analysis are a meaningful peer group for the behavior of interest.

3.4.7 Robustness

Our main analyses focus on sibling pairs who currently reside in the same household, are the oldest two siblings in that household, and where the older sibling is between the ages of 19 and 23. In this subsection, we explore the robustness of our results to different sample selection criteria.

Table 3.9 presents results that remove either the requirement that the siblings currently live

[22] We consider an activity to performed with an older sibling (or similarly, with a friend) if the so-called "who" record in the ATUS data lists at least one older sibling (or friend).

together, the requirement that the siblings are the oldest two siblings in the household, or both. The first panel corresponds to Table 3.6, which uses our main sample. Point estimates are broadly consistent (small, negative, not always statistically significant) across these samples but are closer to zero in the alternative samples. We view this as reassuring since a true spillover effect is likely to be larger when the siblings live together and when the peer is the only older sibling present (as opposed to just the closest older sibling).

Table 3.10 presents estimates of the effect of *younger* sibling's access to alcohol on the *older* sibling's alcohol consumption. Since it is less common for sibling pairs in this age range to live together (here the younger sibling is between the ages of 19 and 22), sample sizes under our typical sample definitions are relatively small in this analysis. For this reason we also present these results with alternative samples as in Table 3.9. Compared to Table 3.9, the sign of these point estimates is less consistent (sometimes negative, sometimes positive) and the estimates are broadly less precise—which is more consistent with a true zero effect. One possible explanation is that the proposed expectations mechanism for the negative effects estimated in our main analysis is less plausible in this context—since older siblings are likely to already have substantial experience with alcohol consumption. We view this as additional suggestive evidence in support of our main results.

All results discussed thus far have relied on a bandwidth of 24 months (i.e., restricting the sample to sibling pairs in which the older sibling is between the ages of 19 years and 0 months and 23 years and 0 months). This follows prior literature utilizing MLDA RDDs, which nearly universally chooses this bandwidth. We demonstrate the robustness of our main results to the choice of bandwidth in two ways. First, Figure 3.6 graphs point estimates and confidence intervals from the preferred model[23] for 36 separate bandwidths (single month increments from 12 to 48 months). Results are stable across bandwidth choices, although (expectedly) less precise as band-

[23]Donut RD specification including a vector of controls, a linear polynomial in the running variable fully interacted with the cutoff dummy, and individual level fixed effects

width shrinks. Second, we verify that our results persist when using the "continuity" framework for estimation and inference in RDDs developed by Cattaneo, Idrobo and Titiunik (2020). This approach involves local polynomial estimation of the discontinuity, bias-corrected standard errors, and a data-driven bandwidth selection approach which is mean squared error optimal. Table 3.12 presents the results of the continuity-based RD models. Column (1) of this table is our preferred specification from Table 3.6 (for reference). Column (2) and (4) use MSE-optimal bandwidths, while column (3) uses our ad-hoc 24-month bandwidth. Consistent to Table 3.6, all coefficients are negative with moderate magnitudes. For models with individual fixed effects, the coefficients are large and statistically significant compared to models without fixed effects, a pattern also present in Table 3.6.

Finally, we visually summarize the robustness of our results in a plot of 64 specifications (most of which have been presented in other parts of this paper), which vary by the inclusion of fixed effects, the inclusion of controls, the order of the running variable polynomial, and the sample, in Figure 3.9 and Figure 3.10. Each regression in Figure 3.9 uses an indicator for any binge drinking as the outcome, and each coefficient is scaled by the corresponding "control" mean (mean of the outcome just below the cutoff)—facilitating comparisons with the EGW point estimate of +19%. Among 64 total specifications, 58 scaled estimates are negative, though the magnitudes vary and confidence intervals usually do not rule out a null effect. Among 64 total specifications, 59 confidence intervals (95%) exclude the EGW point estimate. Figure 3.10 shows similar results with the count of binge drinking days as the outcome.

One additional reassuring pattern in this figure is that the handful of outlier CIs (in both figures) which include large positive effects *all* include quadratic polynomials in the running variable. Visual inspection of various binscatters in this paper suggest that the relationship between the outcome and running variable is likely linear in our data. Further, higher-order polynomials in RD designs are known to increase the risk of detecting spurious effects (Gelman and Imbens, 2019).

3.5 Conclusions

We focus on a population which allows for the estimation of causally interpretable endogenous peer effects in alcohol consumption under relatively weak assumptions: sibling pairs close to the MLDA in the United States. This setting is helpful for two reasons. First, the peer group (sibling pairs) is not chosen. Second, the average alcohol consumption of young adults has been shown to increase discontinuously at the MLDA. This allows for the use of a RDD. While somewhat imprecise, estimates suggest that alcohol consumption of younger siblings decreases when the older sibling gains legal access to alcohol. We are consistently able to rule out large positive effects reported in the prior literature.

Several limitations are worth mentioning. First, as in any RDD, external validity is a concern. Our results apply only to a certain peer group, adolescent siblings residing in the same household where the older sibling is near the MLDA. However, our results are less limited in this way than a typical MLDA-based RDD, given that the running variable and the outcome are taken from different individuals (ages range from 12-20 years for younger siblings in our sample whose older sibling is within one month of the cutoff). Second, an older sibling's legal drinking status could *potentially* affect younger sibling consumption indirectly, e.g., via parental responses or access effects. However, one key set of confounders (parental responses) are, to some extent, ruled out based on observable information in the data, and another (access effects) are likely to bias our results towards zero.

Our results emphasize the important distinction between contextual and endogenous peer effects. A key rationale for policy makers to understand peer effects in alcohol consumption and other costly behaviors, is the potential for the costs of those behaviors[24] to be socially multiplied.

[24] Or the benefits of interventions which successfully decrease the prevalence of the behaviors.

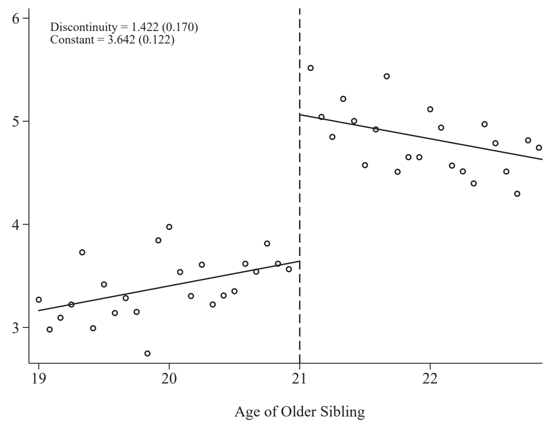
This will occur only if the peer effect in question is endogenous. Many studies on peer effects are only able to identify contextual peer effects. The existence of contextual peer effects in a population does not necessarily imply the existence of endogenous peer effects and policy makers should exercise caution in acting on results which identify only the former and the latter.

An important goal for future work will be to understand the differences between the growing set of well-identified results on peer effects in alcohol consumption. The research designs used in our work, Fletcher and Marksteiner (2017), and the series of papers utilizing randomly assigned college roommates (Duncan et al., 2005; Eisenberg, Golberstein and Whitlock, 2014; Guo et al., 2015) all differ in at least two key dimensions. The peer group studied and the nature of the variation in alcohol consumption exploited by the researcher. Further, the interpretation of the results differs since studies based on randomly assigned roommates identify only contextual peer effects. Future work should aim to understand to what extent these three factors explain the differences in these results and to understand the nature of spillovers in alcohol consumption in more generalizable contexts.

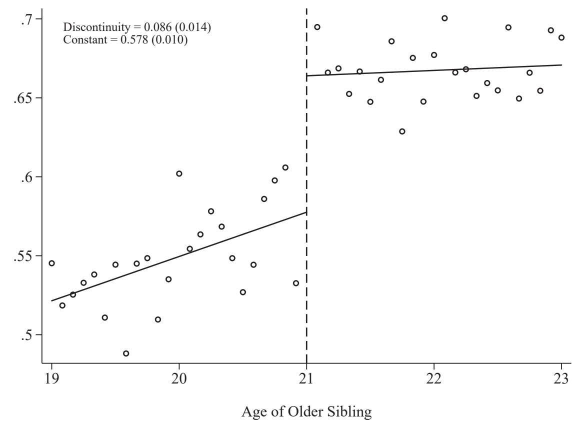
3.6 Figures and Tables

Figure 3.1: Discontinuities in Older Sibling Alcohol Consumption

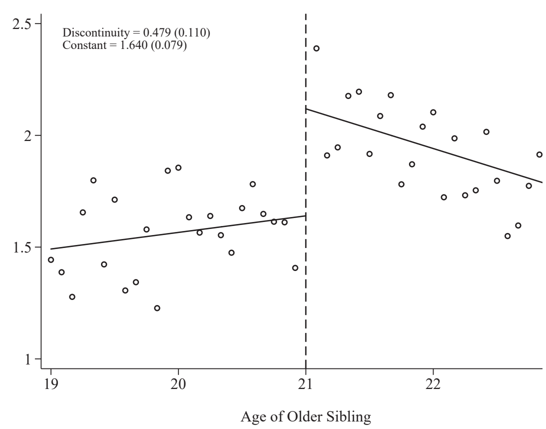
(a) Count of Drinking Days



(b) Prob. of Any Drinking



(c) Count of Binge Drinking Days



(d) Prob. of Any Binge Drinking

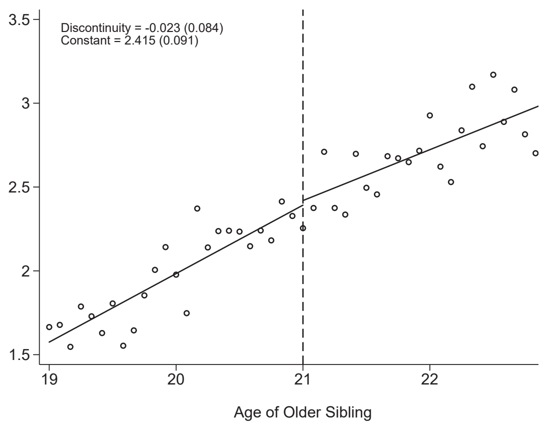
Notes: Each panel shows binscatters of mean alcohol consumption in each older-sibling-age bin (in months), with linear fits estimated separately on each side of the cutoff. Point estimates and standard errors at the top left of each panel are from the corresponding OLS regressions of the relevant alcohol consumption measure age (centered at 21), and an age-21+ indicator, fully interacted (estimated with individual level, not bin-level data). The sample includes all NLSY97 respondents in the age range who have 1 or more siblings and are the oldest siblings in their household.

Figure 3.2: Density of Running Variable

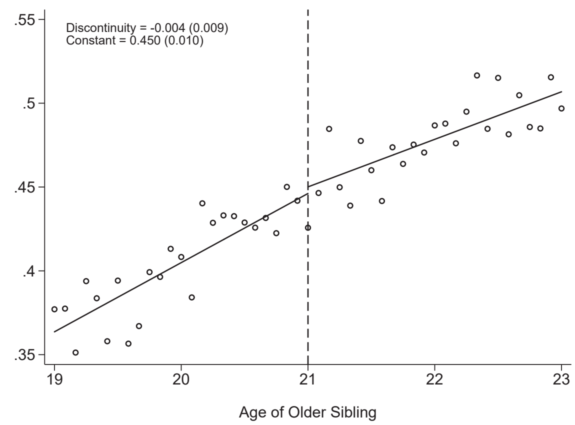
Notes: Y-axis shows the count of sibling-pairs with each running variable value (age in months of older sibling). Point estimates and standard errors are from a regression of the number of observations in each bin on the centered-at-21 running variable, the treatment cutoff, and their interaction.

Figure 3.3: Smoothness of Covariates at Cutoff

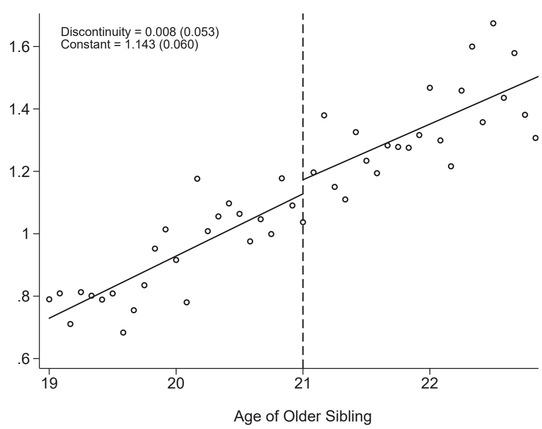
(a) Predicted Count of Drinking Days



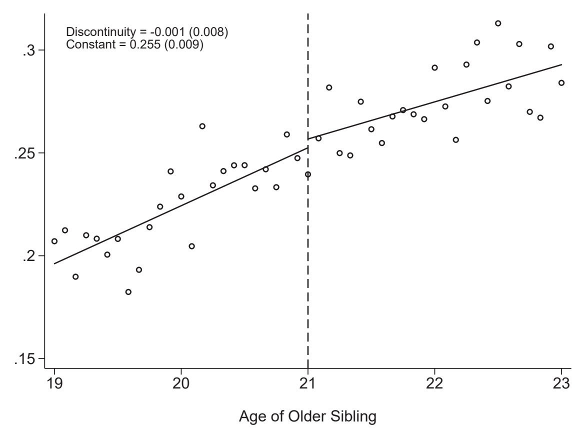
(b) Predicted Prob. of Any Drinking



(c) Predicted Count of Binge Drinking Days



(d) Predicted Prob. of Any Binge Drinking



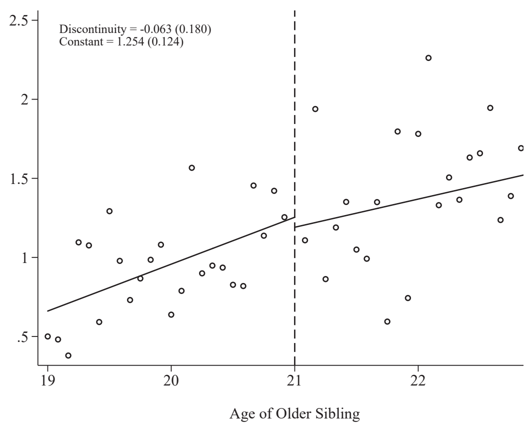
Notes: Each panel shows binscatters of mean predicted alcohol consumption in each older-sibling-age bin (in months), with linear fits estimated separately on each side of the cutoff. Predictions are fitted values from a regression of the alcohol consumption measure on: age, gender of both siblings, educational enrollment, highest completed education, work status, indicators for whether the household lives in an urban area, census region dummies, AFQT score, household size, and interview month/year. (All variables refer to the younger sibling's info unless otherwise noted.) Point estimates and standard errors at the top left of each panel are from corresponding OLS regressions of the prediction on the running variable, the cutoff indicator, and their interaction (estimated with individual level, not bin-level data). Standard errors are calculated via bootstrap. The sample includes all NLSY97 respondents in the age range who are the second oldest siblings in their household.

Figure 3.4: Reduced Form

(a) Count of Drinking Days

(b) Prob. of Any Drinking

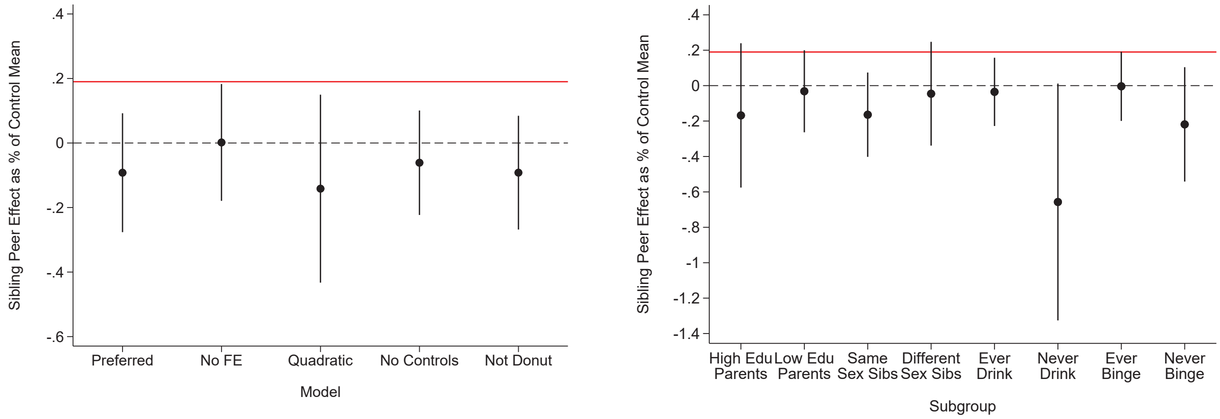
(c) Count of Binge Drinking Days



(d) Prob. of Any Binge Drinking

Notes: Each panel shows binscatters of mean alcohol consumption in each older-sibling-age bin (in months), with linear fits estimated separately on each side of the cutoff. Point estimates and standard errors at the top left of each panel are from corresponding OLS regressions of the younger sibling's alcohol consumption on the running variable, the cutoff indicator, and their interaction (estimated with individual level, not bin-level data). The sample includes all NLSY97 respondents in the age range who are the second oldest siblings in their household.

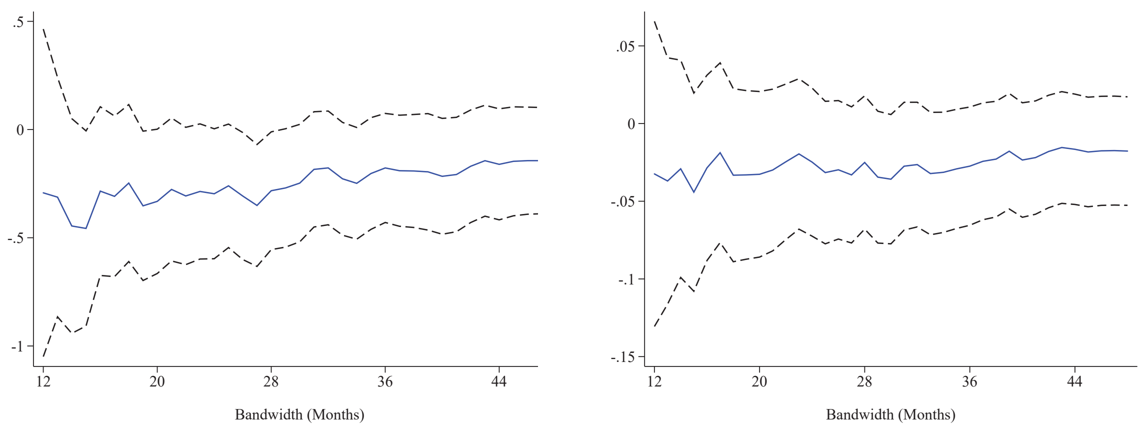
Figure 3.5: Comparison with Eisenberg, Golberstein and Whitlock (2014)



Notes: Figure shows point estimates and confidence intervals from various specifications in Table 3.6 and Table 3.8 alongside the main estimate from EGW (red line). The outcome in each regression (including EGW's) is an indicator for any binge drinking days in the past month.

Figure 3.6: Robustness to Different Bandwidths

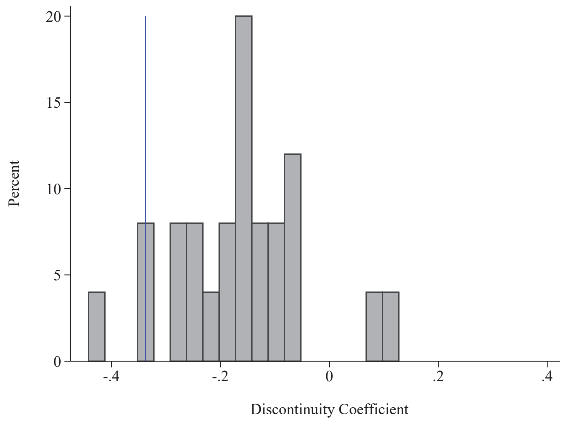
(a) Count of Binge Drinking Days (b) Prob. of Any Binge Drinking



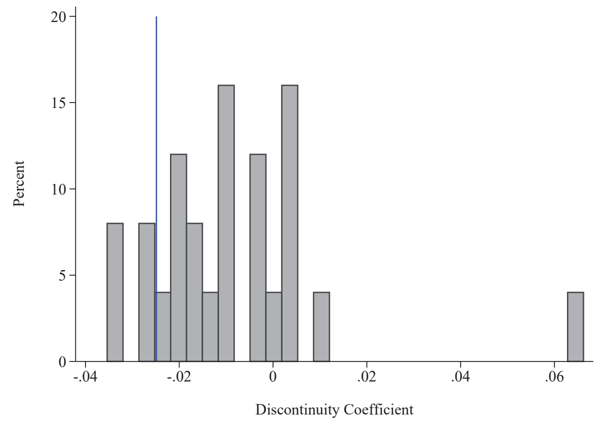
Notes: Figure shows point estimates and confidence intervals from the preferred the specification in Table 3.6 (donut, linear, controls, FE) estimated with varying bandwidths.

Figure 3.7: Placebo Discontinuities

(a) Count of Binge Drinking Days

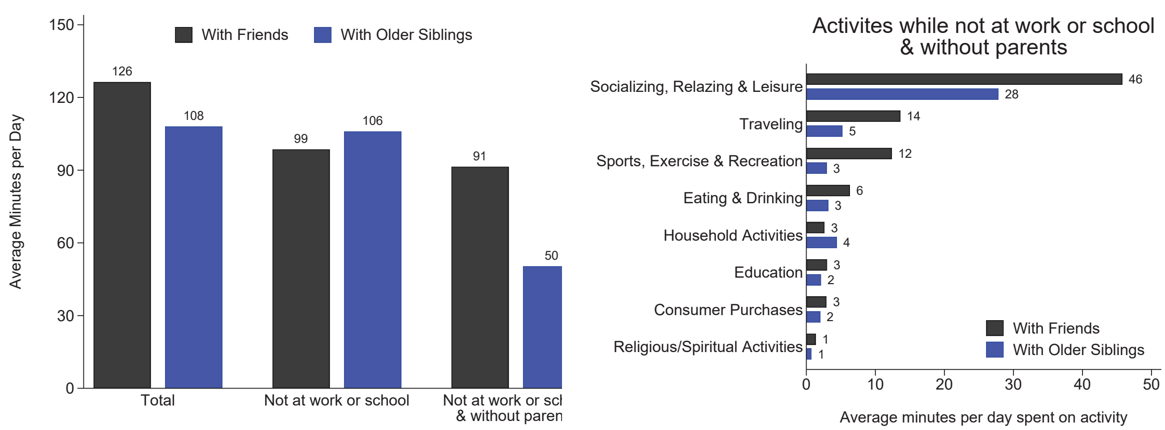


(b) Prob. of Any Binge Drinking



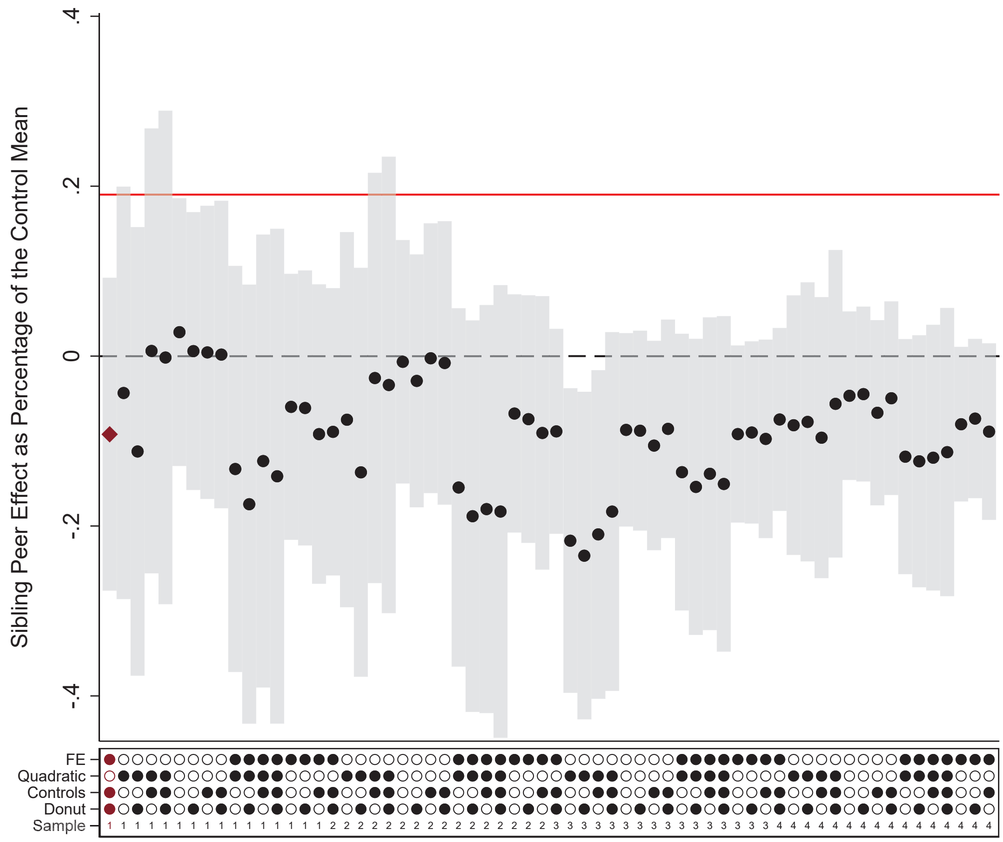
Notes: Figure shows the distribution of point estimates from a series of placebo discontinuity estimates, at running variable values near but not equal to the cutoff. Each underlying regression uses the preferred the specification in Table 3.6 (donut, linear, controls, FE) and a different cutoff up to one year below or above age-21. Actual estimate is shown with a vertical blue line.

Figure 3.8: Time Spent with Older Siblings in the ATUS



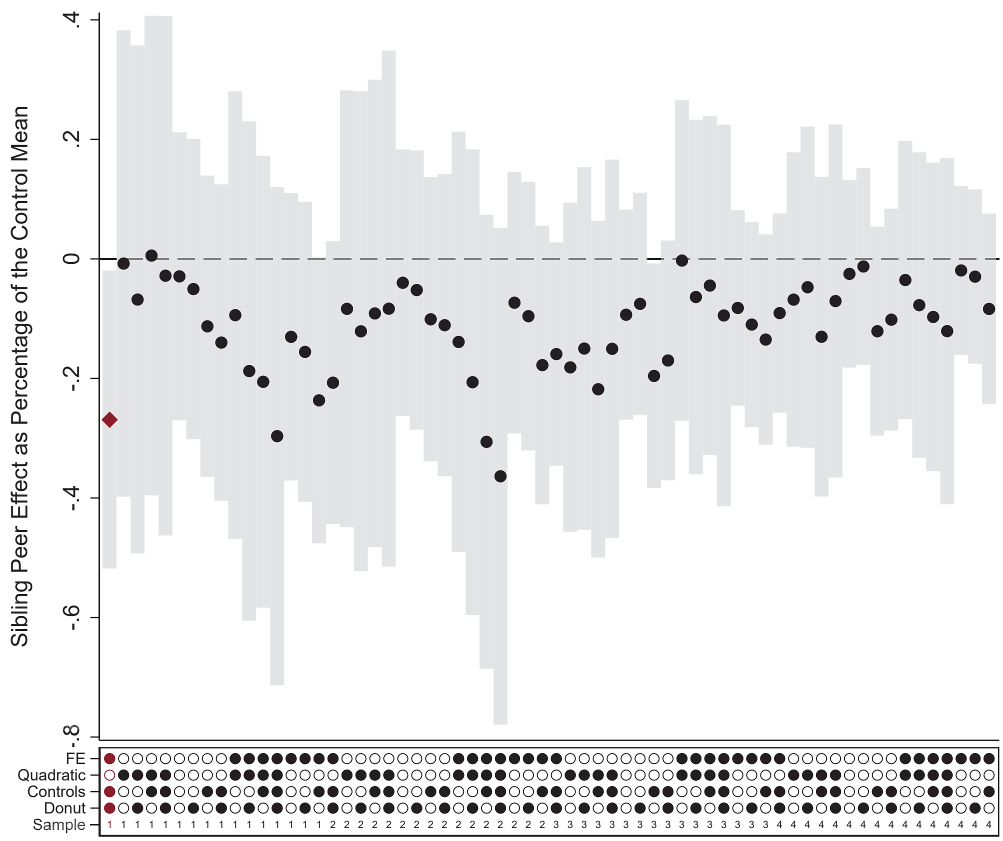
<u>Notes:</u> Descriptive statistics are from a sample of 3,336 respondents in the 2003-2019 waves of the ATUS who have 1 or more older siblings under the age of 24 in their household. Each panel shows the average (across respondent) time per day spent with either friends (blue) or older siblings (black) in various categories.

Figure 3.9: Robustness to Alternative Specifications
(Outcome = Prob. of Any Binge Drinking)



Notes: Figure shows point estimates and confidence intervals from various specifications in Table 3.6, all with an indicator for any binge days as the outcome, and all scaled by the pre-cutoff mean (constant in the corresponding regression on the centered-at-zero running variable, the treatment cutoff, and their interaction). Red line shows the point estimate from Eisenberg, Golberstein and Whitlock (2014) for comparison. "Sample" refers to different subsets of sibling pairs, defined by (i) whether they live together, and (ii) whether they are the oldest two siblings in the household (or family). 1 = same household, oldest two siblings in family. 2 = same household, peer is closest older sibling but not necessarily the only older sibling. 3 = siblings do not necessarily live together, siblings are the oldest two siblings in the family. 4 = siblings do not necessarily live together, peer is closest older sibling but not necessarily the only older sibling.

Figure 3.10: Robustness to Alternative Specifications
(Outcome = Count of Binge Drinking Days)



Notes: Figure shows point estimates and confidence intervals from various specifications in Table 3.6, all with the count of binge drinking days as the outcome, and all scaled by the pre-cutoff mean (constant in the corresponding regression on the centered-at-zero running variable, the treatment cutoff, and their interaction). "Sample" refers to different subsets of sibling pairs, defined by (i) whether they live together, and (ii) whether they are the oldest two siblings in the household (or family). 1 = same household, oldest two siblings in family. 2 = same household, peer is closest older sibling but not necessarily the only older sibling. 3 = siblings do not necessarily live together, siblings are the oldest two siblings in the family. 4 = siblings do not necessarily live together, peer is closest older sibling but not necessarily the only older sibling.

Table 3.1: NLSY97 Sample Summary Statistics

	Full Sample mean/sd	RDD Sample mean/sd
Younger Sibling's Past Month Drinking:		
Drinking Days	2.41	2.31
	(4.98)	(4.76)
Any Drinking Days	0.41	0.42
	(0.49)	(0.49)
Binge Days	1.02	1.14
	(3.03)	(3.25)
Any Binge Days	0.23	0.24
	(0.42)	(0.43)
Younger Sibling's Characteristics:		
Age	18.72	17.99
	(3.97)	(1.72)
Female	0.47	0.48
	(0.50)	(0.50)
Race: Black	0.28	0.26
	(0.45)	(0.44)
Race: Hispanic	0.26	0.24
	(0.44)	(0.43)
In High School	0.53	0.55
	(0.50)	(0.50)
In College	0.15	0.20
	(0.35)	(0.40)
Urban	0.78	0.77
	(0.41)	(0.42)
Worked Past Yr	0.66	0.71
	(0.47)	(0.46)
Household Income (Unit: Dollars)	63,484.53	61,559.45
	(58,841.17)	(55,247.96)
AFQT score	42,842.18	44,616.99
	(28,720.84)	(29,150.67)
Older Sibling's Characteristics:		
Age	22.15	20.87
	(5.29)	(1.17)
Female	0.47	0.47
	(0.50)	(0.50)
Worked Past Yr	0.38	0.39
	(0.49)	(0.49)
Currently Enrolled In School	0.53	0.50
	(0.50)	(0.50)
N	15,230	5,702

Notes: The full sample consists of all NLSY respondents. Our analysis sample ("RDD sample") consists of co-resident sibling pairs who are the oldest siblings in the household, where the younger sibling is a NLSY97 respondent and the older sibling is between the ages of 19 and 23.

Table 3.2: Correlations in Alcohol Consumption Between Siblings

	Younger Sibling Consumption			
	Count of Drinking Days		Any Drinking Days	
	(1)	(2)	(1)	(2)
Older Sibling Consumption	0.161***	0.136***	0.178***	0.168***
	(0.020)	(0.019)	(0.018)	(0.016)
Constant	3.419***	5.074***	0.485***	0.487***
	(0.918)	(0.997)	(0.113)	(0.091)
Mean	1.938	2.356	0.374	0.421
N	3808	6096	3808	6096
	Count of Binge Drinking Days		Any Binge Drinking Days	
Older Sibling Consumption	0.163***	0.133***	0.166***	0.140***
	(0.026)	(0.021)	(0.018)	(0.016)
Constant	2.043**	2.874***	0.326***	0.434***
	(0.642)	(0.604)	(0.091)	(0.081)
Mean	0.881	1.084	0.198	0.236
N	3806	6096	3806	6096
Lag		X		X

Notes: Each model has the younger sibling's consumption as the outcome and the older sibling's consumption as the covariate of interest. Model 1 uses the contemporaneous consumption of the older sibling and excludes sibling pairs in which the older sibling was interviewed after the younger sibling. Model 2 uses the lagged (prior survey year) consumption. All models are estimated via OLS with cluster robust standard errors at the household level and are limited to households with exactly two NLSY97 respondent siblings, in which the older sibling is 23 years old or younger. Lagged models use the prior survey's measure of older sibling consumption. All models include a vector of controls for both siblings similar to those described in the Data section. Means of each outcome (for younger siblings) are shown for each model. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.3: ATUS Sample Summary Statistics

	Mean	SD	Min	Max
Respondent Characteristics:				
Age	17.25	(1.92)	15.0	23.0
Female	0.48	(0.50)	0.0	1.0
In high school	0.57	(0.49)	0.0	1.0
In college	0.19	(0.39)	0.0	1.0
Employed	0.39	(0.49)	0.0	1.0
Surveyed on a weekend or holiday	0.31	(0.46)	0.0	1.0
Age of closest older sibling	20.00	(2.12)	16.0	24.0
Number of older siblings	1.18	(0.43)	1.0	4.0
Minutes Spent (during ATUS-diary day):				
At home	336.35	(224.15)	0.0	1440.0
At school	164.85	(224.15)	0.0	1020.0
At work	71.21	(168.16)	0.0	890.0
Alone	224.11	(199.87)	0.0	1370.0
With an older sibling	109.68	(176.83)	0.0	1200.0
With a parent	145.55	(189.74)	0.0	1200.0
With a friend	129.56	(204.66)	0.0	1140.0
With anyone else	190.24	(230.87)	0.0	1210.0
N	3,336			

Notes: Summary statistics from time-diaries for all ATUS respondents in the 2003-2019 waves who lived with one older sibling 24 years old or younger. All summary statistics calculated using survey weights. Data retrieved from Hofferth et al. (2020).

Table 3.4: Discontinuities in Older Sibling Alcohol Consumption

	Count of Drinking Days				
	(1)	(2)	(3)	(4)	(5)
Age 21+	1.447***	1.585***	1.834***	1.339***	1.328***
	(0.152)	(0.169)	(0.241)	(0.133)	(0.146)
Constant	4.801***	3.205***	4.658***	3.720***	4.785***
	(0.727)	(0.502)	(0.733)	(0.083)	(0.723)
Observations	16682	16682	16682	21376	17049
	Any Drinking Days				
Age 21+	0.083***	0.090***	0.105***	0.083***	0.079***
	(0.013)	(0.013)	(0.019)	(0.011)	(0.012)
Constant	0.545***	0.434***	0.536***	0.585***	0.556***
	(0.057)	(0.041)	(0.058)	(0.008)	(0.056)
Observations	16682	16682	16682	21376	17049
	Count of Binge Drinking Days				
Age 21+	0.454***	0.587***	0.678***	0.404***	0.405***
	(0.100)	(0.111)	(0.161)	(0.089)	(0.095)
Constant	2.406***	1.496***	2.282***	1.667***	2.437***
	(0.463)	(0.322)	(0.464)	(0.051)	(0.469)
Observations	16601	16601	16601	21268	16964
	Any Binge Drinking Days				
Age 21+	0.057***	0.068***	0.073***	0.058***	0.050***
	(0.012)	(0.013)	(0.020)	(0.011)	(0.012)
Constant	0.427***	0.305***	0.423***	0.342***	0.422***
	(0.053)	(0.041)	(0.054)	(0.007)	(0.053)
Observations	16601	16601	16601	21268	16964
Fixed Effects	X		X	X	X
Quadratic			X		
Controls	X	X	X		X
Donut	X	X	X	X	

Notes: Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. All models are estimated in a sample of NLSY97 respondents who are the oldest siblings currently residing in their household who are between the ages of 19 and 23. All models include cluster robust standard errors at individual level. Age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.5: Smoothness of Covariates at Cutoff

	Drinking Days	Any Drinking Days	Binge Drinking Days	Any Binge Drinking Days
Constant	2.415***	0.450***	1.143***	0.255***
	(0.091)	(0.010)	(0.060)	(0.009)
Sibling 21+	-0.023	-0.004	0.008	-0.001
	(0.084)	(0.009)	(0.053)	(0.008)
N	4373	4373	4369	4369

Notes: We test for smoothness of predicted values of outcomes through the cutoff with bootstrapped standard errors. Outcome are predicted with a set of covariates, including the month and year of the survey, the race of the respondent, the gender of the respondent and siblings, educational attainment and enrollment of the respondent, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. Smoothness is tested by regressing each predicted value of outcome variable on the age of the older sibling in months, whether the older sibling is over 21, and their interaction. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.6: Reduced form

	(1)	(2)	(3)	(4)	(5)
	Count of Drinking Days				
Sibling 21+	-0.137	0.098	-0.288	0.156	-0.058
	(0.249)	(0.249)	(0.431)	(0.229)	(0.241)
Constant	2.577	4.555***	2.505	2.389***	2.885
	(2.023)	(0.830)	(2.008)	(0.150)	(1.890)
Observations	4282	4282	4282	5527	4373
	Any Drinking Days				
Sibling 21+	-0.031	-0.009	-0.059	0.004	-0.035
	(0.028)	(0.028)	(0.046)	(0.025)	(0.027)
Constant	0.828***	0.657***	0.840***	0.459***	0.904***
	(0.171)	(0.089)	(0.173)	(0.016)	(0.163)
Observations	4282	4282	4282	5527	4373
	Count of Binge Drinking Days				
Sibling 21+	-0.337*	-0.176	-0.372	-0.195	-0.297+
	(0.159)	(0.169)	(0.267)	(0.161)	(0.153)
Constant	1.512	2.846***	1.440	1.313***	1.567
	(1.485)	(0.599)	(1.465)	(0.107)	(1.383)
Observations	4278	4278	4278	5521	4369
	Any Binge Drinking Days				
Sibling 21+	-0.025	0.000	-0.038	-0.017	-0.025
	(0.025)	(0.025)	(0.040)	(0.022)	(0.024)
Constant	0.408*	0.439***	0.402*	0.282***	0.443*
	(0.183)	(0.078)	(0.185)	(0.015)	(0.184)
Observations	4278	4278	4278	5521	4369
FE	X		X	X	X
Quadratic			X		
Controls	X	X	X		X
Donut	X	X	X	X	

Notes: Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. All models are estimated in a sample of NLSY97 respondents who are the second oldest siblings, where their oldest sibling is between the ages of 19 and 23 and resides in the same household with them. All models include cluster robust standard errors at individual level. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.7: Reduced Form in Subsamples, Binge Days

	Parental Education		Sibling Sex Composition	
	Low (either $<$ HS)	High (both $\geq$ HS)	Same Sex Siblings	Different Sex Siblings
Sibling 21+	-0.124	-0.417^{+}	-0.616**	-0.040
	(0.277)	(0.217)	(0.231)	(0.223)
Constant	-1.695	2.791	-0.194	0.614
	(1.476)	(2.454)	(1.812)	(1.716)
Observations	1045	2206	2379	1899
	Older Sibling's Sex		Sibling Age Diff	
	Male Older Sibling	Female Older Sibling	$\geq$ 30 Months	$<$ 30 Months
Sibling 21+	-0.662**	0.064	-0.439*	-0.290
	(0.245)	(0.208)	(0.202)	(0.261)
Constant	-6.995*	-0.268	1.284	13.177***
	(2.725)	(1.087)	(1.583)	(3.087)
Observations	2266	2012	2312	1966
	Experience with Alcohol (younger sibling in prior year)		Experience with Alcohol (younger sibling in prior year)	
	Ever Drinkers	Never Drinkers	Ever Binge Drinkers	Never Binge Drinkers
Sibling 21+	-0.488*	-0.166	-0.644	-0.195^{+}
	(0.243)	(0.117)	(0.560)	(0.112)
Constant	-0.325	0.750	0.869	1.005
	(1.845)	(0.483)	(2.909)	(0.672)
Observations	2511	1492	1069	2783

Notes: Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. All models are estimated in a sample of NLSY97 respondents who are the second oldest siblings, where their oldest sibling is between the ages of 19 and 23 and resides in the same household with them. All models include cluster robust standard errors at individual level. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.8: Reduced Form in Subsamples, Any Binge Days

	Parental Education		Sibling Sex Composition	
	Low (either $<$ HS)	High (both $\geq$ HS)	Same Sex Siblings	Different Sex Siblings
Sibling 21+	-0.046	-0.009	-0.046	-0.012
	(0.057)	(0.034)	(0.034)	(0.038)
Constant	-0.000	0.645^{**}	0.011	0.795
	(0.305)	(0.239)	(0.196)	(0.540)
Observations	1045	2206	2379	1899
	Older Sibling's Sex		Sibling Age Diff	
	Male Older Sibling	Female Older Sibling	$\geq$ 30 Months	$<$ 30 Months
Sibling 21+	-0.039	-0.009	-0.020	-0.018
	(0.035)	(0.038)	(0.035)	(0.037)
Constant	-2.136^{***}	0.436	0.979^{*}	0.615^{*}
	(0.254)	(0.370)	(0.408)	(0.312)
Observations	2266	2012	2312	1966
	Experience with Alcohol (younger sibling in prior year)		Experience with Alcohol (younger sibling in prior year)	
	Ever Drinkers	Never Drinkers	Ever Binge Drinkers	Never Binge Drinkers
Sibling 21+	-0.014	-0.043^{+}	-0.002	-0.036
	(0.039)	(0.023)	(0.060)	(0.027)
Constant	0.324	0.572^{+}	-0.000	0.475^{**}
	(0.204)	(0.308)	(0.451)	(0.183)
Observations	2511	1492	1069	2783

Notes: Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. All models are estimated in a sample of NLSY97 respondents who are the second oldest siblings, where their oldest sibling is between the ages of 19 and 23 and resides in the same household with them. All models include cluster robust standard errors at individual level. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.9: Alternative samples

	Drinking Days	Any Drinking Days	Binge Drinking Days	Any Binge Drinking Days
	Two oldest siblings living in the same household			
Sibling 21+	-0.137	-0.031	-0.337*	-0.025
	(0.249)	(0.028)	(0.159)	(0.025)
Constant	2.577	0.828***	1.512	0.408*
	(2.023)	(0.171)	(1.485)	(0.183)
Observations	4282	4282	4278	4278
	Two oldest siblings			
Sibling 21+	0.026	-0.025	-0.252+	-0.024
	(0.229)	(0.026)	(0.147)	(0.023)
Constant	5.618**	0.659***	2.205*	0.580*
	(1.861)	(0.195)	(1.069)	(0.247)
Observations	5100	5100	5094	5094
	Two siblings living in the same household			
Sibling 21+	-0.230	-0.015	-0.202+	-0.026
	(0.186)	(0.020)	(0.121)	(0.018)
Constant	2.232	0.451***	0.564	0.296**
	(1.709)	(0.104)	(1.139)	(0.099)
Observations	6988	6988	6975	6975
	Two siblings			
Sibling 21+	-0.079	-0.025	-0.107	-0.020
	(0.161)	(0.016)	(0.101)	(0.015)
Constant	2.354+	0.190*	0.248	0.525***
	(1.218)	(0.077)	(0.820)	(0.074)
Observations	9975	9975	9958	9958

Notes: The first panel is estimated in a sample of NLSY97 respondents who are the second oldest siblings, where their oldest sibling is between the ages of 19 and 23 and resides in the same household with them. The second panel is estimated in a sample of NLSY97 respondents who are the second oldest siblings, where their oldest sibling is between the ages of 19 and 23, regardless of whether the oldest sibling resides in the same household with them. The third panel is estimated in a sample of NLSY97 respondents where their older sibling is between the ages of 19 and 23 and resides in the same household with them, regardless of whether they are the second oldest sibling and regardless of whether the older sibling is the oldest sibling. The last panel is estimated in a sample of NLSY respondents where their older sibling is between the ages of 19 and 23, regardless of whether they are the second oldest sibling, regardless of whether the older sibling is the oldest sibling, and regardless of whether the older sibling resides in the same household with them. Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether the respondent has children, and an indicator for whether the respondent worked in the past year. All models include cluster robust standard errors at individual level. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.10: Effect of younger sibling on older sibling

	Drinking Days	Any Drinking Days	Binge Drinking Days	Any Binge Drinking Days
	Two oldest siblings living in the same household			
Sibling 21+	-0.117	0.003	0.060	0.055
	(0.467)	(0.034)	(0.258)	(0.036)
Constant	12.259***	0.236	0.620	0.122
	(2.778)	(0.260)	(1.550)	(0.260)
Observations	2380	2380	2333	2333
	Two oldest siblings			
Sibling 21+	-0.013	0.016	-0.016	0.062*
	(0.325)	(0.026)	(0.214)	(0.029)
Constant	10.813**	0.513**	2.064+	0.152
	(3.939)	(0.177)	(1.158)	(0.277)
Observations	3758	3758	3702	3702
	Two siblings living in the same household			
Sibling 21+	0.006	0.007	-0.112	0.001
	(0.227)	(0.017)	(0.136)	(0.018)
Constant	4.968***	0.648***	2.764**	0.353*
	(1.420)	(0.119)	(1.032)	(0.139)
Observations	7557	7557	7398	7398
	Two siblings			
Sibling 21+	-0.056	0.002	0.000	0.011
	(0.163)	(0.013)	(0.101)	(0.013)
Constant	5.380***	0.670***	2.464***	0.362***
	(1.022)	(0.098)	(0.728)	(0.097)
Observations	13367	13367	13145	13145

Notes: The first panel is estimated in a sample of NLSY97 respondents who are the oldest siblings, where their second oldest sibling is between the ages of 19 and 23 and resides in the same household with them. The second panel is estimated in a sample of NLSY97 respondents who are the oldest siblings, where their second oldest sibling is between the ages of 19 and 23, regardless of whether the second oldest sibling resides in the same household with them. The third panel is estimated in a sample of NLSY97 respondents where their younger sibling is between the ages of 19 and 23 and resides in the same household with them, regardless of whether they are the oldest sibling and regardless of whether the younger sibling is the second oldest sibling. The last panel is estimated in a sample of NLSY respondents where their younger sibling is between the ages of 19 and 23, regardless of whether they are the oldest sibling, regardless of whether the younger sibling is the second oldest sibling, and regardless of whether the younger sibling resides in the same household with them. Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether the respondent has children, and an indicator for whether the respondent worked in the past year. All models include cluster robust standard errors at individual level. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.11: Parenting style changes at the cutoff

	Degree of Parental Monitoring				
	(1)	(2)	(3)	(4)	(5)
Sibling 21+	-0.243	-0.722*	-0.311	-0.421	-0.198
	(0.321)	(0.354)	(0.508)	(0.292)	(0.294)
Constant	11.307***	9.525***	11.337***	8.653***	9.482***
	(2.290)	(1.418)	(2.324)	(0.167)	(2.353)
Observations	1467	1467	1467	1937	1493
	Parents: Authoritarian/Authoritative				
Sibling 21+	-0.010	-0.000	-0.017	0.007	-0.010
	(0.025)	(0.023)	(0.041)	(0.024)	(0.024)
Constant	0.348	-0.135*	0.369	0.262***	0.347
	(0.364)	(0.069)	(0.369)	(0.016)	(0.368)
Observations	4317	4317	4317	5575	4408

Notes: Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. All models are estimated in a sample of NLSY97 respondents who are the second oldest siblings currently residing in their household, where the oldest sibling is between the ages of 19 and 23. All models include cluster robust standard errors at individual level. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Table 3.12: Continuity Models

	Count of Binge Drinking Days				
Sibling 21+	-0.337*	-0.451***	-0.443**	-0.120	-0.008
	(0.159)	(0.108)	(0.134)	(0.251)	(0.308)
N	4278	2874	4124	3931	4124
Bandwidth	24.00	16.64	24.00	22.80	24.00
	Any Binge Drinking Days				
Sibling 21+	-0.025	-0.032*	-0.037+	0.007	0.019
	(0.025)	(0.015)	(0.021)	(0.035)	(0.045)
N	4278	2555	4124	4278	4124
Bandwidth	24.00	14.07	24.00	24.73	24.00
Continuity		X	X	X	X
FE	X	X	X		
Quadratic					
Controls	X	X	X	X	X
Donut	X	X	X	X	X

Notes: Controls include the month and year of the survey, the gender and age of the respondent and siblings, race of respondents, educational attainment and enrollment of respondents, geography (urban/rural, census region), household size, AFQT score of the respondent, an indicator for whether or not the respondent has children, and an indicator for whether the respondent worked in the past year. All models except for model 1 include bias-corrected estimates and standard errors which are cluster robust at individual level. (Model 1 is the same with model 1 in **??**.) Models 2 and 4 use MSE-optimal bandwidths, and the other models use our ad-hoc 24 month bandwidth. Sibling age is centered at 21 years so that a rough estimate of the percentage increase at the cutoff for any model can be obtained by dividing by the constant term in the corresponding model without controls. +, *, **, and *** denote statistical significance at the 10%, 5%, 1%, and 0.1% levels, respectively.

Bibliography

Altmejd, Adam, Andrés Barrios-Fernández, Marin Drlje, Joshua Goodman, Michael Hurwitz, Dejan Kovac, Christine Mulhern, Christopher Neilson, and Jonathan Smith. 2021. "O brother, where start thou? Sibling spillovers on college and major choice in four countries." *The Quarterly Journal of Economics*.

Altonji, Joseph G, Sarah Cattan, and Iain Ware. 2016. "Identifying sibling influence on teenage substance use." *Journal of Human Resources*, 0714–6474R1.

Anderson, Patricia M, and Bruce D Meyer. 1997. "Unemployment Insurance Takeup Rates and the After-tax Value of Benefits." *Quarterly Journal of Economics*, 112(3): 913–937.

Anderson, Patricia M, and Bruce D Meyer. 2000. "The effects of the unemployment insurance payroll tax on wages, employment, claims and denials." *Journal of public Economics*, 78(1-2): 81–106.

Armour, Philip. 2018. "The Role of Information in Disability Insurance Application: An Analysis of the Social Security Statement Phase-In." *American Economic Journal: Economic Policy*, 10(3): 1–41.

Auray, Stephane, David L Fuller, and Damba Lkhagvasuren. 2019. "Unemployment insurance take-up rates in an equilibrium search model." *European Economic Review*, 112: 1–31.

Baily, Martin Neil. 1978. "Some Aspects of Optimal Unemployment Insurance." *Journal of Public Economics*, 10(3): 379–402.

Barr, Andrew, and Sarah Turner. 2018. "A Letter and Encouragement: Does Information Increase Postsecondary Enrollment of UI Recipients?" *American Economic Journal: Economic Policy*, 10(3): 42–68.

Bartfeld, Judith, Craig Gundersen, Timothy Smeeding, and James Ziliak. 2015. *SNAP Matters: How Food Stamps Affect Health and Well-Being.* Stanford University Press.

Bell, Alex, Thomas J Hedin, Geoffrey Schnorr, and Till Von Wachter. 2020. "An Analysis of Unemployment Insurance Claims in California During the COVID-19 Pandemic." *California Policy Lab policy brief.*

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-differences Estimates?" *Quarterly Journal of Economics*, 119(1): 249–275.

Bingley, Paul, Petter Lundborg, and Stéphanie Vincent Lyk-Jensen. 2021. "Brothers in Arms Spillovers from a Draft Lottery." *Journal of Human Resources*, 56(1): 225–268.

Bitler, Marianne, and Hilary Hoynes. 2016. "The More Things Change, the More They Stay the Same? The Safety Net and Poverty in the Great Recession." *Journal of Labor Economics*, 34(S1): S403–S444.

Blank, Rebecca M, and David E Card. 1991. "Recent trends in insured and uninsured unemployment: is there an explanation?" *The Quarterly Journal of Economics*, 106(4): 1157–1189.

Boone, Christopher, Arindrajit Dube, Lucas Goodman, and Ethan Kaplan. 2021. "Unemployment insurance generosity and aggregate employment." *American Economic Journal: Economic Policy*, 13(2): 58–99.

Breining, Sanni Nørgaard. 2014. "The presence of ADHD: Spillovers between siblings." *Economics Letters*, 124(3): 469–473.

Burda, Michael, Katie R Genadek, and Daniel S Hamermesh. 2020. "Unemployment and Effort at Work." *Economica*, 87(347): 662–681.

Bureau of Labor Statistics. 2003-2014. "Local Area Unemployment Statistics." U.S. Department of Labor. https://www.bls.gov/lau/, accessed 2017-11-05.

Bureau of Labor Statistics. 2015. "Time spent working by full- and part-time status, gender, and location in 2014." U.S. Department of Labor. *The Economics Daily*, https://www.bls.gov/opub/ted/2015/time-spent-working-by-full-and-part-time-status-gender-and-location-in-2014.htm, accessed 2020-09-24.

Bureau of Labor Statistics. 2017*a*. "Labor Force Statistics from the Current Population Survey." U.S. Department of Labor. https://www.bls.gov/cps/aa2017/cpsaat08.htm, accessed 2020-09-24.

Bureau of Labor Statistics. 2017*b*. "May 2017 National Occupational Employment and Wage Estimates." U.S. Department of Labor. https://www.bls.gov/oes/2017/may/oes_nat.htm, accessed 2020-09-24.

Caldwell, Sydnee, Scott Nelson, and Daniel Waldinger. 2020. "Tax Refund Uncertainty: Evidence and Welfare Implications." *Working paper*.

Card, David, Andrew Johnston, Pauline Leung, Alexandre Mas, and Zhuan Pei. 2015. "The Effect of Unemployment Benefits on the Duration of Unemployment Insurance Receipt: New Evidence from a Regression Kink Design in Missouri, 2003-2013." *American Economic Review*, 105(5): 126–30.

Carpenter, Christopher, and Carlos Dobkin. 2009. "The effect of alcohol consumption on mortality: regression discontinuity evidence from the minimum drinking age." *American Economic Journal: Applied Economics*, 1(1): 164–182.

Carpenter, Christopher, and Carlos Dobkin. 2011. "The minimum legal drinking age and public health." *The Journal of Economic Perspectives*, 25(2): 133–156.

Carpenter, Christopher, and Carlos Dobkin. 2015. "The minimum legal drinking age and crime." *Review of economics and statistics*, 97(2): 521–524.

Carpenter, Christopher, and Carlos Dobkin. 2017. “The minimum legal drinking age and morbidity in the United States.” *Review of Economics and Statistics*, 99(1): 95–104.

Carpenter, Christopher S, Carlos Dobkin, and Casey Warman. 2016. “The mechanisms of alcohol control.” *Journal of human resources*, 51(2): 328–356.

Carrell, Scott E, Mark Hoekstra, and James E West. 2011. “Does drinking impair college performance? Evidence from a regression discontinuity approach.” *Journal of public Economics*, 95(1): 54–62.

Cattaneo, Matias D, Nicolás Idrobo, and Rocío Titiunik. 2019. *A practical introduction to regression discontinuity designs: Foundations.* Cambridge University Press.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations. Elements in Quantitative and Computational Methods for the Social Sciences*, Cambridge University Press.

Cawley, John, Euna Han, Jiyoon Kim, and Edward C Norton. 2019. “Testing for family influences on obesity: The role of genetic nurture.” *Health economics*, 28(7): 937–952.

Chetty, Raj. 2006. “A General Formula for the Optimal Level of Social Insurance.” *Journal of Public Economics*, 90(10): 1879–1901.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. “Salience and Taxation: Theory and Evidence.” *American Economic Review*, 99(4): 1145–77.

Chetty, Raj, John N Friedman, and Emmanuel Saez. 2013. “Using Differences in Knowledge across Neighborhoods to Uncover the Impacts of the EITC on Earnings.” *American Economic Review*, 103(7): 2683–2721.

Chiu, W Henry, and Edi Karni. 1998. “Endogenous Adverse Selection and Unemployment Insurance.” *Journal of Political Economy*, 106(4): 806–827.

Chodorow-Reich, Gabriel, John Coglianese, and Loukas Karabarbounis. 2019. “The Macro Effects of Unemployment Benefit Extensions: A Measurement Error Approach.” *Quarterly Journal of Economics*, 134(1): 227–279.

Christofides, Louis N, and Chris J McKenna. 1995. “Unemployment Insurance and Moral Hazard in Employment.” *Economics Letters*, 49(2): 205–210.

Christofides, Louis N, and Chris J McKenna. 1996. “Unemployment Insurance and Job Duration in Canada.” *Journal of Labor Economics*, 14(2): 286–312.

Clark, Andrew E, and Youenn Lohéac. 2007. ““It wasn’t me, it was them!” Social influence in risky behavior by adolescents.” *Journal of health economics*, 26(4): 763–784.

Cohen, Alma, and Rajeev Dehejia. 2004. “The Effect of Automobile Insurance and Accident Liability Laws on Traffic Fatalities.” *Journal of Law and Economics*, 47(2): 357–393.

Coile, Courtney, Peter Diamond, Jonathan Gruber, and Alain Jousten. 2002. “Delays in claiming social security benefits.” *Journal of Public Economics*, 84(3): 357–385.

Crost, Benjamin, and Daniel I Rees. 2013. “The minimum legal drinking age and marijuana use: New estimates from the NLSY97.” *Journal of health economics*, 32(2): 474–476.

Crost, Benjamin, and Santiago Guerrero. 2012. "The effect of alcohol availability on marijuana use: Evidence from the minimum legal drinking age." *Journal of health economics*, 31(1): 112–121.

Currie, Janet Marion. 2006. "The take-up of social benefits." In *Public Policy and the Distribution of Income*. 80–148. Russell Sage Foundation.

Dahl, Gordon B, Katrine V Løken, and Magne Mogstad. 2014. "Peer effects in program participation." *American Economic Review*, 104(7): 2049–74.

Dave, Dhaval, and Robert Kaestner. 2009. "Health Insurance and Ex ante Moral Hazard: Evidence from Medicare." *International Journal of Health Care Finance and Economics*, 9: 367–390.

Daysal, N Meltem, Marianne Simonsen, Mircea Trandafir, and Sanni Breining. 2019. "Spillover effects of early-life medical interventions." *The Review of Economics and Statistics*, 1–46.

Decker, Sandra L. 2005. "Medicare and the Health of Women with Breast Cancer." *Journal of Human Resources*, 40(4): 948–968.

DeJong, William, and Jason Blanchette. 2014. "Case closed: research evidence on the positive public health impact of the age 21 minimum legal drinking age in the United States." *Journal of Studies on Alcohol and Drugs, Supplement*, , (s17): 108–115.

Deza, Monica. 2015. "The effects of alcohol on the consumption of hard drugs: regression discontinuity evidence from the National Longitudinal Study of Youth, 1997." *Health economics*, 24(4): 419–438.

Dong, Yingying. 2015. "Regression discontinuity applications with rounding errors in the running variable." *Journal of Applied Econometrics*, 30(3): 422–446.

Duncan, Greg J, Johanne Boisjoly, and Kathleen Mullan Harris. 2001. "Sibling, peer, neighbor, and schoolmate correlations as indicators of the importance of context for adolescent development." *Demography*, 38(3): 437–447.

Duncan, Greg J, Johanne Boisjoly, Michael Kremer, Dan M Levy, and Jacque Eccles. 2005. "Peer effects in drug use and sex among college students." *Journal of abnormal child psychology*, 33(3): 375–385.

Dunifon, Rachel, Paula Fomby, and Kelly Musick. 2017. "Siblings and children's time use in the United States." *Demographic Research*, 37: 1611–1624.

Ebenstein, Avraham, and Kevin Stange. 2010. "Does Inconvenience Explain Low Take-up? Evidence from Unemployment Insurance." *Journal of Policy Analysis and Management*, 29(1): 111–136.

Eisenberg, Daniel, Ezra Golberstein, and Janis L Whitlock. 2014. "Peer effects on risky behaviors: New evidence from college roommate assignments." *Journal of health economics*, 33: 126–138.

Ejrnæs, Mette, and Stefan Hochguertel. 2013. "Is Business Failure Due to Lack of Effort? Empirical Evidence from a Large Administrative Sample." *Economic Journal*, 123(571): 791–830.

Employment Development Department. 2008-2011. "Unemployment Insurance Initial Claims [database]." California Labor and Workforce Development Agency. Accessed 2019-12-27.

Employment & Training Administration. 2003-2014*a*. "Benefit Accuracy Measurement." U.S. Department of Labor. https://oui.doleta.gov/unemploy/bam/2002/bam_fact.asp, accessed 2020-09-24.

Employment & Training Administration. 2003-2014*b*. "Data Downloads, Report 207." U.S. Department of Labor. https://oui.doleta.gov/unemploy/DataDownloads.asp, accessed 2020-01-29.

Employment & Training Administration. 2019. "Comparison of State Unemployment Laws." U.S. Department of Labor. https://oui.doleta.gov/unemploy/comparison/2010-2019/comparison2019.asp, accessed 2020-09-24.

Fagan, Abigail A, and Jake M Najman. 2005. "The relative contributions of parental and sibling substance use to adolescent tobacco, alcohol, and other drug use." *Journal of Drug issues*, 35(4): 869–883.

Farber, Henry S, and Robert G Valletta. 2015. "Do Extended Unemployment Benefits Lengthen Unemployment Spells? Evidence from Recent Cycles in the US Labor Market." *Journal of Human Resources*, 50(4): 873–909.

Farber, Henry S, Jesse Rothstein, and Robert G Valletta. 2015. "The Effect of Extended Unemployment Insurance Benefits: Evidence from the 2012–2013 Phase-Out." *American Economic Review*, 105(5): 171–176.

Finkelstein, Amy, and Matthew J Notowidigdo. 2019. "Take-up and targeting: Experimental evidence from SNAP." *The Quarterly Journal of Economics*, 134(3): 1505–1556.

Fletcher, Jason, and Ryne Marksteiner. 2017. "Causal spousal health spillover effects and implications for program evaluation." *American Economic Journal: Economic Policy*, 9(4): 144–66.

Fletcher, Jason M. 2012. "Peer influences on adolescent alcohol consumption: evidence using an instrumental variables/fixed effect approach." *Journal of Population Economics*, 25(4): 1265–1286.

Flood, Sarah M, Miriam King, Steven Ruggles, and Robert J. Waren. 2017. "Integrated Public Use Microdata Series, Current Population Survey: Version 5.0 [dataset]." *Minneapolis, MN: IPUMS*. https://doi.org/10.18128/D030.V5.0.

Food and Nutrition Service. 2008-2011. "SNAP Data Tables, report FNS-388." U.S. Department of Agriculture. https://www.fns.usda.gov/pd/supplemental-nutrition-assistance-program-snap, accessed 2018-03-23.

Fortin, Bernard, and Paul Lanoie. 2000. "Incentive Effects of Workers' Compensation: A Survey." In *Handbook of Insurance*. 421–458.

Ganong, Peter, and Pascal Noel. 2019. "Consumer spending during unemployment: Positive and normative implications." *American Economic Review*, 109(7): 2383–2424.

Gaviria, Alejandro, and Steven Raphael. 2001. "School-based peer effects and juvenile behavior." *The review of economics and statistics*, 83(2): 257–268.

Gelman, Andrew, and Guido Imbens. 2019. "Why high-order polynomials should not be used in regression discontinuity designs." *Journal of Business & Economic Statistics*, 37(3): 447–456.

Glover, Dylan, Amanda Pallais, and William Pariente. 2017. "Discrimination as a Self-fulfilling Prophecy: Evidence from French Grocery Stores." *Quarterly Journal of Economics*, 132(3): 1219–1260.

Goerke, Laszlo. 2000. "On the Structure of Unemployment Benefits in Shirking Models." *Labour Economics*, 7(3): 283–295.

Goodman, Jack. 2008. "Who Does the Grocery Shopping, and When Do They Do It?" The Time Use Institute White Paper.

Goodman, Joshua, Michael Hurwitz, Jonathan Smith, and Julia Fox. 2015. "The relationship between siblings' college choices: Evidence from one million SAT-taking families." *Economics of Education Review*, 48: 75–85.

Google Trends. 2008-2009. https://trends.google.com/, accessed 2019-07-31.

Graves, John A. 2012. "Better methods will be needed to project incomes to estimate eligibility for subsidies in health insurance exchanges." *Health Affairs*, 31(7): 1613–1622.

Green, David A, and Timothy C Sargent. 1998. "Unemployment Insurance and Job Durations: Seasonal and Non-seasonal Jobs." *Canadian Journal of Economics*, 31(2): 247–278.

Griffith, Rachel, Martin O'Connell, and Kate Smith. 2016. "Shopping Around: How Households Adjusted Food Spending Over the Great Recession." *Economica*, 83(330): 247–280.

Guo, Audrey. 2021. "The Effects of State Business Taxes on Plant Closures: Evidence from Unemployment Insurance Taxation and Multi-Establishment Firms." *The Review of Economics and Statistics*, 1–45.

Guo, Guang, Yi Li, Craig Owen, Hongyu Wang, and Greg J Duncan. 2015. "A natural experiment of peer influences on youth alcohol use." *Social science research*, 52: 193–207.

Hagedorn, Marcus, Fatih Karahan, Iourii Manovskii, and Kurt Mitman. 2013. "Unemployment Benefits and Unemployment in the Great Recession: The Role of Macro Effects." National Bureau of Economic Research, Working Paper No. w19499.

Hansen, Benjamin, and Glen R Waddell. 2018. "Legal access to alcohol and criminality." *Journal of health economics*, 57: 277–289.

Hansen, Benjamin, Tuan Nguyen, and Glen R Waddell. 2017. "Benefit Generosity and Injury Duration: Quasi-Experimental Evidence from Regression Kinks." IZA Discussion Paper No. 10621.

Hardy, Bradley, and James P Ziliak. 2014. "Decomposing trends in income volatility: The "wild ride" at the top and bottom." *Economic Inquiry*, 52(1): 459–476.

Hardy, Bradley L. 2017. "Income instability and the response of the safety net." *Contemporary Economic Policy*, 35(2): 312–330.

Harris, Jeffrey E, and Beatriz González López-Valcárcel. 2008. "Asymmetric peer effects in the analysis of cigarette smoking among young people in the United States, 1992–1999." *Journal of health economics*, 27(2): 249–264.

Heissel, Jennifer A. 2021. "Teen Fertility and Siblings' Outcomes Evidence of Family Spillovers Using Matched Samples." *Journal of Human Resources*, 56(1): 40–72.

Henriques, Alice M. 2018. "How does social security claiming respond to incentives? Considering husbands' and wives' benefits separately." *Journal of Human Resources*, 53(2): 382–413.

Ho, Cheuk Yin. 2017. "Estimating sibling spillovers in health: Evidence on symptoms." *Economics & Human Biology*, 27: 93–101.

Hofferth, Sandra L, Sarah M Flood, and Matthew Sobek. 2017. "American Time Use Survey Data Extract Builder: version 2.6 [dataset]." *College Park, MD: University of Maryland and Minneapolis, MN: IPUMS*. https://doi.org/10.18128/D060.V2.6.

Hofferth, Sandra L, Sarah M Flood, Matthew Sobek, and Daniel Backman. 2020. "American Time Use Survey Data Extract Builder: version 2.8 [dataset]." *College Park, MD: University of Maryland and Minneapolis, MN: IPUMS*. https://doi.org/10.18128/D060.V2.8.

Hong, Long, and Corina Mommaerts. 2021. "Time Aggregation in Health Insurance Deductibles." National Bureau of Economic Research.

Hsu, Joanne W, David A Matsa, and Brian T Melzer. 2018. "Replication data for: Unemployment Insurance as a Housing Market Stabilizer." American Economic Review [publisher], Interuniversity Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E116160V1.

Hurd, Michael D., and Susann Rohwedder. 2010. "Effects of the Financial Crisis and Great Recession on American Households." National Bureau of Economic Research, Working Paper 16407.

Isaacs, Katelin P. 2019. "Unemployment Insurance: Consequences of Changes in State Unemployment Compensation Laws." Congressional Research Service, Library of Congress R41859.

Isaacs, Katelin P, and Julie M Whittaker. 2014. "Emergency Unemployment Compensation (EUC08): Status of Benefits Prior to Expiration." Congressional Research Service, Library of Congress R42444.

Jäger, Simon, Benjamin Schoefer, and Josef Zweimüller. 2018. "Marginal Jobs and Job Surplus: Evidence from Separations and Unemployment Insurance." Working Paper.

Joensen, Juanna Schrøter, and Helena Skyt Nielsen. 2018. "Spillovers in education choice." *Journal of Public Economics*, 157: 158–183.

Johnston, Andrew C. 2021. "Unemployment insurance taxes and labor demand: Quasi-experimental evidence from administrative data." *American Economic Journal: Economic Policy*, 13(1): 266–93.

Johnston, Andrew C, and Alexandre Mas. 2018. "Potential Unemployment Insurance Duration and Labor Supply: The Individual and Market-level Response to a Benefit Cut." *Journal of Political Economy*, 126(6): 2480–2522.

Karbownik, Krzysztof, and Umut Özek. 2019. "Setting a good example? Examining sibling spillovers in educational achievement using a regression discontinuity design." National Bureau of Economic Research.

Katz, Lawrence F, and Bruce D Meyer. 1990. "Unemployment Insurance, Recall Expectations, and Unemployment Outcomes." *Quarterly Journal of Economics*, 105(4): 973–1002.

Kiefer, Nicholas M, Shelly J Lundberg, and George R Neumann. 1985. "How Long is a Spell of Unemployment? Illusions and Biases in the Use of CPS Data." *Journal of Business & Economic Statistics*, 3(2): 118–128.

Kilts Center for Marketing. 2009-2014. "The Nielsen Datasets." University of Chicago Booth School of Business. https://www.chicagobooth.edu/research/kilts/datasets/nielsen, accessed 2019-11-19.

Kleven, Henrik Jacobsen. 2016. "Bunching." *Annual Review of Economics*, 8: 435–464.

Kleven, Henrik J, and Mazhar Waseem. 2013. "Using notches to uncover optimization frictions and structural elasticities: Theory and evidence from Pakistan." *The Quarterly Journal of Economics*, 128(2): 669–723.

Kumcu, Aylin, and Phillip R Kaufman. 2011. "Food Spending Adjustments During Recessionary Times." *Amber Waves, USDA, Economic Research Service*, 9(3).

Lachowska, Marta, Isaac Sorkin, and Stephen Woodbury. 2021. "Firms and Unemployment Insurance Take-Up." *Working paper*.

Landais, Camille. 2015. "Assessing the welfare effects of unemployment benefits using the regression kink design." *American Economic Journal: Economic Policy*, 7(4): 243–78.

Lazear, Edward P, Kathryn L Shaw, and Christopher Stanton. 2016. "Making Do with Less: Working Harder During Recessions." *Journal of Labor Economics*, 34(S1): S333–S360.

Leung, Pauline, and Zhuan Pei. 2020. "Further Education During Unemployment."

Lindo, Jason M, Isaac D Swensen, and Glen R Waddell. 2013. "Alcohol and student performance: Estimating the effect of legal access." *Journal of health economics*, 32(1): 22–32.

Lundborg, Petter. 2006. "Having the wrong friends? Peer effects in adolescent substance use." *Journal of health economics*, 25(2): 214–233.

Luttmer, Erzo FP, and Andrew A Samwick. 2018. "The welfare cost of perceived policy uncertainty: evidence from social security." *American Economic Review*, 108(2): 275–307.

Manski, Charles F. 1993. "Identification of endogenous social effects: The reflection problem." *The review of economic studies*, 60(3): 531–542.

Manski, Charles F. 2000. "Economic Analysis of Social Interactions." *The Journal of Economic Perspectives*, 14(3): 115–136.

Marinescu, Ioana. 2017. "The General Equilibrium Impacts of Unemployment Insurance: Evidence from a Large Online Job Board." *Journal of Public Economics*, 150: 14–29.

Mas, Alexandre, and Enrico Moretti. 2009. “Peers at Work.” *American Economic Review*, 99(1): 112–145.

Mastrobuoni, Giovanni. 2011. “The role of information for retirement behavior: Evidence based on the stepwise introduction of the Social Security Statement.” *Journal of Public Economics*, 95(7-8): 913–925.

McCall, Brian P. 1995. “The impact of unemployment insurance benefit levels on recipiency.” *Journal of Business & Economic Statistics*, 13(2): 189–198.

Meyer, Bruce D. 2002. “Unemployment and Workers’ Compensation Programmes: Rationale, Design, Labour Supply and Income Support.” *Fiscal Studies*, 23(1): 1–49.

Meyer, Bruce D, and Wallace KC Mok. 2014. “A short review of recent evidence on the disincentive effects of unemployment insurance and new evidence from New York State.” *National Tax Journal*, 67(1): 219.

Milton, Ross, and Corina Mommaerts. 2020. “Income aggregation, taxation, and labor supply.” *Working paper.*

National Employment Law Project. 2015. “Unemployment Insurance 101: A Basic Glossary of Terms.” https://www.nelp.org/wp-content/uploads/2015/03/Basic-Glossary-of-Terms.pdf, accessed 2020-09-24.

Nevo, Aviv, and Arlene Wong. 2019. “The Elasticity of Substitution between Time and Market Goods: Evidence from the Great Recession.” *International Economic Review*, 60(1): 25–51.

Newhouse, Joseph P., and Rand Corporation. Insurance Experiment Group. 1993. *Free for All?: Lessons from the RAND Health Insurance Experiment.* Harvard University Press.

Nolo. n.d.. “Unemployment Benefits: What If You’re Fired?” https://www.nolo.com/legal-encyclopedia/unemployment-benefits-when-fired-32449.html, accessed 2020-09-24.

Office of Policy Development and Research. 2011. “HUD-USPS Zip to County Crosswalk File.” U.S. Department of Housing and Urban Development. https://www.huduser.gov/portal/datasets/usps_crosswalk.html#data, accessed 2020-09-24.

Office of Unemployment Insurance. 2008-2011. “Unemployment Insurance Weekly Claims.” U.S. Department of Labor. https://oui.doleta.gov/unemploy/claims_arch.asp, accessed 2020-09-24.

Pei, Zhuan. 2017. “Eligibility Recertification and Dynamic Opt-In Incentives in Income-Tested Social Programs: Evidence from Medicaid/CHIP.” *American Economic Journal: Economic Policy*, 9(1): 241–76.

Polasik, Michal, Jakub Górka, Gracjan Wilczewski, Janusz Kunkowski, Karolina Przenajkowska, and Natalia Tetkowska. 2012. “Time Efficiency of Point-of-Sale Payment Methods: Empirical Results for Cash, Cards and Mobile Payments.” *International Conference on Enterprise Information Systems*, 306–320.

Prell, Mark A. 2008. “Income volatility and certification duration for WIC children.” *Income Volatility and food Assistance in the united States*, 259–94.

Rebollo-Sanz, Yolanda. 2012. "Unemployment Insurance and Job Turnover in Spain." *Labour Economics*, 19(3): 403–426.

Rothstein, Jesse. 2011. "Unemployment Insurance and Job Search in the Great Recession." *Brookings Papers on Economic Activity*, 2011(2): 143–213.

Rothstein, Jesse, and Robert G Valletta. 2017. "Scraping by: Income and program participation after the loss of extended unemployment benefits." *Journal of Policy Analysis and Management*, 36(4): 880–908.

Sass, Steven A, Wei Sun, and Anthony Webb. 2013. "Social security claiming decision of married men and widow poverty." *Economics Letters*, 119(1): 20–23.

Schmieder, Johannes F, and Till Von Wachter. 2016. "The Effects of Unemployment Insurance Benefits: New Evidence and Interpretation." *Annual Review of Economics*, 8(1): 547–581.

Shapiro, Carl, and Joseph E Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74(3): 433–444.

Shore-Sheppard, Lara D. 2014. "Income dynamics and the affordable care act." *Health services research*, 49(S2): 2041–2061.

SIEPR-Giannini Data Center. n.d.. https://are.berkeley.edu/SGDC/, accessed 2020-09-24.

Taylor, Rebecca L. C. 2020. "A Mixed Bag: The Hidden Time Costs of Regulating Consumer Behavior." *Journal of the Association of Environmental and Resource Economists*, 7(2): 345–378.

Trim, Ryan S, Eileen Leuthe, and Laurie Chassin. 2006. "Sibling influence on alcohol use in a young adult, high-risk sample." *Journal of studies on alcohol*, 67(3): 391–398.

U.S. Census Bureau. 2017. "Retail Trade: Summary Statistics for the U.S., States, and Selected Geographies: 2017." https://data.census.gov/cedsci/, accessed 2020-09-24.

Valletta, Robert G. 2014. "Recent Extensions of US Unemployment Benefits: Search Responses in Alternative Labor Market States." *IZA Journal of Labor Policy*, 3(1): 18.

Van Der Vorst, Haske, Rutger CME Engels, Wim Meeus, Maja Deković, and Jan Van Leeuwe. 2007. "Similarities and bi-directional influences regarding alcohol consumption in adolescent sibling pairs." *Addictive behaviors*, 32(9): 1814–1825.

Vickrey, William. 1939. "Averaging of income for income-tax purposes." *Journal of Political Economy*, 47(3): 379–397.

Whiteman, Shawn D, Alexander C Jensen, and Jennifer L Maggs. 2013. "Similarities in adolescent siblings' substance use: Testing competing pathways of influence." *Journal of studies on alcohol and drugs*, 74(1): 104–113.

Whittaker, Julie M, and Katelin P Isaacs. 2013. "Extending Unemployment Compensation Benefits During Recessions." Congressional Research Service, Library of Congress RL34340.

Whittaker, Julie M, and Katelin P Isaacs. 2014. "Unemployment Insurance: Legislative Issues in the 113th Congress." Congressional Research Service, Library of Congress R42936.

Wikle, Jocelyn S, and Alex Hoagland. 2020. "Adolescent interactions with family and emotions during interactions: Variation by family structure." *Journal of Family Psychology*, 34(5): 544.

Winter-Ebmer, Rudolf. 2003. "Benefit Duration and Unemployment Entry: A Quasi-experiment in Austria." *European Economic Review*, 47(2): 259–273.

Yörük, Barış K, and Ceren Ertan Yörük. 2011. "The impact of minimum legal drinking age laws on alcohol consumption, smoking, and marijuana use: Evidence from a regression discontinuity design using exact date of birth." *Journal of health economics*, 30(4): 740–752.

Yörük, Barış K, and Ceren Ertan Yörük. 2013. "The impact of minimum legal drinking age laws on alcohol consumption, smoking, and marijuana use revisited." *Journal of health economics*, 32(2): 477–479.

Yörük, Ceren Ertan, and Barış K Yörük. 2012. "The impact of drinking on psychological well-being: Evidence from minimum drinking age laws in the United States." *Social Science & Medicine*, 75(10): 1844–1854.

Yörük, Ceren Ertan, and Barış Yörük. 2015. "Alcohol consumption and risky sexual behavior among young adults: evidence from minimum legal drinking age laws." *Journal of Population Economics*, 28(1).