Finance and Economics Discussion Series

Federal Reserve Board, Washington, D.C.
ISSN 1936-2854 (Print)
ISSN 2767-3898 (Online)

Social Security and High-Frequency Labor Supply: Evidence from Uber Drivers

Timothy K. M. Beatty, Joakim A. Weill

2024-079

Please cite this paper as:
Beatty, Timothy K. M., and Joakim A. Weill (2024). "Social Security and High-Frequency Labor Supply: Evidence from Uber Drivers," Finance and Economics Discussion Series 2024-079. Washington: Board of Governors of the Federal Reserve System, https://doi.org/10.17016/FEDS.2024.079.

NOTE: Staff working papers in the Finance and Economics Discussion Series (FEDS) are preliminary materials circulated to stimulate discussion and critical comment. The analysis and conclusions set forth are those of the authors and do not indicate concurrence by other members of the research staff or the Board of Governors. References in publications to the Finance and Economics Discussion Series (other than acknowledgement) should be cleared with the author(s) to protect the tentative character of these papers.

SOCIAL SECURITY AND HIGH-FREQUENCY LABOR SUPPLY: EVIDENCE FROM UBER DRIVERS

Timothy K. M. Beatty* Joakim A. Weill†‡

September 2024

Abstract

We estimate the impact of anticipated transfers on labor supply using confidential driver-level data from Uber. Leveraging the staggered timing of Social Security retirement benefits within each month and a novel identification strategy, we find that the labor supply of older drivers declines by 2%, on average, during the week of benefit receipt—a precisely estimated but economically small effect. Individual-level analyses reveal that the average effect obscures heterogeneous micro-behavior: while the majority of drivers do not meaningfully adjust labor supply in response to social security benefits, a small group reduces labor supply by more than 40%. The results suggest that departures from standard models of labor supply can be substantial, but only for a small number of individuals.

*Agricultural and Resource Economics, University of California, Davis, USA.

†Federal Reserve Board of Governors, Washington, DC, USA. Email: joakim.a.weill@frb.gov

‡We thank Uber for generously providing data access, Libby Mishkin for invaluable insights and support with this project, Mariya Shappo for constructive feedback, and Felipe Vial and Tianxia Zhou for expert technical assistance. We also thank Marianne Bitler, Eirik Brandsaas, David Cho, Edmund Crawley, Thomas Crossley, John Driscoll, Hank Farber, Peter Ganong, Joaquin Garcia-Cabo Herrero, Hilary Hoynes, Paul Lengermann, Jim Poterba, Brendan Price and Takuya Ura for thoughtful suggestions. We thank numerous seminar participants for their helpful comments. Uber data were obtained by Joakim Weill prior to his employment at the Federal Reserve Board, while he was a PhD candidate at UC Davis. The views expressed in this paper are solely the responsibility of the authors and should not be interpreted as reflecting the opinions of the Federal Reserve Board or of any other person associated with the Federal Reserve System.

1 Introduction

Americans rely on Social Security to fund the majority of their retirement expenses, but for many, the benefits are not enough to make ends meet (Porell and Bond, 2020; Dushi and Trenkamp, 2021). A small but growing number of retirees are turning to nontraditional, or "gig," work for income replacement or to feel productive (Ramnath, Shoven, and Slavov, 2021; Wettstein and Rutledge, 2023). But we know little about how Social Security and gig work interact. Does the receipt of Social Security retirement benefits alter labor-supply decisions?

This paper examines whether and how Uber drivers adjust daily labor-supply decisions in response to the timing of monthly Social Security retirement-benefit receipt. Gig work offers an ideal setting for tackling this question because there are no transaction costs associated with searching for work and workers can costlessly vary hours worked. Most older Americans are eligible for Medicare, which disentangles employment decisions from access to health insurance (Cohen, Cha, Terlizzi, and Martinez, 2021).

We use a well-established research design that leverages the plausibly exogenous assignment of Social Security disbursement dates based on the day of birth (see, among others, Evans and Moore, 2011; Leary and Wang, 2014; Baugh, Leary, and Wang, 2017; Gross, Layton, and Prinz, 2022; Akesaka, Eibich, Hanaoka, and Shigeoka, 2023). Social Security benefits are perfectly anticipated, economically important, and do not fluctuate in value during the course of a calendar year. While we do not observe Social Security transfers themselves, we observe each driver's age, which is the main determinant of Social Security eligibility, and their day of birth, which determines the timing of benefit receipt. Building on recent econometric work focused on estimating average treatment effects when treatment is staggered in time (de Chaisemartin and D'Haultfœuille, 2020; Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; Sun and Abraham, 2021), we develop a novel estimation strategy that allows for both ex ante anticipation effects and ex post receipt effects by cohort. We construct explicit comparisons between drivers who are within a week of treatment and drivers who are at least two weeks away from treatment: identification rests on the assumption that the disbursement of Social Security benefits distorts labor supply at most two weeks per month. We scrutinize this assumption using recent developments in the difference-in-differences literature (Butts, 2021; Gardner, Thakral, Tô, and Yap, 2024).

We find that, on average, Uber drivers older than 62 years old reduce labor supplied by between 2% and 5% during the week of benefit receipt. Change occurs at the extensive margin, and drivers are less likely to work on the days both immediately before and after the receipt of benefits. We find precise estimates of zero effect for a placebo group of younger drivers (58 to 61 years old) who are not yet eligible for Social Security. Drivers who use the Instant Pay feature on the app more frequently, which allows for near-immediate transfer of earnings into their bank account, are more likely to reduce their labor supply in response to the timing of benefit receipt.

At first glance, these findings might suggest that a key prediction of the standard model of labor supply — that workers should not respond to an anticipated lump-sum transfer — does not hold, even in a setting where workers can costlessly adjust when and how much to work. But estimated effect sizes are small and not economically important. Small estimated average effects can be explained in two ways: either a large group of drivers adjusts their labor supply around the date of benefit receipt by a little, or a small group adjusts their labor supply by a lot. Our results point toward the latter: estimating individual-level treatment effects to assess unobserved heterogeneity, we find that while most drivers do not respond in a meaningful way to the receipt of Social Security benefits, a small group of drivers (about 1%) sharply curtails hours worked around the date of benefit receipt, forgoing roughly 20% of average earnings. In short, individual-level results show that the standard model offers a reasonable approximation for most drivers but is rejected for a small group of drivers. For these drivers, the large labor supply distortions cannot be explained by liquidity constraints, as these would cause labor supply to *increase* on the days preceding the receipt of benefits. Instead, our results point toward nonstandard preferences such as inconsistent time preferences or target-earning behavior.

This research question fits into a large literature on whether the receipt of anticipated lump-sum transfers distorts consumption and labor supply decisions. The life-cycle model of labor supply implies that labor supply should not respond to anticipated lump-sum transfers at the time of receipt. Despite strong theoretical predictions that lump-sum transfers will not distort behavior, empirical evidence often finds they do. Studies of transfer payments from safety net programs often document changes in labor supply. Studies of earned income tax credit (EITC) receipt document a decrease in job-search intensity among some groups (LaLumia, 2013) and an economically meaningful reduction in labor supply (Yang, 2018) in the months following the likely receipt of an EITC refund. Deshpande (2016) finds the loss of Supplemental Security Income induces an increase in labor earnings, and Gelber, Moore, and Strand (2017) find increases in Social Security Disability Insurance benefits lead to reductions in labor earnings. Studies of lottery winners (Imbens, Rubin, and Sacerdote, 2001; Cesarini, Lindqvist, Notowidigdo, and Östling, 2017) document reductions in labor supply, though estimated magnitudes vary. On the other hand, other lump-sum transfers do not appear to affect behavior such as transfer payments from casino profits (Akee, Copeland, Keeler, Angold, and Costello, 2010) or the receipt of a cash dividend from the Alaska Permanent Fund (Jones and Marinescu, 2022). Our results here may have lessons for these earlier studies that couldn't distinguish between average effects for a large population vs large effects for a small population.

The effects of lump-sum transfers have also been studied on consumption behavior. The permanent-income hypothesis implies that expenditure should not respond to the receipt of a predictable transfer. However, most empirical tests of this prediction find that spending increases following the receipt of an anticipated transfer such as a tax refund (Souleles, 1999), Social Security payment (Stephens Jr., 2003; Stephens Jr. and Unayama, 2011; Gross et al., 2022), or other government transfer (Shapiro, 2005; Beatty and Tuttle, 2015; Hastings and Shapiro, 2018).

We add to a recent literature examining the unintended consequences of the monthly benefit cycles of multiple social-assistance programs. Prior work documents a within-month spending cycle associated with food stamps / SNAP benefit-receipt (Shapiro, 2005; Smith, Berning, Yang, Colson, and Dorfman, 2016; Beatty, Bitler, Cheng, and van der Werf, 2019; Zaki and Todd, 2021; Goldin, Homonoff, and Meckel, 2022), with knock-on effects for food insecurity (Gregory and Todd, 2021), crime (Carr and Packham, 2019), domestic violence (Carr and Packham, 2021), emergency room visits (Cotti, Gordanier, and Ozturk, 2020), and test scores (Bond, Carr, Packham, and Smith, 2022). Relative to the transfers studied in earlier work, Social Security benefits are typically larger and the timing of benefit disbursement is more precisely identified.

This paper also contributes to an emerging literature on the rise of alternative work arrangements and the gig economy (Cook, Diamond, and Oyer, 2019). Gig work is common: a 2021 survey finds that 16% of Americans have earned money through online gig work at some point and 4% are currently doing so, with transportation being the industry of fastest growth (Abraham, Haltiwanger, Sandusky, and Spletzer, 2019; Anderson, McClain, Faverio, and Gelles-Watnick, 2021). One of the most important benefits of gig work is the ability to smooth income by adjusting labor supply. In a survey of Uber drivers, Hall and Krueger (2018) find that "drivers often cite the desire to smooth fluctuations in their income as one of their reasons for partnering with Uber." Gig work offers flexible work arrangements to older workers who wish to supplement retirement savings (Ameriks, Briggs, Caplin, Lee, Shapiro, and Tonetti, 2020). Given the increasing popularity of alternative work arrangements, understanding the interplay between monthly benefit-issuance cycles and the gig economy is important (Katz and

Krueger, 2019).

We focus on older adults who are likely eligible to receive Social Security. Across much of the developed world, the number of adults over 60 will double by 2050 (United Nations, 2017). The share of adults over 65 in the labor force has doubled over the last 35 years(Fry, Richard and Braga, Dana, 2023). Given the evolving landscape of retirement and pension systems, understanding how stable transfers affect labor supply among older adults is of macroeconomic and policy importance.

Finally, our paper is related to the large literature looking at the labor supply decisions of New York City taxi drivers (Camerer, Babcock, Loewenstein, and Thaler, 1997; Farber, 2005, 2008; Crawford and Meng, 2011; Farber, 2015; Thakral and Tô, 2021) and Fehr and Goette (2007)'s seminal field experiment with bicycle messengers in Zurich. This literature asks how workers vary their labor supply decisions in response to a change in the effective wage rate. In contrast, our paper looks at how labor supply is affected by the anticipation and receipt of a predictable monthly lump-sum transfer.

The rest of the paper is organized as follows. Section 2 provides additional background on Social Security eligibility and treatment timing. Section 3 presents the data and shows graphical evidence of an effect of Social Security benefits on labor supply. Section 4 presents our research design and event study estimates, while individual-level analyses are presented in Section 5. Section 6 discusses policy implications.

2 Background

Social Security retirement benefits, or Old-Age, Survivors, and Disability Insurance (OASDI), are paid monthly. These payments are predictable and do not vary month to month. Benefits are indexed for inflation (CPI-W) and updated every January. Social Security benefits are ubiquitous and economically important: they cover more than 93% of workers and provide the majority of retirement income to individuals

65 and older (Poterba, 2014; Dushi, Iams, and Trenkamp, 2017). During our sample period, 2018–2019, the average monthly benefit was $1,479, paid to more than 60 million individuals (Social Security Administration, 2020).

Unlike the EITC or tax refunds more broadly, the timing of Social Security benefit disbursement is perfectly anticipated by participants and known to researchers. Beginning in 1997, benefits for new claimants have been paid based on day of birth. Specifically, individuals born between the 1st and 10th of the month receive their benefits on the second Wednesday of each month, individuals born between the 11th and 20th of the month receive their benefits on the third Wednesday, and individuals born on or after the 21st of the month receive their benefits on the fourth Wednesday.[1] We refer to these three groups as *day-of-birth cohorts*. Payments are made via direct deposit to individuals. While timing is dictated by day-of-birth, some banks and credit unions make benefits available to individuals a day or two early (Gross et al., 2022).[2]

Individuals can begin collecting Social Security benefits at age 62, albeit at a reduced rate relative to the full retirement age (FRA) of 66. Most drivers in our sample would experience a roughly 25% benefit reduction were they to claim benefits before FRA.[3] There are benefits to deferring Social Security beyond FRA: every year of delay leads to an 8% increase in benefits, up to a maximum of 132% of FRA benefits at age 70. Despite incentives to defer benefits, recent work explores why a large share of Americans claim Social Security as soon as possible (Shu and Payne, 2023). There is some evidence that workers may use gig work to postpone receipt of benefits in order to access higher benefit levels (Jackson, 2022).

[1] There are a few exceptions to this rule. If the payment date is a federal holiday, disbursement occurs on the Tuesday. Individuals who claimed Social Security before 1997 receive their benefits on the third day of the month. Finally, individuals who participate in both Supplemental Security Income and Social Security retirement benefits are also paid on the third of the month.

[2] For example, Wells Fargo advertises that eligible government benefits, such as Social Security, may be made available up to two days early; see `https://www.wellsfargo.com/checking/early-pay-day/`

[3] For the drivers in our sample born between 1955 and 1959, the FRA increased from 66 to 67 in two-month increments.

Social Security recipients below FRA face an income-clawback provision: for every dollar of income net of expenses earned above a threshold ($17,040 in 2018 and $17,640 in 2019), benefits are withheld at 50% of the excess.[4] In the year a recipient reaches FRA, the limits are higher ($45,360 in 2018 and $46,920 in 2019), and benefits are withheld at 33% of the excess. Benefits withheld are subsequently applied as a retirement credit once the worker reaches FRA. There are no earning limits for workers at or above FRA. Individuals declare their anticipated annual income at the beginning of the year, and all Social Security benefits are withheld until the appropriate share of the excess earnings is reached. The practical implication is that a driver who expects to earn above the limit could receive zero (or reduced) benefits during the first months of the year. Finally, self-employed workers below FRA cannot work more than 45 hours a week and still collect Social Security. Anecdotal evidence from driver message boards suggests self-employment income limits are rarely binding. Drivers deduct depreciation or rental costs, insurance, fuel, and other expenses to reduce taxable income below thresholds.

As with an earlier literature focused on the impacts of EITC receipt (for example, McGranahan and Schanzenbach, 2013; Goodman-Bacon and McGranahan, 2008; LaLumia, 2013; Yang, 2018; Baugh, Ben-David, Park, and Parker, 2021; Aladangady, Aron-Dine, Cashin, Dunn, Feiveson, Lengermann, Richard, and Sahm, 2023), we do not observe whether a given driver receives benefits. Rather, we observe eligibility and treatment timing; younger drivers may be deferring participation, or drivers may be recent immigrants who are not yet eligible for benefits. The implication is that we recover intent-to-treat estimates. However, survey and administrative data tell us that most older Americans collect Social Security benefits. According to the American Community Survey (ACS), most older Taxi Drivers and Chauffeurs (occupation code

[4]For drivers in our sample, 84 % and 80 % of drivers below FRA earned less than these thresholds in 2018 and 2019, respectively. Note that this is likely a lower bound on the share of drivers unaffected by these thresholds as as we observe gross earnings and taxable income is likely lower because drivers can deduct costs such as depreciation, fuel, and insurance.

9140) collect Social Security benefits, and the share quickly increases with age: 25% of drivers 62 years old, 60% of drivers 65 years old, and 80% of drivers 68 and older receive Social Security retirement income (see Figure A18 in the appendix). Figure A19 shows that Social Security benefits make up an increasingly large share of driver income. This lines up with reports from the Social Security Administration, suggesting that most drivers in our sample are treated. In 2018, 27.4% of new male OASDI and 31% of female claimants were 62, 49.7% of male and 54.6% of female claimants were less than FRA, and only 14.6% of men and 14.9% of women deferred claiming past FRA (Social Security Administration, 2020).

3 Data

3.1 Sources and Summary Statistics

We use proprietary data from Uber. The data consist of both a main (Retirement Age) and a placebo (Working Age) sample observed in 2018 and 2019. The Retirement Age sample is a random sample of 49,515 Uber drivers born in 1956 or earlier. Drivers in this sample were at least 62 years old in January 2018 and potentially eligible to collect Social Security retirement benefits.[5] The Working Age sample is a random sample of 41,446 younger drivers, born between 1958 and 1961, and therefore not eligible for Social Security retirement benefits during our sample period. This sample is used to conduct placebo tests. For every driver, we drop quarters when the individual is not online. This results in an unbalanced panel of 90,961 drivers, with a maximum of 730 unique daily observations per driver. Consistent with earlier work on the gig economy, the number of drivers in each age cell is decreasing in age (Figure A10). Drivers are distributed across all US states.

[5] We exclude the small number of drivers who turned 65 prior to the implementation of the staggered payment schedule based on day of birth in 1997 (and would be at least 86 years old in 2018), as these drivers are likely to receive their benefits on the third day of the month.

Each driver falls in one of three day-of-birth cohorts, where cohort membership is determined by day of birth (1–10, 11–20, or 21–31). These cohorts determine the timing of potential receipt of Social Security retirement benefits for Retirement Age drivers and serve as our primary source of identifying variation. For example, a driver born on March 7 belongs to the first birthday cohort and potentially receives OASDI benefits on the second Wednesday of each month. As detailed above, we do not observe whether individual drivers currently receive Social Security, so results should be interpreted as intent-to-treat estimates. More broadly, we do not observe whether individuals are ineligible for Social Security, as might be the case for a recent immigrant who lacks the requisite work history.

We observe each driver's daily usage of the Uber app and driving behavior over the 24 months between January 2018 and December 2019. Table 1 presents summary statistics for our Retirement Age and Working Age samples. On average, drivers in both samples are online about 15 days a month, spend an average of 2.3 hours online, and drive 1.3 hours per day. The difference between time online and time driving is largely due to time between trips. Relative to drivers in our Working Age sample, drivers in the Retirement Age sample are more likely to use the app during working hours and during the workweek—consistent with the fact that drivers in the main sample are older and thus more likely to be retired from full-time employment.

We focus on the probability of being online ("probability online") on any given day as our measure of labor supply at the extensive margin and the time spent on the app ("hours online") as our preferred measure of the intensive margin. While informative, the daily number of hours spent driving is an equilibrium outcome that also depends on the demand for trips. All results are robust to alternative empirical choices, such as using time spent driving ("hours active") as the measure of labor supply. Consistent with Cook et al. (2019), we see that the older drivers (in the Retirement Age sample) are less efficient—that is, they work slightly more but earn slightly less—than the younger

Table 1: Summary table of Uber data

Variable	Observations	Working Age	Retirement Age
Female	90,961	0.22 (0.41)	0.15 (0.36)
Share online 10am-4pm	90,961	0.34 (0.23)	0.42 (0.25)
Share online workweek	90,961	0.71 (0.19)	0.74 (0.18)
Earnings (wk)	5,745,710	248.06 (320.91)	244.52 (295.36)
Instant pay usage (wk)	5,745,710	0.34 (0.47)	0.3 (0.46)
Tips (wk)	5,745,710	17.31 (26.05)	18.01 (25.05)
Hours active (day)	39,123,035	1.32 (2.1)	1.36 (2.04)
Hours online (day)	39,123,035	2.18 (3.23)	2.27 (3.15)
Probability active (0/1)	39,123,035	0.41 (0.49)	0.44 (0.5)
Probability online (0/1)	39,123,035	0.45 (0.5)	0.48 (0.5)
Rides, trips (day)	39,123,035	3.62 (6.01)	3.65 (5.7)

The Working Age sample includes drivers born between 1958 and 1961, while the Retirement Age sample includes drivers born before 1956. Conditional on these age restrictions, both samples are randomly drawn from the population of Uber drivers in the United States.

drivers (in the Working Age sample).

Earnings are calculated weekly and automatically deposited in a driver's account on Wednesdays. Drivers have the option to request the immediate transfer of their available balance by using the Instant Pay option. Depending on the driver's financial institution, this transfer can take up to three days to be completed. Earnings and Instant Pay usage are recorded weekly by Uber. Drivers use the Instant Pay option a third of the time, on average, and earn around $250 per week from Uber.

We observe roughly the same number of drivers in each sample and in each day-of-birth cohort within each sample. The labor supply behaviors of drivers in the Retirement Age and Working Age samples are significantly different along all characteristics: Working Age drivers spend less time driving and online than Retirement Age drivers, work slightly less from 10 a.m. to 4 p.m. and during the workweek, and earn slightly more (Table 1). Differences also exist between day-of-birth cohorts within each driver

Table 2: Balance table between drivers' day-of-birth cohorts

Variable	Observations	Cohort 1	Cohort 2	Cohort 3
A. Retirement Age:				
Year of birth	49,515	1,950.86 (4.49)	1,950.97 (4.46)	1,950.87 (4.51)
Female	49,515	0.15 (0.36)	0.15 (0.36)	0.15 (0.36)
Share online 10am-4pm	49,515	0.42 (0.25)	0.41 (0.25)	0.41 (0.25)
Share online workweek	49,515	0.74 (0.18)	0.74 (0.18)	0.74 (0.18)
Earnings (wk)	3,136,033	246.92 (297.41)	244.59 (296.69)	242.11 (292.05)
Instant pay usage (wk)	3,136,033	0.3 (0.46)	0.3 (0.46)	0.3 (0.46)
Tips (wk)	3,136,033	17.97 (25.07)	18.18 (25.48)	17.9 (24.61)
Hours active (day)	21,361,837	1.37 (2.04)	1.35 (2.04)	1.35 (2.03)
Hours online (day)	21,361,837	2.29 (3.16)	2.27 (3.15)	2.25 (3.14)
Probability active (0/1)	21,361,837	0.44 (0.5)	0.44 (0.5)	0.44 (0.5)
Probability online (0/1)	21,361,837	0.48 (0.5)	0.48 (0.5)	0.47 (0.5)
Rides, trips (day)	21,361,837	3.68 (5.71)	3.64 (5.72)	3.62 (5.67)
B. Working Age:				
Year of birth	41,446	1,959.6 (1.11)	1,959.61 (1.12)	1,959.59 (1.12)
Female	41,446	0.21 (0.41)	0.22 (0.41)	0.22 (0.42)
Share online 10am-4pm	41,446	0.33 (0.23)	0.33 (0.23)	0.34 (0.23)
Share online workweek	41,446	0.71 (0.19)	0.71 (0.19)	0.71 (0.19)
Earnings (wk)	2,609,677	253.92 (328.07)	248.45 (320.04)	241.72 (314.24)
Instant pay usage (wk)	2,609,677	0.34 (0.47)	0.33 (0.47)	0.34 (0.47)
Tips (wk)	2,609,677	17.56 (26.74)	17.36 (26.23)	17.02 (25.13)
Hours active (day)	17,761,198	1.34 (2.13)	1.32 (2.1)	1.29 (2.08)
Hours online (day)	17,761,198	2.22 (3.26)	2.18 (3.23)	2.14 (3.2)
Probability active (0/1)	17,761,198	0.42 (0.49)	0.41 (0.49)	0.41 (0.49)
Probability online (0/1)	17,761,198	0.46 (0.5)	0.45 (0.5)	0.45 (0.5)
Rides, trips (day)	17,761,198	3.69 (6.08)	3.63 (6.03)	3.54 (5.93)

The balance table presents means and standard deviations by day-of-birth cohorts, separately for the Working Age and Retirement Age samples. Cohort 1 consists of drivers born between the 1st and the 10th day of the month; Cohort 2, between the 11th and the 20th; and Cohort 3, between the 21st and the 31st.

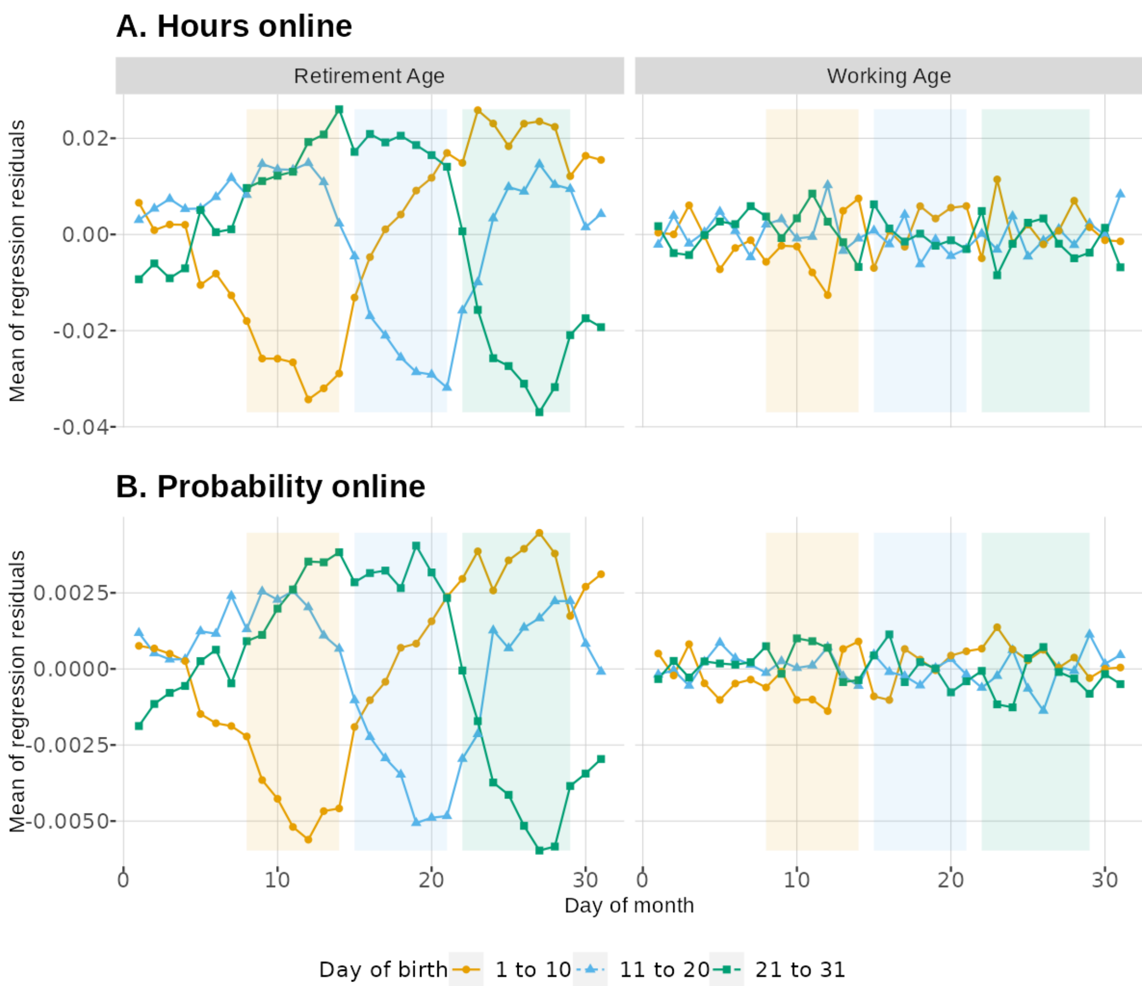
sample: Table 2 highlights that for both samples, drivers in the first cohort spend more time online and earn more money than drivers in the second cohort, who themselves work slightly more than drivers in the third cohort. However, behavioral differences between cohorts within each driver sample are substantially smaller than differences between samples (Figures A21 and A22 in the appendix present the distributions of these variables, further highlighting differences between driver samples). These stylized facts motivate us to compare day-of-birth cohorts within each sample to each other rather than comparing Retirement Age to Working Age drivers.

3.2 Graphical Analysis

We begin with a simple graphical analysis of driver behavior on the days around benefit receipt. Figure 1 presents deviations from the day mean and cohort mean of time spent online (Panel A) and probability online (Panel B) for our Retirement Age sample (left) and Working Age sample (right). As the exact day of benefit receipt varies between months, rectangles indicate the date range when benefits are paid during the months for each cohort. Looking at the Retirement Age sample (left panel), across both measures of labor supply and for both samples, a clear pattern emerges: drivers are less likely to be online and spend less hours online on the days around benefit receipt for their cohort but not during the disbursement windows for other cohorts. No similar pattern is observed for our Working Age placebo group (right panel).

This variation is also clear in event time, in which the event ($t = 0$) is defined as the day of Social Security receipt. Figure 2 shows that for the main sample, the probability of being online and hours online decline one or two days *prior* to the receipt of benefits, are lowest on the day of benefit receipt, and slowly increase on the days after benefits are paid. The effect of benefit receipt appears to be asymmetric around the event date. The probability of being online is below the date mean between two and three days ahead of benefit receipt, suggesting that anticipation effects are larger

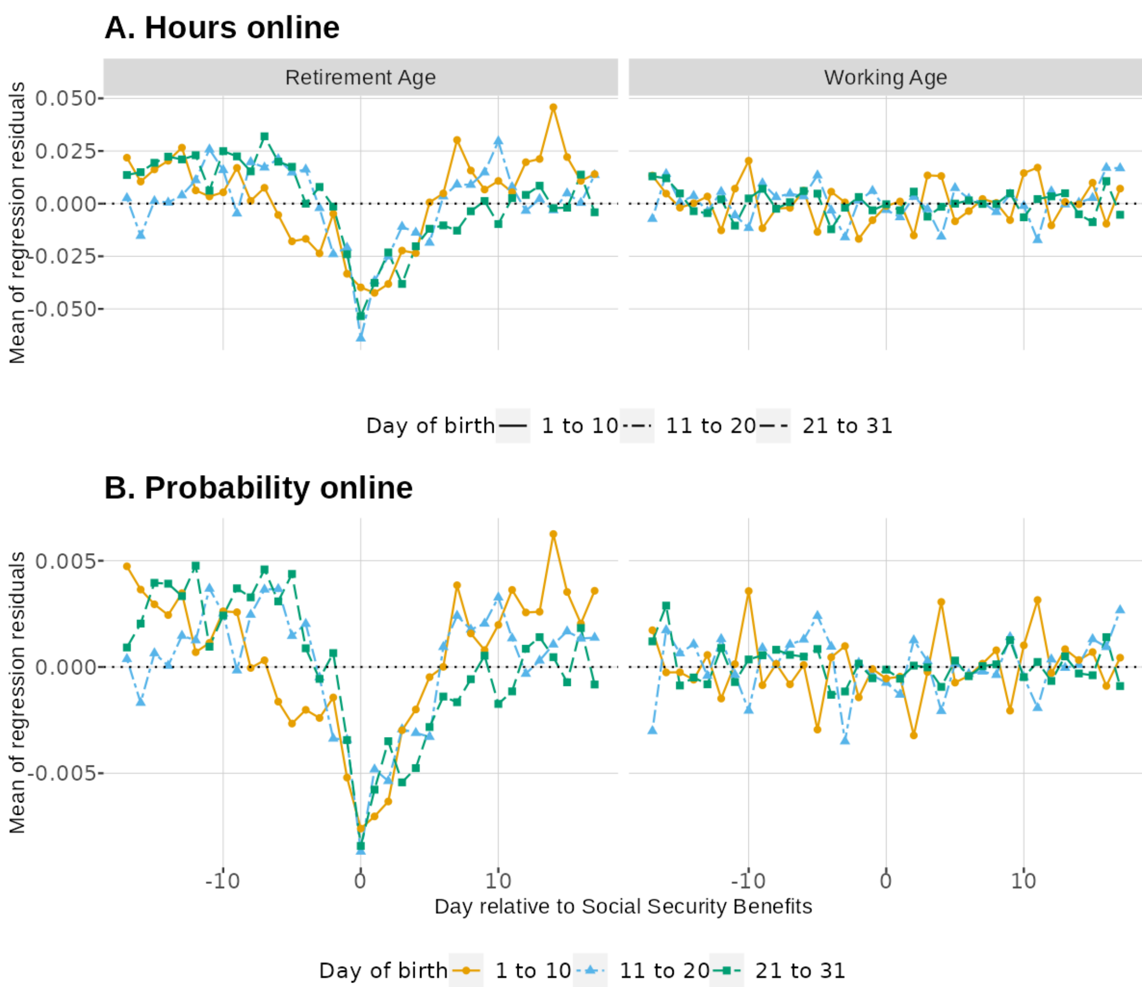
Figure 1: Labor supply residualized by day of month



The figure presents the mean residuals of the regressions $Y_{it} = D_t + B_i + \epsilon_{it}$, where Y_{it} is a measure of labor supply on day t for driver i (number of hours spent online in Panel A, and probability of being online in Panel B), D_t are day-of-month fixed effects, and B_i are day-of-birth-cohort fixed effects (1 to 10, 11 to 20, or 21 to 31), with residuals computed separately for the Working Age and Retirement Age samples. The colored rectangles outline the time windows when each day-of-birth cohort is expecting to receive their benefits (these are windows rather than points, as Wednesdays fall on different dates each month).

than can be explained by the fact that banks make benefits available a day early. After treatment, the effect exhibits sharp dynamics, with drivers returning to baseline about five or six days after benefit receipt. As above, Working Age drivers do not respond, either before or after, to the timing of Social Security disbursement. Comparing the two standardized measures of labor supply suggests that most of the variation is driven by the extensive margin (decision to work or not work on a given day) rather than by the intensive margin (time spent online).

Figure 2: Labor supply residualized by date



The figure presents the mean residuals of the regressions $Y_{it} = Date_t + B_i + \epsilon_{it}$, where Y_{it} is a measure of labor supply on day t for driver i (number of hours spent online in Panel A, and probability being online in Panel B), $Date_t$ are date fixed effects, and B_i are day-of-birth-cohort fixed effects (1 to 10, 11 to 20, or 21 to 31), with residuals computed separately for the Working Age and Retirement Age samples and plotted relative to the day of receipt of benefits.

In sum, the simple graphical analysis above suggests that labor supply responds, at least a little, to the receipt of a perfectly predictable lump sum transfer. However, this analysis does not account for potential confounders at the driver level, nor does it control for demand shocks that might systematically occur on Wednesdays. The next section formalizes the analysis to tackle these issues and conduct hypothesis testing.

4 Research Design

4.1 Paired-Event-Study Framework

Our setting involves a large number of units (90,961 drivers) potentially receiving treatment (Social Security benefits) at three different times throughout the month—the second, third, or fourth Wednesday—depending on day of birth. While the timing of treatment is clearly exogenous—that is, people do not choose their day of birth—identifying the causal effect of Social Security benefits presents a number of challenges.

First, treatment in our context is both *staggered* within a month and *repeated* every month. The result is that there is no obvious choice of what constitutes a pre- or post-treatment period. An emerging and active econometrics literature tackles the issue of recovering average-treatment-effect estimates under staggered treatment (Goodman-Bacon, 2021; Sun and Abraham, 2021; Callaway and Sant'Anna, 2021). However, this literature tends to focus on "absorbing treatments"—that is, treatments in which treated units do not change their status once they become treated.[6] Second, there is no natural control group—that is, we do not observe a group of otherwise identical Retirement Age drivers who do not participate in Social Security. While at first blush our placebo (Working Age) drivers might seem like a natural control group, Working Age drivers exhibit sharply different labor supply behavior—driving many more evening and week-

[6] de Chaisemartin and D'Haultfœuille (2020) allow for the possibility of units' switching from treated to control but do not explicitly tackle the case of repeated switchers.

end hours than Retirement Age drivers—making them a poor control group choice to identify the effects of labor supply at the daily scale. Finally, because Social Security benefits are perfectly anticipated, it is important to explicitly estimate pre-treatment effects to allow for anticipation effects.

To overcome these challenges, we impose a restriction on treatment-effect dynamics. In order to construct control groups, we assume the timing of Social Security benefits distorts labor supply, at most, 15 days per month, each and every month. This is consistent with the patterns depicted in Figure 1. It is also in line with the earlier literature on the impacts of Social Security receipt on spending, which typically finds spending returns to baseline levels within a week of benefit receipt (Stephens Jr., 2003; Gross et al., 2022).[7] Our approach differs from earlier work using the Social Security calendar in that we explicitly allow for anticipation effects, which may be more relevant in the case of labor supply than spending.

More formally, let $ATT(g,t,m)$ denote the average treatment effect on the treated for group g t days away from treatment for occurrence m, where, as before, groups are defined based on the day of birth. In keeping with the definition presented in Callaway and Sant'Anna (2021), this effect can be written in terms of potential outcomes:

$$ATT(g,t,m) = E[Y_t(g) - Y_t(0) \,|\, G_g = 1, M_m = 1] \tag{1}$$

Here, $Y_t(g)$ is the potential outcome of units in group g t days away from treatment, and $Y_t(0)$ is the potential outcome of these units had they not been treated. G_g and M_m are indicator functions denoting the group membership of drivers and the numbered occurrence of their treatment (January 2018 is the first treatment in our sample, February 2018 is the second, and so on). To identify these effects using observed outcomes, we assume no anticipation effects more than seven days prior to treatment and

[7] In their regression specification, Gross et al. (2022) specify a five-day post-treatment window.

no treatment effects more than seven days after treatment:

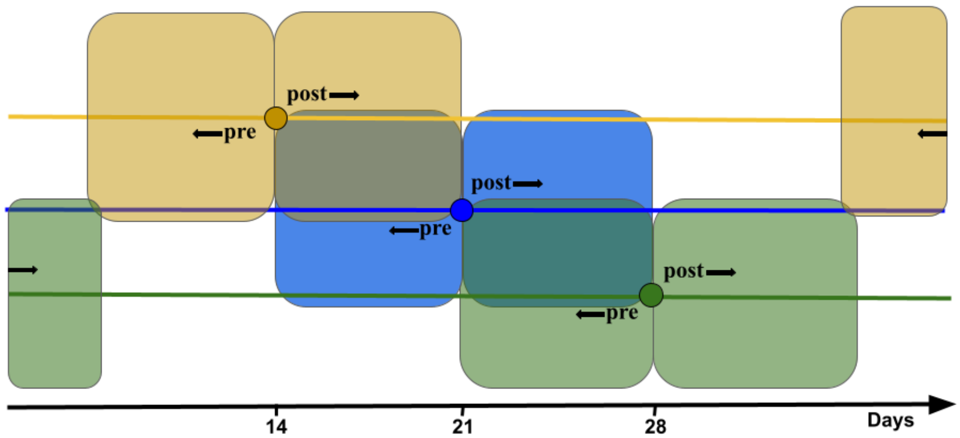
$$ATT(g, t, m) = 0 \ \forall \{g, m\} \ \& \ \forall t \notin [-7, 7] \tag{2}$$

This offers a transparent means to address the problem of estimating average treatment effects for the receipt of Social Security benefits: compare each group of drivers to other drivers who are more than 7 days away from receiving their benefits. This leaves one remaining issue if we allow for both anticipation and treatment effects: for each cohort of drivers, the other two cohorts cannot simultaneously be valid control units for both pre- and post-treatment periods. To see this, recall that Cohorts 2 and 3 receive their benefits exactly 7 and 14 days after Cohort 1, respectively. If we were to use Cohort 2 as a control for Cohort 1, estimates of post-treatment effects for Cohort 1 would be confounded with any pre-treatment (anticipation) effects for Cohort 2. Alternatively, if we were to choose Cohort 3 as a control, estimates of pre-treatment effects for Cohort 1 would be confounded by any post-treatment effects for Cohort 3.

As a solution to this Catch-22, we estimate pre- and post-treatment effects *separately*. That is, continuing with the example above, we compare Cohort 1 to Cohort 2 to estimate the pre-treatment effects for Cohort 1. Because these comparisons occur more than seven days away from the treatment of Cohort 2, conditional on our assumptions above, these estimates recover pre-treatment-effect estimates for Cohort 1. Similarly, we compare Cohort 1 and Cohort 3 to recover the post-treatment effects for Cohort 1. Figure 3 shows the assignment pattern for all before and after comparisons.

In total, for each driver sample and labor supply outcome we estimate six different regressions, where we alternate both the cohorts used and the period under considera-

Figure 3: Construction of treatment-control comparisons in event studies



The lines represent the evolution of outcome variables for Cohort 1 (yellow), Cohort 2 (blue), and Cohort 3 (green). Each circle represents the timing of payment of Social Security benefits in a hypothetical month. We construct treatment-control comparisons so that the control cohort is more than seven days away from its treatment. For instance, to estimate pre-treatment anticipation of Social Security benefits for Cohort 1, we use Cohort 2 (blue) as a control.

tion. Formally, for each sample we estimate the following specifications:

$$\begin{aligned} Y_{it} &= \alpha_i + \delta_t + \sum_{k=-8}^{k=0} \theta_k \cdot \tau_{ik} + \epsilon_{it} \\ Y_{it} &= \alpha_i + \delta_t + \sum_{k=1}^{k=8} \theta_k \cdot \tau_{ik} + \epsilon_{it} \end{aligned} \tag{3}$$

where α_i and δ_t denote driver and day fixed effects, τ_{ik} is a dummy variable equal to 1 if driver i is k days away from treatment, and θ_k are the coefficients of interest. The units included in each of these regressions follow the constructions presented in Figure 3, while the samples include observations up to 10 days prior to treatment and 10 days after treatment. Therefore, pre-treatment estimates are relative to 9 and 10 days prior to treatment, while post-treatment estimates are relative to 9 and 10 days after treatment.

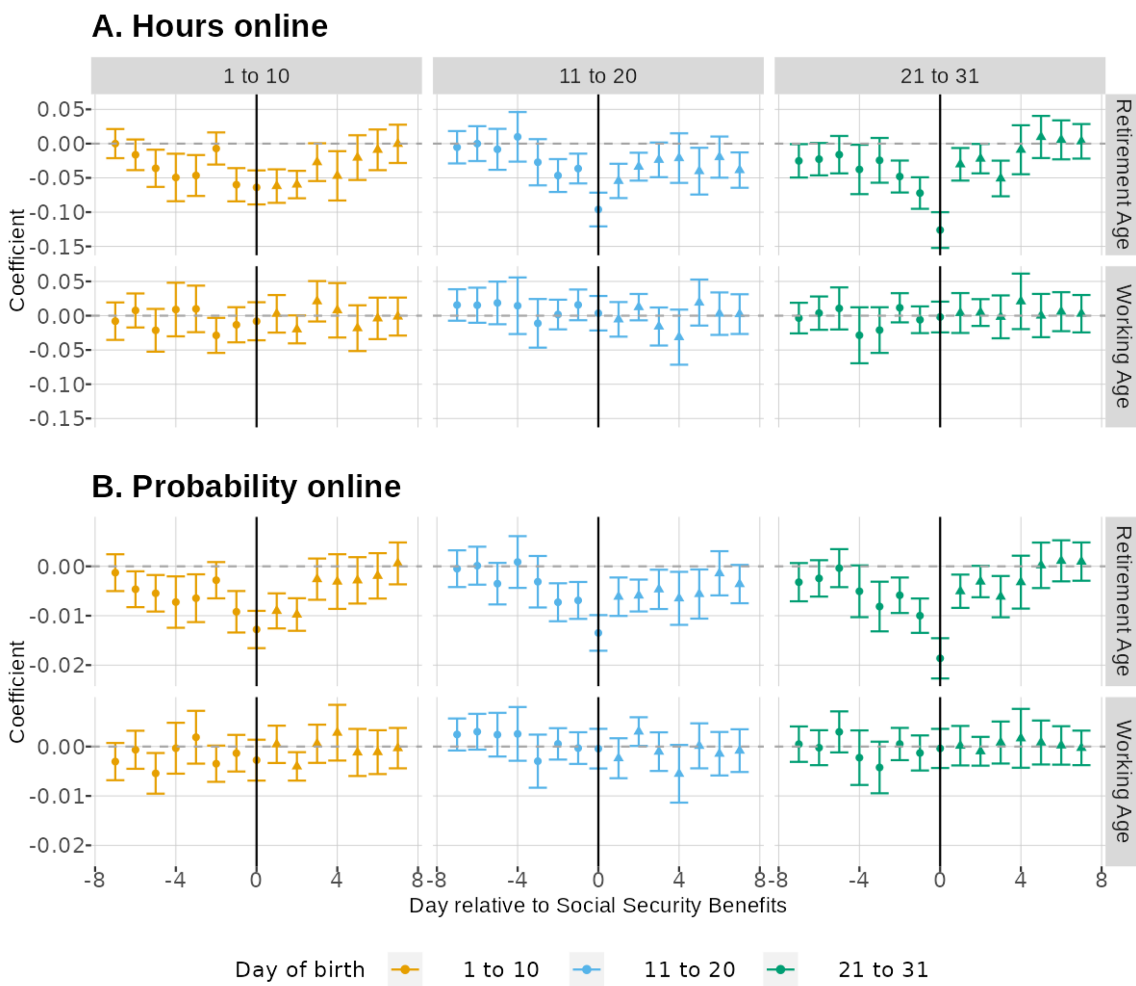
4.2 Paired-Event-Study Results

Figure 4 shows the results of estimating equation 3 for our two preferred measures of labor supply—hours spent online in Panel A and probability of being online on a given day in Panel B. Similar results for two additional measures of labor supply (hours spent driving and probability of driving) are presented in Figure A13. Recall that before and after coefficients are estimated separately and must be interpreted relative to $[-10, -9]$ for pre-treatment coefficients and $[9, 10]$ for post-treatment coefficients.

As with the graphical evidence presented above, we find clear evidence that, on average, Retirement Age drivers are less likely to work on the days leading up to and following the date of receipt of Social Security benefits. Effect sizes are precisely estimated but economically small. On the day of benefit receipt, the probability of being online decreases by 1.25 percentage points for the first cohort and 2 percentage points for the last cohort relative to the control groups, from a baseline level of 0.48—a decline of 2.6% to 4.2%. In contrast, results for the placebo sample (Working Age drivers) show no evidence of labor supply dynamics in the period leading up to or the period following benefit receipt.

Given that we do not observe Social Security participation, a reasonable concern is that the estimates are driven by workers who reached FRA: younger drivers might be choosing to delay claiming Social Security. To test this possibility, Figure A14 presents event-study estimates separated by FRA group for the Retirement Age sample. Drivers who reached FRA respond with similar magnitudes to drivers in the other age groups, although older drivers born later in the month exhibit the strongest response to benefit receipt. While we cannot rule out that some drivers delay receipt, the results suggest that *if* the effect of Social Security benefits on labor supply does not vary with age, then a substantial share of drivers below FRA are also receiving these benefits, which

Figure 4: Paired-event-study estimates



The figure depicts event-study estimates of the impact of days relative to benefit receipt on labor supply (hours online in Panel A, and probability online in Panel B). The event studies are estimated separately for the Retirement Age sample (top rows) and Working Age sample (bottom rows) and separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels.

is consistent with evidence from the ACS described above.

We observe drivers' use of Uber's Instant Pay feature. As noted above, Instant Pay transfers drivers' current earnings to their bank account. According to Uber, most

Figure 5: Paired-event-study estimates, Instant Pay–usage heterogeneity

The figure depicts event-study estimates of the impact of days relative to benefit receipt on probability online, by Instant Pay–usage groups. The event studies are estimated separately for the Retirement Age sample (Panel A) and Working Age sample (Panel B) and separately for the before and after periods and Instant Pay–group heterogeneity, resulting in six different regressions following equations 3 for each sample and heterogeneity group. The high– and low–Instant Pay groups are above and below the median of Instant Pay usage within each sample. The reference dummies are −8 for the before period and +8 for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels.

drivers receive the transfer immediately.[8] Instant Pay can be thought of as a proxy for a cash-on-hand constraint. One important caveat is that because Instant Pay can only be used when a driver worked recently, weekly usage of Instant Pay will be mechanically correlated with labor supply. To partially address this concern, we segment the sample based on whether a driver uses Instant Pay more or less than the sample median over the two-year period, thus capturing cash constraints at the driver (rather than driver-day) level. Results are presented in Figure 5. A clear pattern emerges: only drivers who use Instant Pay more than the median respond to Social Security benefit receipt. As before, we observe a decline in labor supply on the days before and after benefit receipt for this group.

Comparing magnitudes between Figures 4 and 5, we see that point estimates in Figure 5 are roughly double those in Figure 4. This is consistent with averaging over groups that respond to the receipt of Social Security benefits and groups that do not. Section 5.1 explores this in greater detail.

The negative anticipation effects are surprising. A model of dynamic labor supply with imperfect foresight would predict an increase in labor supply on the days before benefit payment to alleviate an imperfectly anticipated cash constraint. The negative anticipation effect we observe is even stronger in the group of drivers who use Instant Pay the most, the group we would expect to be the most cash constrained. Observed behavior is more consistent with a model of targeted-earning behavior, where the target is averaged over a multi-day window rather than a single point in time.

4.3 Robustness: Two-stage difference-in-differences

The paired event-study approach provides a transparent way to estimate cohort-specific treatment effects. However, causal identification flows from the assumption that bene-

[8]Information about Instant Pay is available on Uber's website at `https://www.uber.com/us/en/drive/driver-app/instant-pay/`. Transfers to a standard debit card incur a $1.25 fee and drivers are limited to six transfers per day.

fit receipts has no impact on labor supply more than a week away from treatment and makes comparing ex-ante and ex-post coefficients challenging. To scrutinize our identifying assumptions, we use the two-stage difference-in-difference estimator presented in Butts (2021); Gardner et al. (2024).

This alternative approach offers an heterogeneity-robust and efficient estimation of average treatment effects on the treated in the presence of staggered treatment timing. While this estimator does not yield cohort-specific estimates, it offers a check of the identifying assumption of the previous paired event-study. Consider the following event-study specification, where g is the treatment cohort:

$$Y_{igt} = \alpha_g + \delta_t + \sum_{k=-17}^{k=17} \theta_k \cdot \tau_{ik} + \epsilon_{igt} \tag{4}$$

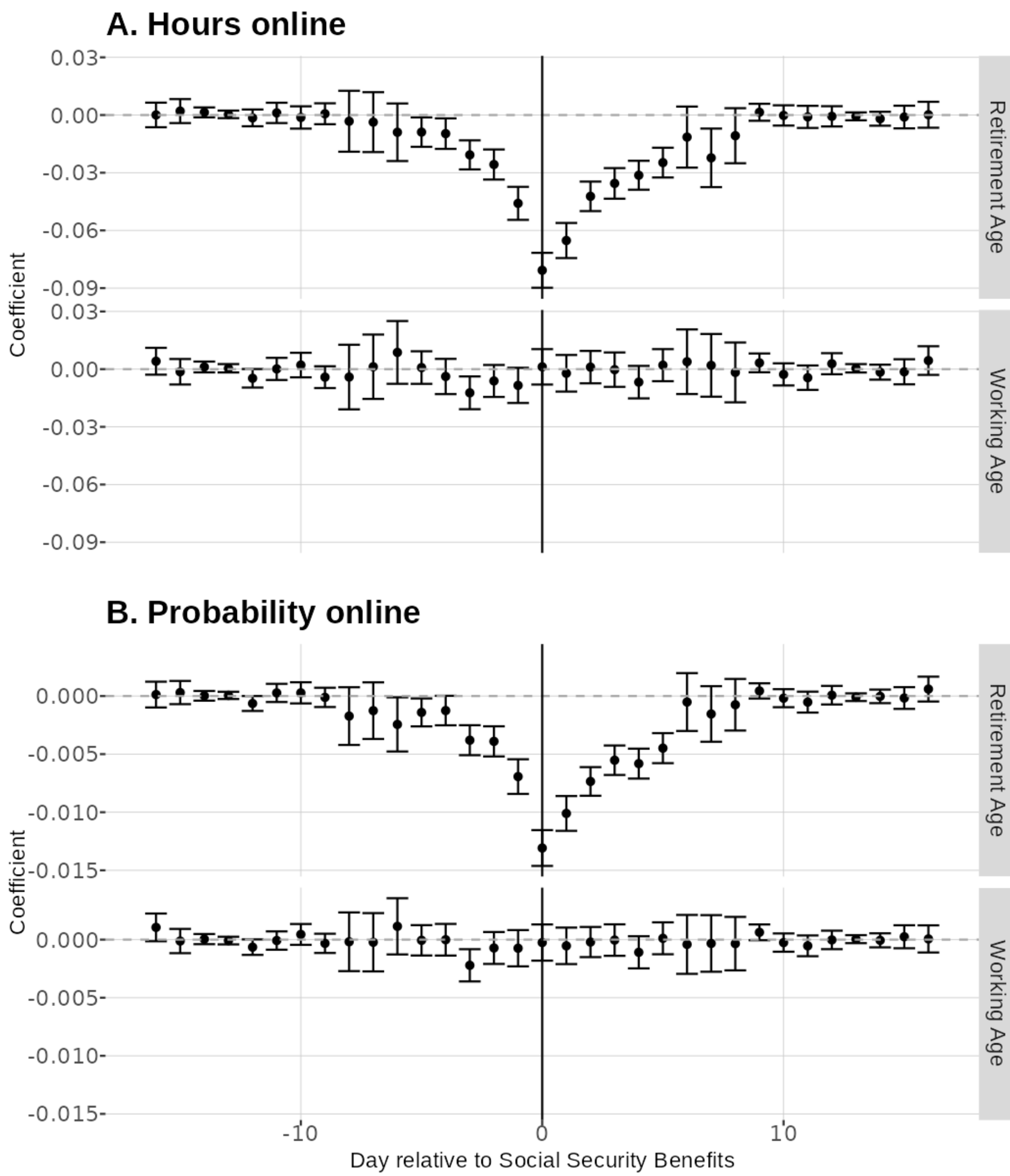
The two-step difference-in-differences procedure is as follows. First, regress the outcome on group and time fixed effects using only observations from outside of a two-week (seven days before and seven days after the day of benefit receipt) treatment window for each cohort:

$$Y_{igt} = \alpha_g + \delta_t + \epsilon_{igt} \tag{5}$$

Second, regress the adjusted outcomes $Y_{igt} - \hat{\alpha_g} - \hat{\delta_t}$ on the treatment dummies $\sum_{k=-17}^{k=17} \tau_{ik}$.

Figure 6 presents the results for both samples and for our two main outcomes. Within each sample and for each outcome, the average treatment effects (obtained over the combined cohorts of drivers) are similar in magnitude to the cohort-specific estimates presented in Figure 4. In the Retirement Age sample, labor supply decreases in the four days leading up to the receipt of benefits and slowly recovers back to baseline in the following week. Important for our main results, treatment effects estimates are close to zero and not statistically significant outside of the two-week window around the receipt of benefits. These estimates increase the credibility of the limited anticipation

Figure 6: Two-Stage Difference-in-Differences Estimates



Regression results using the two-stage difference-in-differences estimator described in Gardner et al. (2024), implemented with the package developed by Butts (2021). The event studies are estimated separately for the Retirement Age sample and Working Age sample. The reference time periods are 17 days away from treatment. Error bars display 95% confidence intervals using standard errors clustered at the driver-level.

assumption of the paired event-study above. As above, estimates in the placebo sample of Working Age drivers are small and not statistically significantly different from zero.

5 Unobserved Driver Heterogeneity

5.1 Individual-Level Treatment Effects

The graphical analyses and paired event studies offered evidence that the receipt of Social Security benefits decreases measures of labor supply on the days leading up to and following the disbursement date. However, the average effects are small, and our heterogeneity analysis provides evidence that a subset of drivers may be driving average results. Put differently, it is unclear the extent to which the estimates above are the result of small changes made by many drivers or large changes made by a few drivers. Distinguishing between these scenarios is fundamental: if most drivers respond, this behavior is inconsistent with the neoclassical model of labor supply; alternatively, most drivers behave as predicted by the neoclassical model.

In Section 4.2, we followed the standard approach to investigating heterogeneity by estimating *conditional* average treatment effects. This standard approach has two main drawbacks. First, it is a priori unclear which covariates should be used to estimate conditional average treatment effects. Economic theory suggests we should estimate heterogeneity with respect to age, as the probability of participation increases with age, but Figure A14 shows that average treatment effects on the treated are broadly similar between groups. Alternatively, Figure 5 shows important heterogeneity with regard to Instant Pay use. One might estimate heterogeneity with respect to all available covariates, but this increases the probability of false discovery. The second drawback, specific to our setting, is that we observe few individual-driver characteristics (only gender and state of residence).

As a way forward, we propose an approach that allows for *unobserved* driver heterogeneity and estimate treatment effects at the individual level. For each driver, we regress our measure of labor supply on a dummy variable equal to 1 during the window $[t-3,\ t+3]$ around the date of Social Security receipt and 0 at other times.[9] Specifically, we run the regression

$$\text{Labor Supply}_t = \alpha + \tau D_t + u_t, \tag{6}$$

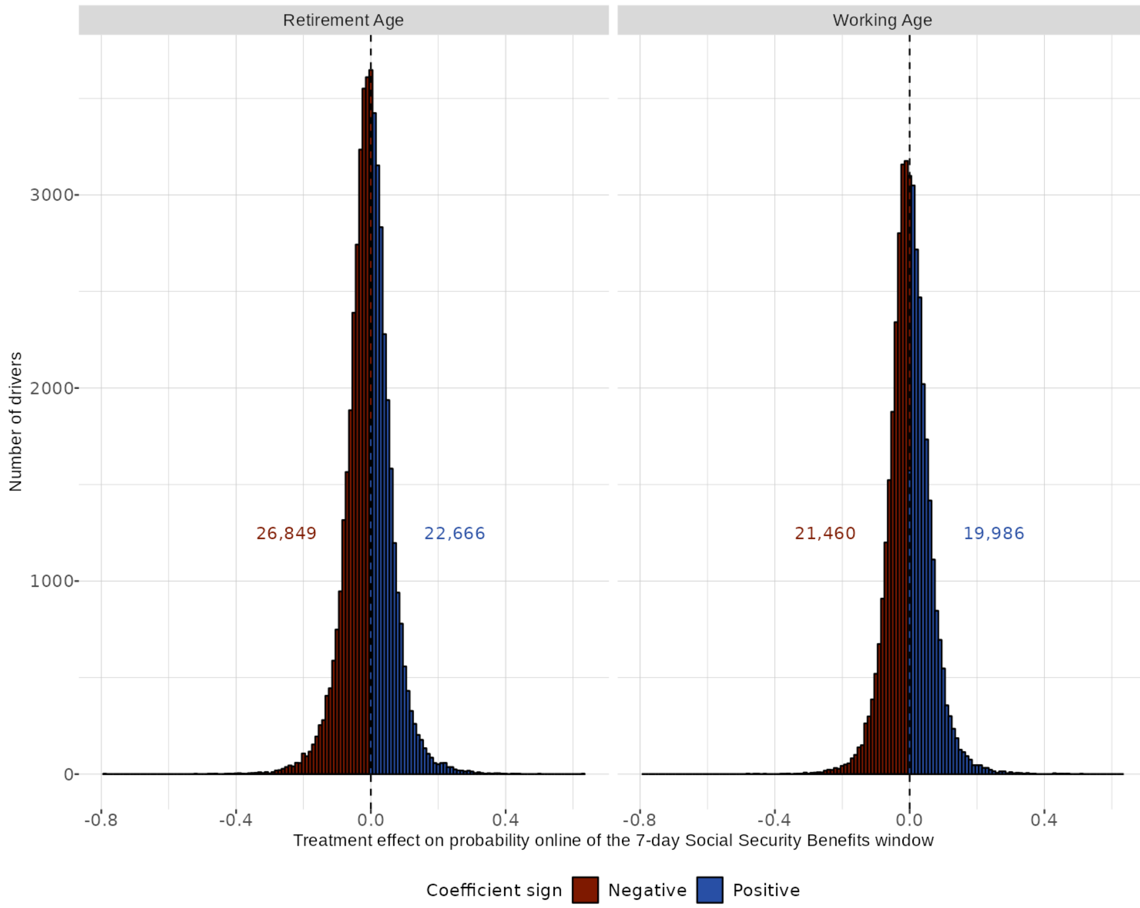
where Labor Supply_t is our measure of labor supply at the extensive margin (a dummy equal to 1 if hours online is positive on a given day and 0 otherwise, residualized by day of month and cohort fixed effects as in Figure A23), D is a dummy equal to 1 within the seven-day window around the dates of Social Security receipt, and τ denotes the treatment effect of interest, which is estimated *within* driver.

Figure 7 presents a histogram of the estimated individual treatment effects estimated for the 90,961 Retirement Age and Working Age drivers. At first glance, there appears to be considerable heterogeneity in individual-driver responses to the timing of Social Security benefits. For both samples, the distribution of $\hat{\tau}$ is asymmetric, with roughly 4,000 more negative treatment effects than positive treatment effects for the Retirement Age sample and about 1,500 more negative treatment effects than positive treatment effects for the Working Age sample. This asymmetry may be due to noise and sampling variation, or it might represent a statistically significant difference.

Figure 8 presents histograms of p-values for the Retirement Age and Working Age samples, using *probability online* as the outcome. Because we estimate more than 90,000 coefficients, we use a multiple hypotheses testing adjustment to control the risk of false discovery. Figure 8 shows clear evidence of bunching of p-values below 0.1 for our Retirement Age sample but not for the Working Age sample, providing additional support that we estimate a response to Social Security benefits. Further, bunching only occurs for *negative* individual treatment effects, consistent with the idea that Social Security benefits might decrease labor supply for some drivers, but they do not increase

[9]Results are robust to the choice of alternative windows.

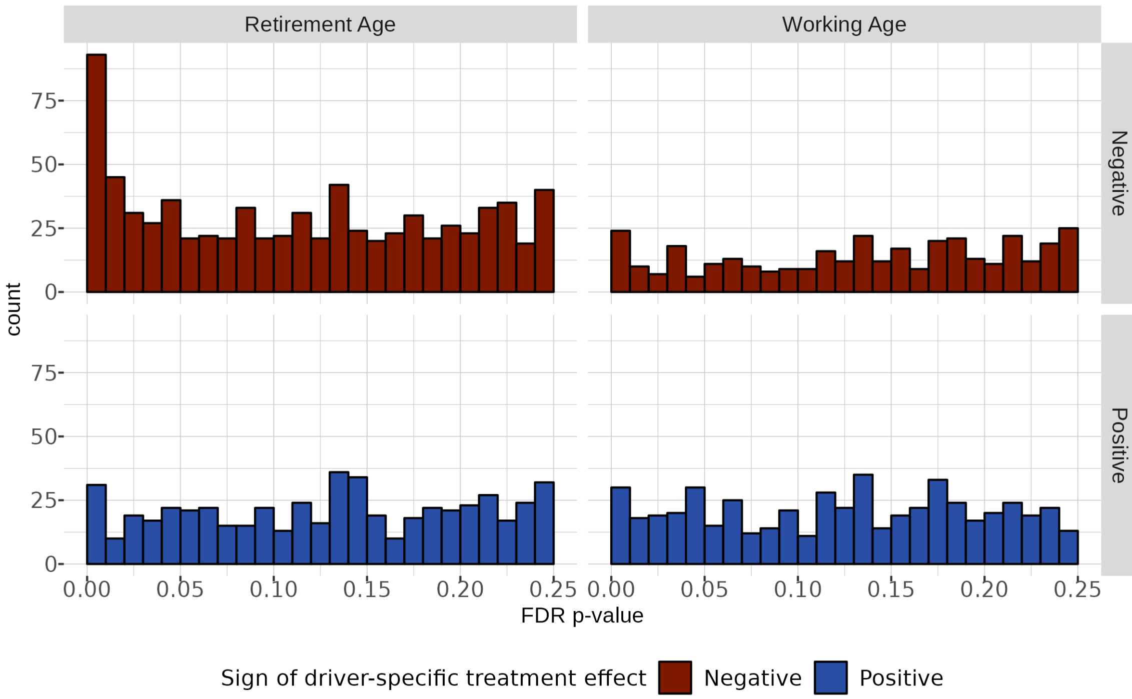
Figure 7: Histogram of individual-level treatment effects



The histogram depicts driver-level treatment-effect estimates of the impact of the week-of-benefit timing on the probability of being online, following regression 6. The colored numbers represent the number of negative (in red) and positive (in blue) treatment effects within each driver sample.

it for any drivers. Finally, we note that only about 1% of drivers have an FDR-corrected p-value below 0.10. While the FDR correction can be overly conservative, these results suggest that only a small number of drivers respond strongly to Social Security benefits. Results are similar if using *hours online* as the outcome. In the appendix, we show these findings and the following are robust to using a Bonferroni correction (Figures A16 and A17). We also present and implement a large scale finite-sample inference approach, which yields similar results (Figures A28 and A29).

Figure 8: Histogram of the individual treatment-effect FDR-adjusted p-values



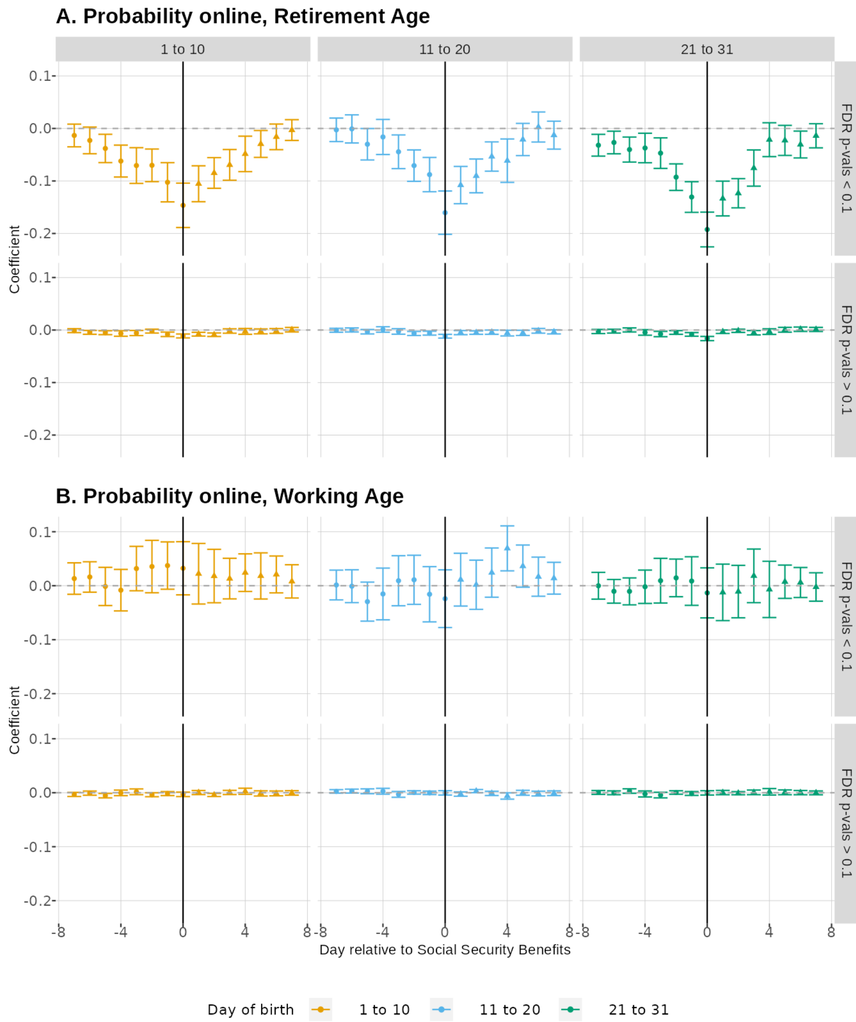
Individual treatment-effect FDR-adjusted p-values, using probability online as the outcome. Each driver-level treatment effect i is estimated with regression 6, and the FDR correction applies to the entire sample of drivers. The x-axis is truncated at 0.25 for visibility.

5.2 Paired Event Studies Conditional on Individual Treatment Effects

We extend our investigation by rerunning our paired event studies separately for two groups of drivers, based on whether their adjusted p-values are above or below 0.1. We present the standard asymptotic (and clustered) confidence intervals for reference only, but note that these intervals are too small: they fail to account for the fact that the groups are the result of an estimation procedure. The construction of valid confidence intervals to account for sample preselection is an active area of econometric research but beyond the scope of the current paper.

Figure 9 presents the pattern of estimates from paired event studies for our preferred

Figure 9: Paired-event-study estimates, heterogeneity by driver-level response



The figure depicts event-study estimates of the impact of days relative to benefit receipts on probability online, by individual-treatment-effects group. Within each driver sample, driver-level individual treatment effects are estimated following regression 6 and p-values are adjusted to control the False Discovery Rate. Within each panel (A and B), the top and bottom rows present event studies for the drivers with a FDR p-value below and above 0.10, respectively. The event studies are estimated separately for the Retirement Age sample (Panel A) and Working Age sample (Panel B) and separately for the before and after periods and individual-treatment-effect p-value group, resulting in six different regressions following equations 3 for each sample or individual-treatment-effect group. The error bars represent 95% confidence intervals with standard errors clustered at the day and individual levels. They are not adjusted for postselection and are presented as auxiliary evidence only. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3.

measure of labor supply for each of the two samples. The bottom row in each panel reports estimated coefficients for the group of Retirement Age drivers for whom the FDR-adjusted p-value is above 0.1—"non-responders.'. We find the effects of benefit receipt are small and negative only the day before, the day of, and the day after benefit receipt—largely supportive of the standard life-cycle model of labor supply. Point estimates are exactly zero for the Working Age sample.

These results are in marked contrast to estimated effects for the individuals with adjusted p-values below 0.1—"responders." The decline in labor supply around the date of benefit disbursement for this group is economically meaningful, with a 20 percentage point decline in the probability of working during the day of benefit receipt—a 40% decline relative to baseline. The effects last for up to one week before and after benefit receipt, with treatment effects gradually approaching zero outside this window. We find no labor supply responses in our placebo sample of Working Age drivers.

Finally, we use the weekly earnings data to assess foregone income. On average, responders earn \$216 during the week of benefit receipts, against \$269 for the other weeks. This suggests that Social Security-induced labor supply distortions cause a 22% decline in the week of benefits relative to a \$244 weekly average. Retirement Age non-responders weekly earnings are essentially unchanged during the week of Social Security benefit receipts. This is also the case for all Working Age drivers.

Overall, this analysis based on individual-level treatment effects reveals substantial *unobserved heterogeneity* — that is, heterogeneity that we cannot explain with driver-level characteristics, as we observe a very limited set of covariates. The results suggest that statistically significant but economically small average effects estimated in Section 4.3 do not represent average micro-behavior. Rather, a small subset of drivers responds substantially to the timing of Social Security benefits. For the large majority of drivers, departures from the standard model of neoclassical supply are extremely small and do not lead to significantly different earnings during the week of benefit receipt.

6 Conclusion

This paper investigates the labor supply response of Retirement Age Uber drivers to the receipt of monthly Social Security benefits. Using a well-established research design that leverages variation in the timing of benefit payments and a novel identification strategy that allows for ex ante and ex post effects, we find that, on average, Retirement Age drivers reduce their labor supply both in anticipation of and following their receipt of benefits. These effects are precisely estimated, robust to the use of a recently developed two-stage estimator, and driven by the extensive margin (decision to work or not work on any given day) but economically small on average. We find no comparable effects for a placebo group of Working Age drivers.

Leveraging individual-level treatment effects to assess unobserved heterogeneity, we find that small average treatment effects are largely due to a small group of drivers (about 1% of the sample) experiencing substantial treatment effects: they reduce their labor supply by as much as 40% on the day of benefit receipt. In contrast, most drivers do not exhibit any substantial departure from the standard model of labor supply.

Our results indicate that while most drivers in our sample are forward looking and do not meaningfully respond to a lump-sum transfer, there are substantial violations of the standard model of labor supply among a small group of drivers. These violations cannot be explained by liquidity constraints, which would lead to increasing labor supply on the days leading to the receipt of Social Security benefits, but instead reflect nonstandard preferences such as target-earning behavior or inconsistent time preferences. While our data do not allow us to rule out competing behavioral explanations, observed patterns are consistent with hand-to-mouth consumption: responding drivers reduce labor supply ahead of benefit receipt, and labor supply returns to baseline between two and six days after benefit receipt.

For most Retirement Age drivers, the welfare consequences of the labor supply reductions are likely small. The average driver reduces hours online by 2% and effects are only statistically significant in a small window on the days around benefit receipt. For this group earnings are essentially unchanged. However, for a small group of drivers, estimated magnitudes are economically important – drivers spend roughly 40% less time online on the day of benefit receipt. For responding drivers, anticipation effects begin nearly a week before benefit receipt and end nearly a week after, with foregone earnings during the week of benefits amounting to more than 20% of weekly earnings on average. To the extent that excess sensitivity to benefit receipt affects nearly two weeks of the benefit month, this suggests that the arbitrary timing and lump sum nature of OASDI benefits likely comes at a significant welfare cost for these drivers.

A key contribution of our paper is to distinguish between small average effects and large effects for a small group of Retirement Age drivers. With the latter, broad-based policies are likely to be less cost-effective than those focused on the few responders. However, our setting also highlights that a policymaker might not be able to accurately target those responders: treatment effects are fairly similar along most of the covariates we observe for Uber drivers. In this context, targeting might be best achieved through self-selection. For instance, the Social Security Administration might pay benefits once a month by default but offer individuals the possibility to select a bi-weekly payment schedule. Given well-established default biases, we might expect only individuals with the highest benefits from an alternative payment schedule to self-select into one.

While our unique data set provides high-frequency observations of labor supply, it is limited in terms of the number of covariates useful to explain differences in behavior. This limits our ability to precisely identify mechanisms. In particular, we do not observe whether drivers participate in Social Security or, as is the case for a recent immigrant without the requisite employment history, are even eligible to participate. Rather, much like researchers studying EITC receipt, we rely on eligibility and secondary data that confirm high rates of take-up among eligible groups. This implies that some of the null individual-level treatment-effect estimates are the result of non-participation rather than rationality. However, given the small share of responders in our sample, this seems unlikely to undermine our key finding that most drivers behave in a way that is consistent with standard economic theory.

This paper extends a large empirical literature on behavioral changes in response to large lump-sum transfers. Broadly speaking, prior work finds that households are more likely to consume and less likely to work upon receipt of a lump-sum transfer. However, effect sizes are small. Our results raise the question: to what extent are estimates in earlier papers driven by averaging results over a large population of non-responders and a small population of responders? We are able to consider individual-level treatment effects because of unique data that offer more observations per unit and repeated treatments of units. This suggests that revisiting earlier studies with high-frequency data might prove a fruitful direction for future research.

References

Abraham, K. G., J. Haltiwanger, K. Sandusky, and J. Spletzer (2019). The Rise of the Gig Economy. AEA Papers and Proceedings 109(May), 357–361.

Akee, R. K. Q., W. E. Copeland, G. Keeler, A. Angold, and E. J. Costello (2010, January). Parents' incomes and children's outcomes: A quasi-experiment using transfer payments from casino profits. American Economic Journal: Applied Economics 2(1), 86–115.

Akesaka, M., P. Eibich, C. Hanaoka, and H. Shigeoka (2023). Temporal instability of risk preference among the poor: Evidence from payday cycles. American Economic Journal: Applied Economics Forthcoming.

Aladangady, A., S. Aron-Dine, D. Cashin, W. Dunn, L. Feiveson, P. Lengermann, K. Richard, and C. Sahm (2023). Spending responses to high-frequency shifts in payment timing: Evidence from the earned income tax credit. American Economic Journal: Economic Policy Forthcoming.

Ameriks, J., J. Briggs, A. Caplin, M. Lee, M. D. Shapiro, and C. Tonetti (2020). Older Americans Would Work Longer if Jobs were Flexible. American Economic Journal: Macroeconomics 12(1), 174–209.

Anderson, M., C. McClain, M. Faverio, and R. Gelles-Watnick (2021). The State of Gig Work in 2021. Pew Research Center Report, 67.

Baugh, B., I. Ben-David, H. Park, and J. A. Parker (2021). Asymmetric Consumption Smoothing. American Economic Review 111(1), 192–230.

Baugh, B., J. B. Leary, and J. Wang (2017). When Is It Hard to Make Ends Meet? Working Paper, 34.

Beatty, T. K., M. P. Bitler, X. H. Cheng, and C. van der Werf (2019). SNAP and Paycheck Cycles. Southern Economic Journal 86(1), 18–48.

Beatty, T. K. and C. J. Tuttle (2015). Expenditure response to increases in in-kind transfers: Evidence from the supplemental nutrition assistance program. American Journal of Agricultural Economics 97(2), 390–404.

Bond, T. N., J. B. Carr, A. E. Packham, and J. Smith (2022). Hungry for success? snap timing, high-stakes exam performance, and college attendance. American Economic Journal: Economic Policy 14(4), 51–79.

Butts, K. (2021). did2s: Two-stage difference-in-differences following gardner (2021).

Callaway, B. and P. H. Sant'Anna (2021). Difference-in-differences with multiple time periods. Journal of Econometrics 225(2), 200–230.

Camerer, C. F., L. Babcock, G. Loewenstein, and R. H. Thaler (1997). Labor Supply of New York City Cab Drivers: One Day at a Time. Quarterly Journal of Economics 112(2), 407–441.

Carr, J. B. and A. Packham (2019). Snap benefits and crime: Evidence from changing disbursement schedules. The Review of Economics and Statistics 101(2), 310–325.

Carr, J. B. and A. E. Packham (2021). Snap schedules and domestic violence. Journal of Policy Analysis and Management 40(2), 412–452.

Cesarini, D., E. Lindqvist, M. J. Notowidigdo, and R. Östling (2017, December). The effect of wealth on individual and household labor supply: Evidence from swedish lotteries. American Economic Review 107(12), 3917–46.

Cohen, R. A., A. E. Cha, E. P. Terlizzi, and M. E. Martinez (2021). Demographic variation in health insurance coverage: United states, 2019. National Health Statistics Reports 159, Center for Disease Control.

Cook, C., R. Diamond, and P. Oyer (2019). Older Workers and the Gig Economy. AEA Papers and Proceedings 109, 372–376.

Cotti, C. D., J. M. Gordanier, and O. D. Ozturk (2020, May). Hunger pains? snap timing and emergency room visits. Journal of Health Economics 71, 102313.

Crawford, V. P. and J. Meng (2011). New York city cab drivers' labor supply revisited: Reference-dependent Preferences with Rational-Expectations Targets for Hours and Income. American Economic Review 101(5), 1912–1932.

de Chaisemartin, C. and X. D'Haultfœuille (2020, September). Two-way fixed effects estimators with heterogeneous treatment effects. American Economic Review 110(9), 2964–96.

Deshpande, M. (2016, 10). The Effect of Disability Payments on Household Earnings and Income: Evidence from the SSI Children's Program. The Review of Economics and Statistics 98(4), 638–654.

Dushi, I., H. M. Iams, and B. Trenkamp (2017). The importance of social security benefits to the income of the aged population. Social Security Bulletin 77(2).

Dushi, I. and B. Trenkamp (2021, January). Improving the measurement of retirement income of the aged population. Working Paper ORES Working Paper No. 116, Office of Research, Social Security Administration.

Efron, B. (2013). Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction (IMS ed.). Cambridge University Press.

Evans, W. N. and T. J. Moore (2011). The Short-term Mortality Consequences of Income Receipt. Journal of Public Economics 95(11-12), 1410–1424.

Farber, H. S. (2005). Is Tomorrow Another Day? The labor Supply of New York City Cabdrivers. Journal of Political Economy 113(1), 46–82.

Farber, H. S. (2008). Reference-dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers. American Economic Review 98(3), 1069–1082.

Farber, H. S. (2015). Why you Can't Find a Taxi in the Rain and Other Labor Supply Lessons from Cab Drivers. The Quarterly Journal of Economics 130(4), 1975–2026.

Fehr, E. and L. Goette (2007, March). Do workers work more if wages are high? evidence from a randomized field experiment. American Economic Review 97(1), 298–317.

Fry, Richard and Braga, Dana (2023). Older workers are growing in number and earning higher wages. Report, Pew Research Center.

Gardner, J., N. Thakral, L. T. Tô, and L. Yap (2024). Two-stage difference-in-differences.

Gelber, A., T. J. Moore, and A. Strand (2017, August). The effect of disability insurance payments on beneficiaries' earnings. American Economic Journal: Economic Policy 9(3), 229–61.

Goldin, J., T. Homonoff, and K. Meckel (2022). Issuance and Incidence: SNAP Benefit Cycles and Grocery Prices. American Economic Journal: Economic Policy 14(1), 152–178.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics 225(2), 254–277. Themed Issue: Treatment Effect 1.

Goodman-Bacon, A. and L. McGranahan (2008). How do eitc recipients spend their refunds? Economic Perspectives 32(2).

Gregory, C. A. and J. E. Todd (2021). Snap timing and food insecurity. PLOS One 16(2), e0246946.

Gross, T., T. Layton, and D. Prinz (2022). The Liquidity Sensitivity of Healthcare Consumption: Evidence from Social Security Payments. American Economic Review: Insights.

Hall, J. V. and A. B. Krueger (2018). An Analysis of the Labor Market for Uber's Driver-Partners in the United States. ILR Review 71(3), 705–732.

Hastings, J. and J. M. Shapiro (2018, December). How are snap benefits spent? evidence from a retail panel. American Economic Review 108(12), 3493–354–.

Imbens, G. W., D. B. Rubin, and B. I. Sacerdote (2001, September). Estimating the effect of unearned income on labor earnings, savings, and consumption: Evidence from a survey of lottery players. American Economic Review 91(4), 778–794.

Jackson, E. (2022). Availability of the Gig Economy and Long Run Labor Supply Effects for the Unemployed. Working paper, IRS-SOI.

Jones, D. and I. Marinescu (2022, May). The labor market impacts of universal and permanent cash transfers: Evidence from the alaska permanent fund. American Economic Journal: Economic Policy 14(2), 315–40.

Katz, L. F. and A. B. Krueger (2019). The rise and nature of alternative work arrangements in the united states, 1995-2015. ILR Review 72(2), 382–416.

LaLumia, S. (2013). The EITC, Tax refunds, and unemployment spells. American Economic Journal: Economic Policy 5(2), 188–221.

Leary, J. B. and J. Wang (2014). Liquidity Constraints and Budgeting Mistakes: Evidence from Social Security Recipients. Working paper (3), 1–37.

McGranahan, L. and D. W. Schanzenbach (2013). The earned income tax credit and food consumption patterns. Working Paper 14, Federal Reserve Bank of Chicago.

Mu, Y., E. A. Rubin, and E. Zou (2022). What's Missing in Environmental (Self-)Monitoring: Evidence from Strategic Shutdowns of Pollution Monitors. NBER Working Paper, 60.

Porell, F. and T. Bond (2020). Examining the nest egg: The sources of retirement income for older americans. Technical report, National Institute on Retirement Security, https://www.nirsonline.org/reports/examining-the-nest-egg/.

Poterba, J. M. (2014). Retirement security in an aging population. American Economic Review: Papers and Proceedings 104(5), 1–30.

Ramnath, S., J. B. Shoven, and S. N. Slavov (2021). Pathways to retirement through self-employment. Journal of Pension Economics and Finance 20(2), 232–251.

Shapiro, J. M. (2005). Is there a Daily Discount Rate? Evidence from the Food Stamp Nutrition Cycle. Journal of Public Economics 89(2-3), 303–325.

Shu, S. and J. W. Payne (2023). Social security claiming intentions: Psychological ownership, loss aversion, and information displays. Working Paper 31499, National Bureau of Economic Research.

Smith, T. A., J. P. Berning, X. Yang, G. Colson, and J. H. Dorfman (2016). The Effects of Benefit Timing and Income Fungibility on Food Purchasing Decisions among Supplemental Nutrition Assistance Program Households. American Journal of Agricultural Economics 98(2), 564–580.

Social Security Administration (2020). Social Security Bulletin: Annual Statistical Supplement.

Souleles, N. S. (1999). The response of household consumption to income tax refunds. American Economic Review 89(4), 947–958.

Stephens Jr., M. (2003). "3rd of tha month": Do social security recipients smooth consumption between checks? American Economic Review 93(1), 406–422.

Stephens Jr., M. and T. Unayama (2011). The Consumption Response to Seasonal Income: Evidence from Japanese Public Pension Benefits. American Economic Journal: Applied Economics 3(4), 86–118.

Sun, L. and S. Abraham (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics 225(2), 175–199. Themed Issue: Treatment Effect 1.

Thakral, N. and L. T. Tô (2021). Daily Labor Supply and Adaptive Reference Points. American Economic Review 111(8), 2417–2443.

United Nations (2017). World population ageing. Report 31499, Department of Economic and Social Affairs.

Wettstein, G. and M. S. Rutledge (2023). Is nontraditional work at older ages associated with better retirement security? The Journal of Retirement 10(4), 49–68.

Yang, T.-T. (2018). Family labor supply and the timing of cash transfers: Evidence from the earned income tax credit. Journal of Human Resources 53(2), 445–473.

Zaki, M. and J. E. Todd (2021). Price Consciousness at the Peak of "Impatience". Journal of Human Resources, 0121–11411.

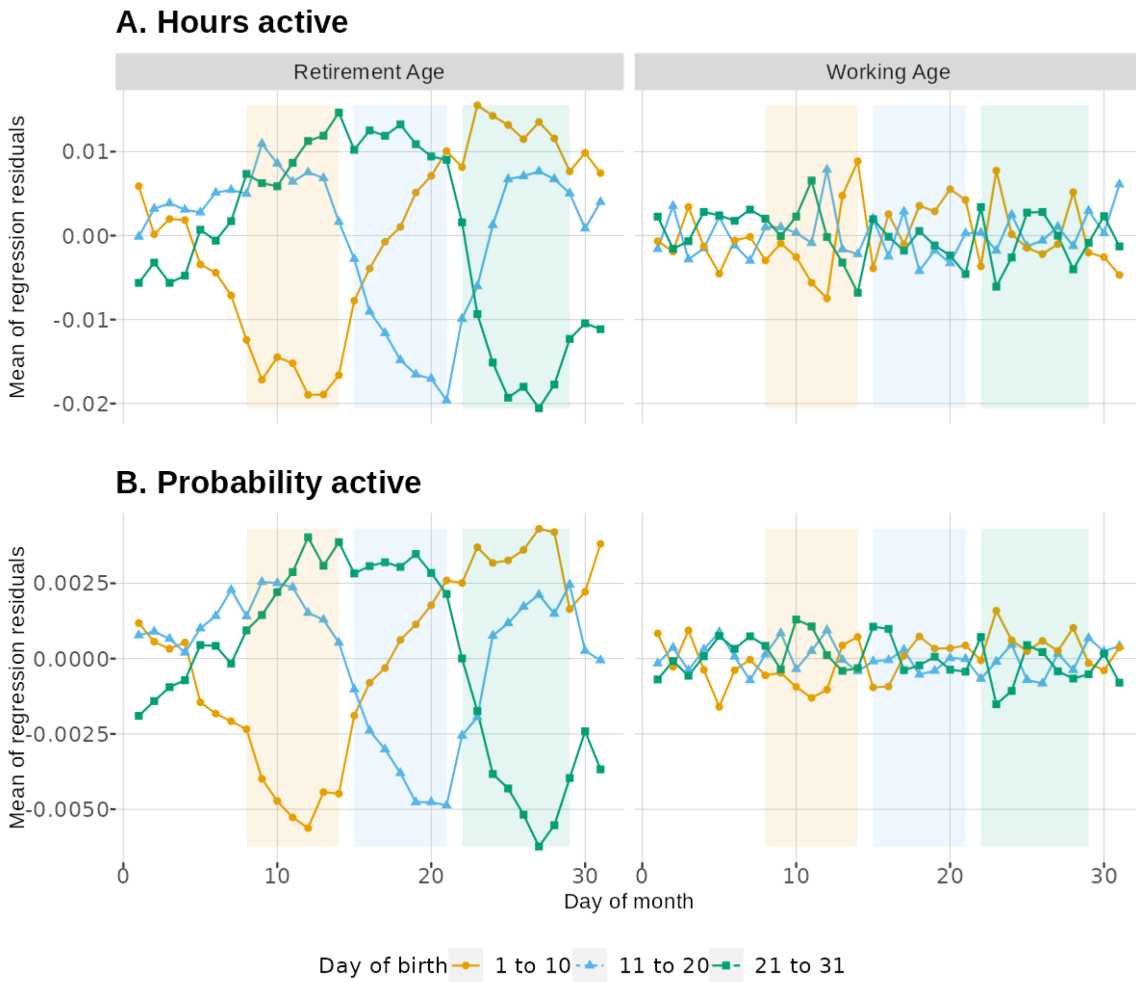
A Online Appendix

A.1 Additional descriptive results

Figure A10: Drivers' year of birth

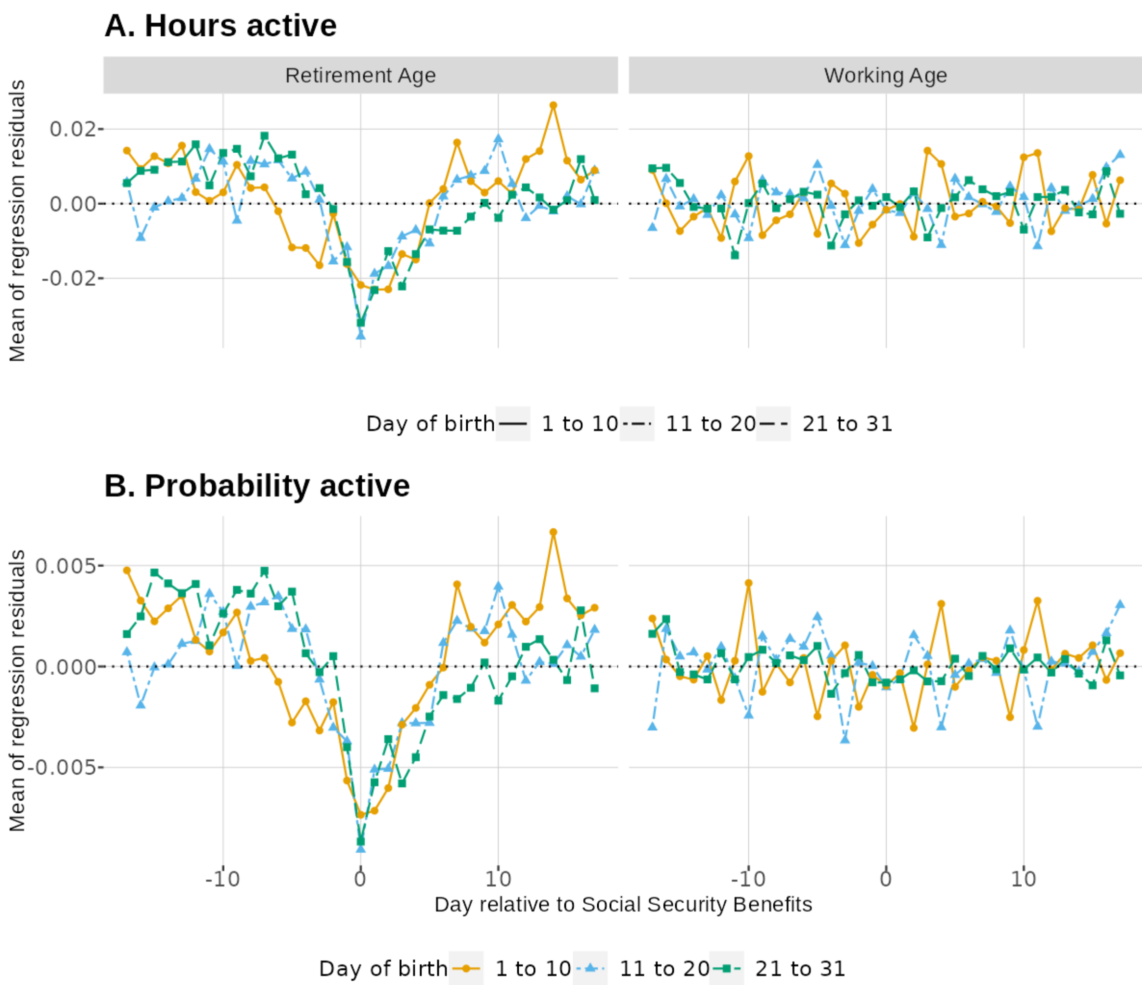
Number of drivers by year of birth bins. The Retirement Age facets comprise a random sample of drivers born in 1956 and earlier, whereas the Working Age facets comprise drivers born between 1958 and 1961.

Figure A11: Labor supply residualized by day of month, alternative outcomes



The figure presents the mean residuals of the regressions $Y_{it} = D_t + B_i + \epsilon_{it}$, where Y_{it} is a measure of labor supply on day t for driver i (number of hours spent driving in panel A, probability of driving in panel B), D_t are day of month fixed effects, and B_i are day of birth cohorts fixed effects (1 to 10, 11 to 20, or 21 to 31), with residuals computed separately for the Working Age and Retirement Age samples. The colored rectangles outline the time windows when each day of birth cohort is expecting to receive their benefits (these are windows rather than points as Wednesdays fall on different dates each month).

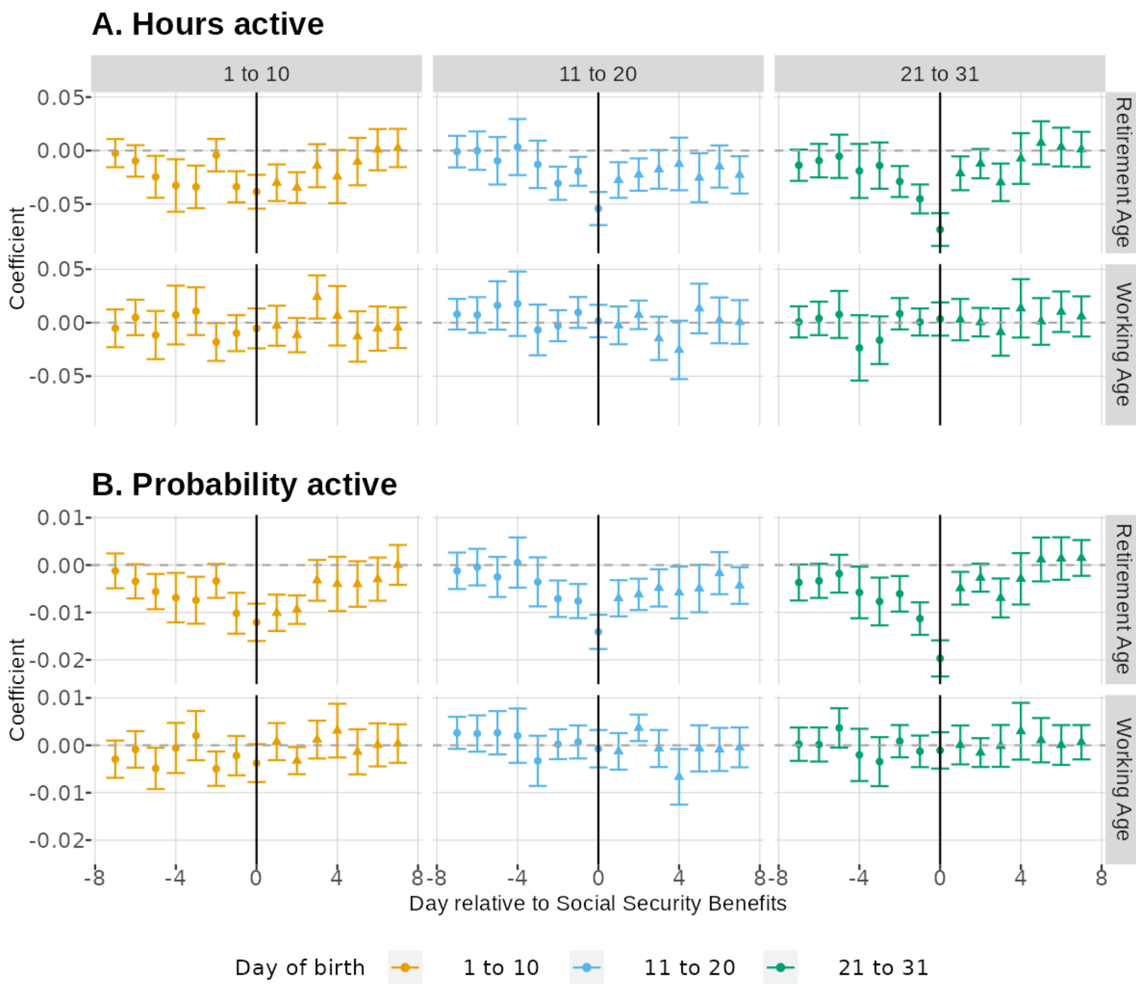
Figure A12: Labor supply residualized by date, alternative outcomes



The figure presents the mean residuals of the regressions $Y_{it} = Date_t + B_i + \epsilon_{it}$, where Y_{it} is a measure of labor supply on day t for driver i (number of hours spent driving in panel A, probability of driving in panel B), $Date_t$ are date fixed effects, and B_i are day of birth cohorts fixed effects (1 to 10, 11 to 20, or 21 to 31), with residuals computed separately for the Working Age and Retirement Age samples and plotted relative to the day of receipt of benefits.

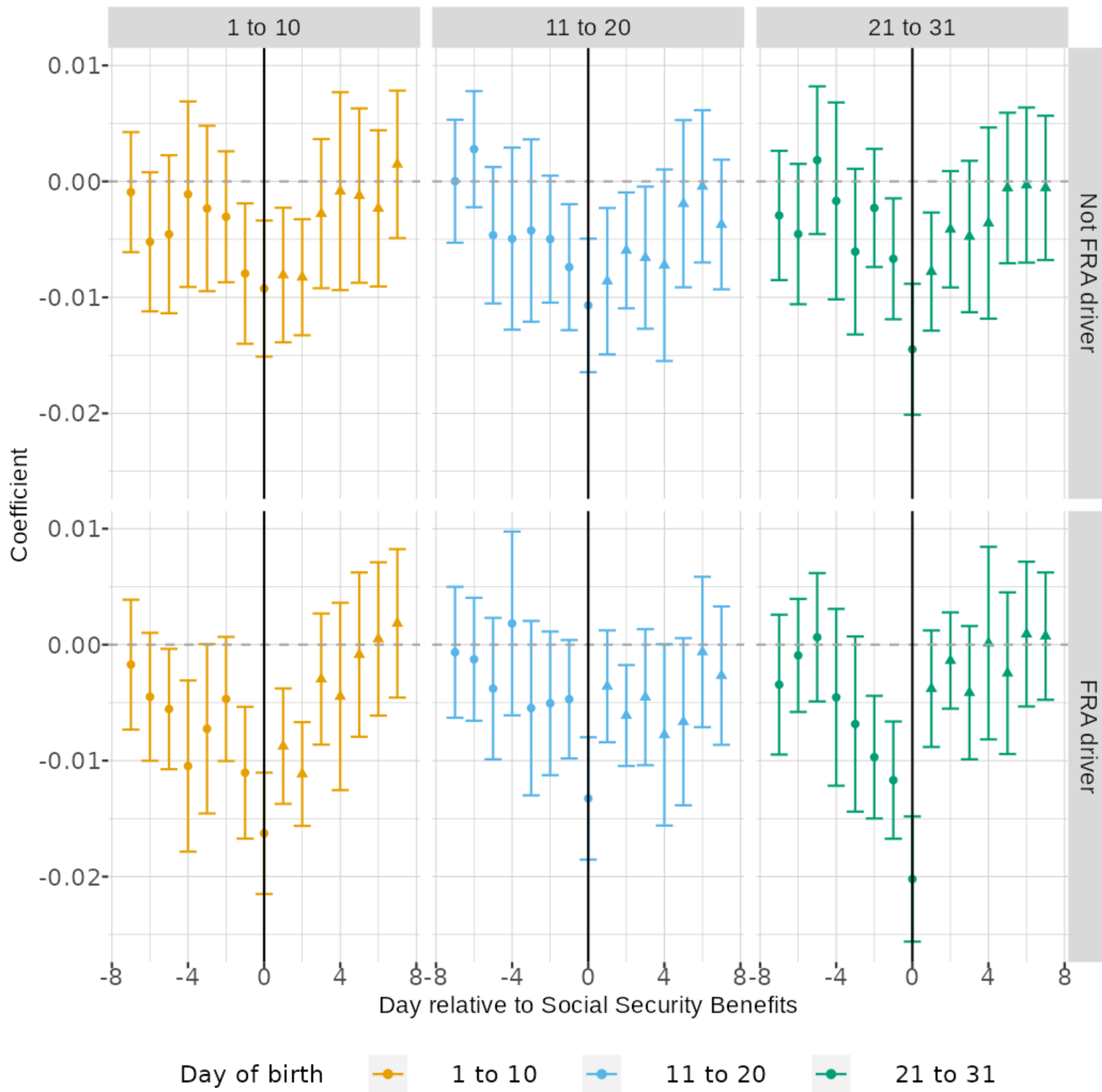
A.2 Event studies

Figure A13: Other outcomes



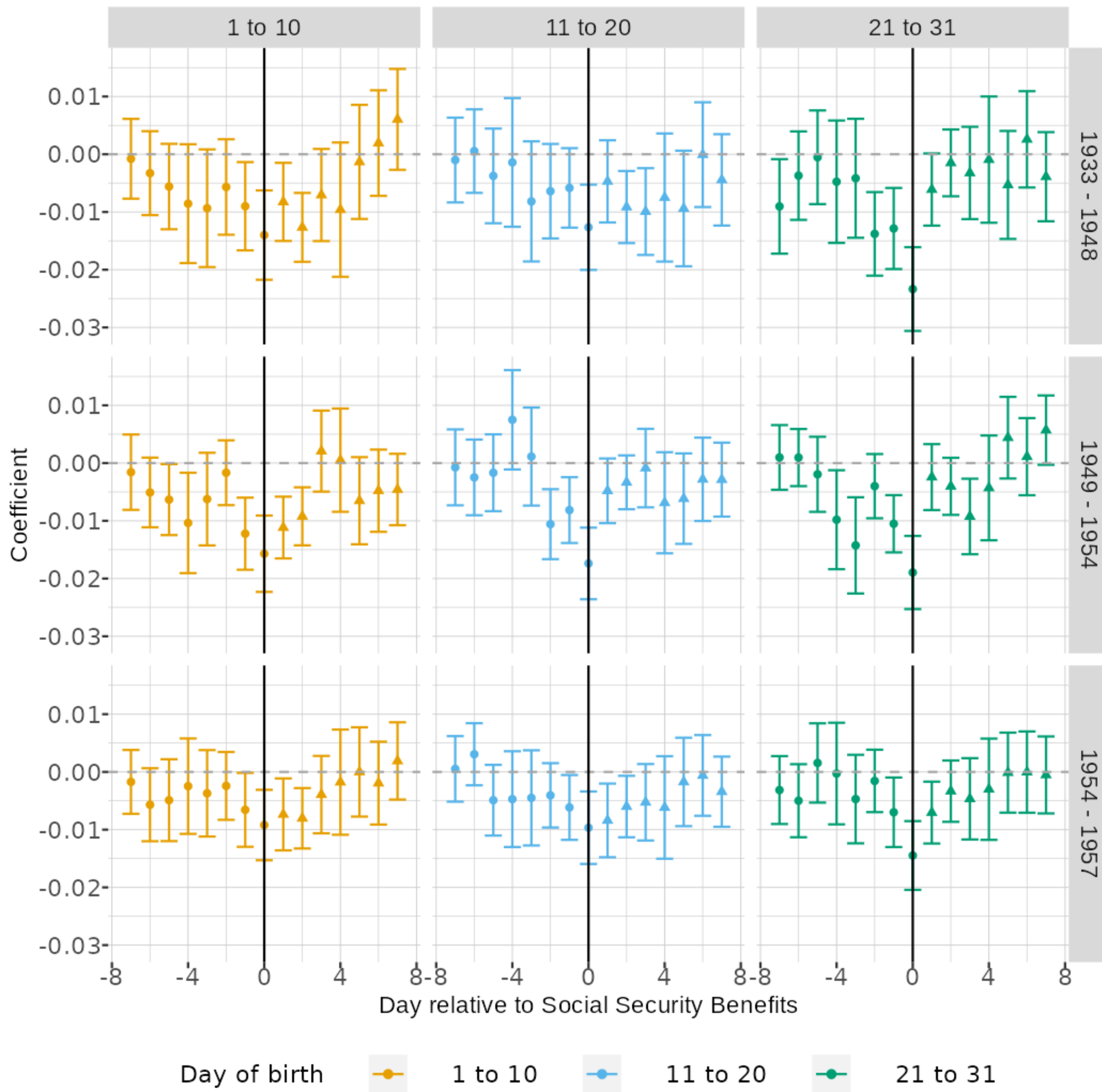
The figure depicts event-study estimates of the impact of days relative to benefit receipt on labor supply (hours active in Panel A, and probability active in Panel B). The event studies are estimated separately for the Retirement Age sample (top rows) and Working Age sample (bottom rows) and separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are −8 for the before period and +8 for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels.

Figure A14: Event studies, probability online heterogeneity by Full Retirement Age



The figure depicts event-study estimates of the impact of days relative to benefit receipt on Probability Online, highlighting heterogeneity for Retirement Age workers before and after the Full Retirement Age. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels. The Full Retirement Age (FRA) is 66 for drivers born from 1943 to 1954, and 67 for drivers born after 1959. For the drivers in our sample born between 1955 and 1959, the FRA increased from 66 to 67 in two-month increments. In this figure, we exclude the small group of drivers who become FRA during our period of study (2018-2019).

Figure A15: Retirement Age drivers, probability online, heterogeneity by age group



The figure depicts event-study estimates of the impact of days relative to benefit receipt on Probability Online, highlighting heterogeneity for Retirement Age by age. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are −8 for the before period and +8 for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels.

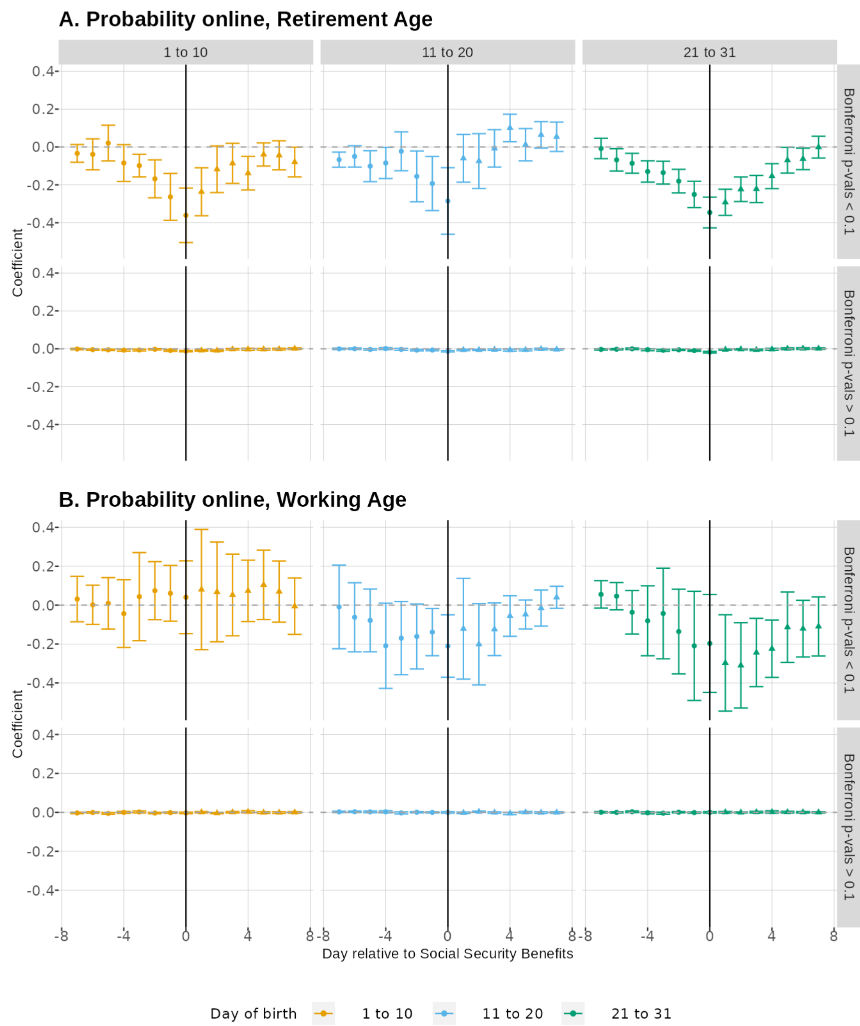
A.3 Individual treatment effects analysis

A.3.1 Bonferonni correction

Figure A16: Histogram of the individual treatment-effect Bonferroni-adjusted p-values

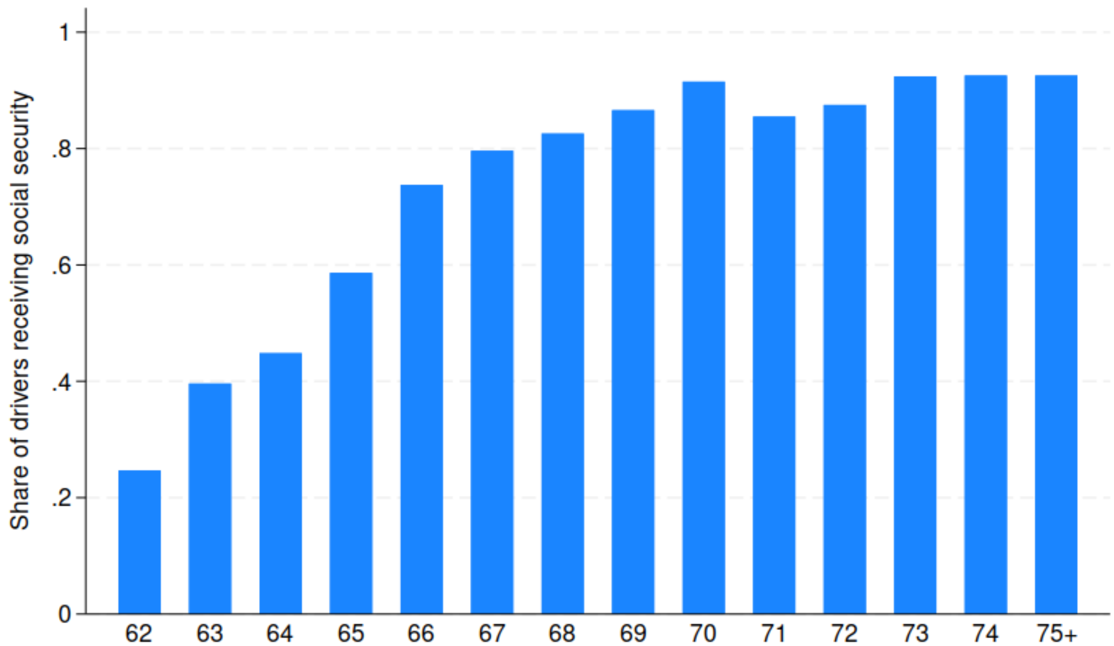
Individual treatment-effect Bonferonni-adjusted p-values, using probability online as the outcome. Each driver-level treatment effect i is estimated with regression 6, and the Bonferonni correction applies to the entire sample of drivers. Because of the large mass of p-values equal to 1, we truncate the x-axis at 0.99 for visibility.

Figure A17: Paired-event-study estimates, heterogeneity by driver-level response, Bonferonni correction



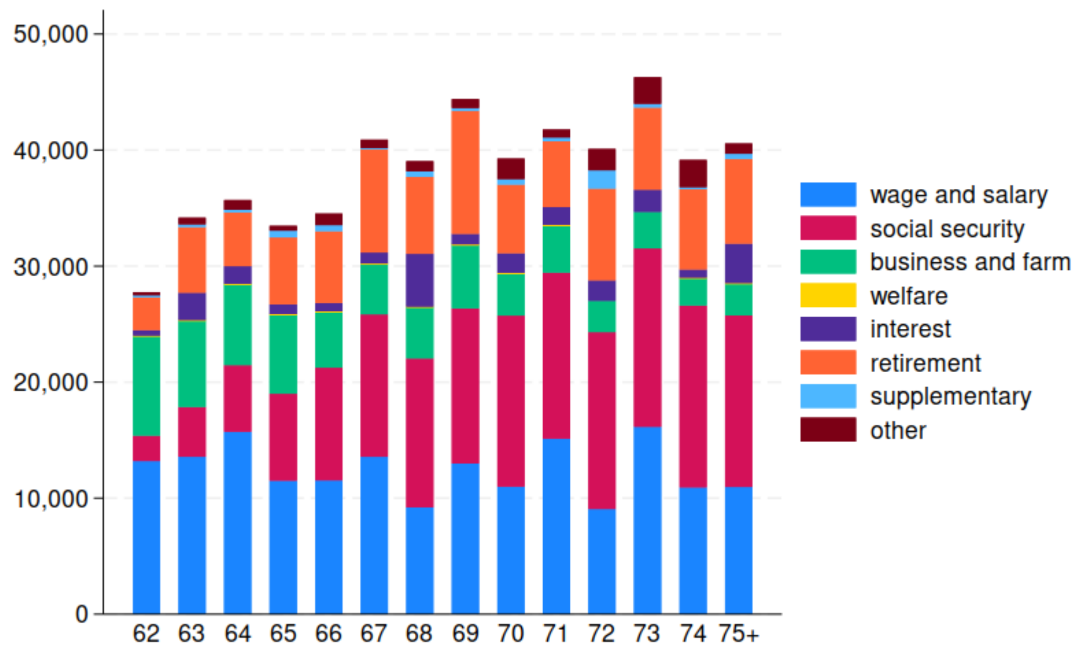
The figure depicts event-study estimates of the impact of days relative to benefit receipts on probability online, by individual-treatment-effects group. Within each driver sample, driver-level individual treatment effects are estimated following regression 6 and p-values are Bonferonni-adjusted. Within each panel (A and B), the top and bottom rows present event studies for the drivers with a Bonferonni p-value below and above 0.10, respectively. The event studies are estimated separately for the Retirement Age sample (Panel A) and Working Age sample (Panel B) and separately for the before and after periods and individual-treatment-effect p-value group, resulting in six different regressions following equations 3 for each sample or individual-treatment-effect group. The error bars represent 95% confidence intervals with standard errors clustered at the day and individual levels. They are not adjusted for postselection and are presented as auxiliary evidence only. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3.

Figure A18: Social Security Receipt by Drivers by Age, ACS 2018-2019.



Data is from the the American Community Survey (ACS) for 2018 and 2019. We restrict the sample to survey respondents aged 62 and up, with occupation code corresponding to taxi drivers or shuttle drivers and chauffeurs. We generate an indicator variable equal to one if the respondent received positive income from social security, and zero otherwise. We compute the share of respondents that received social security by age, grouping those aged 75 and up into one category. The bar graph shows the share of drivers reporting social security income by age.

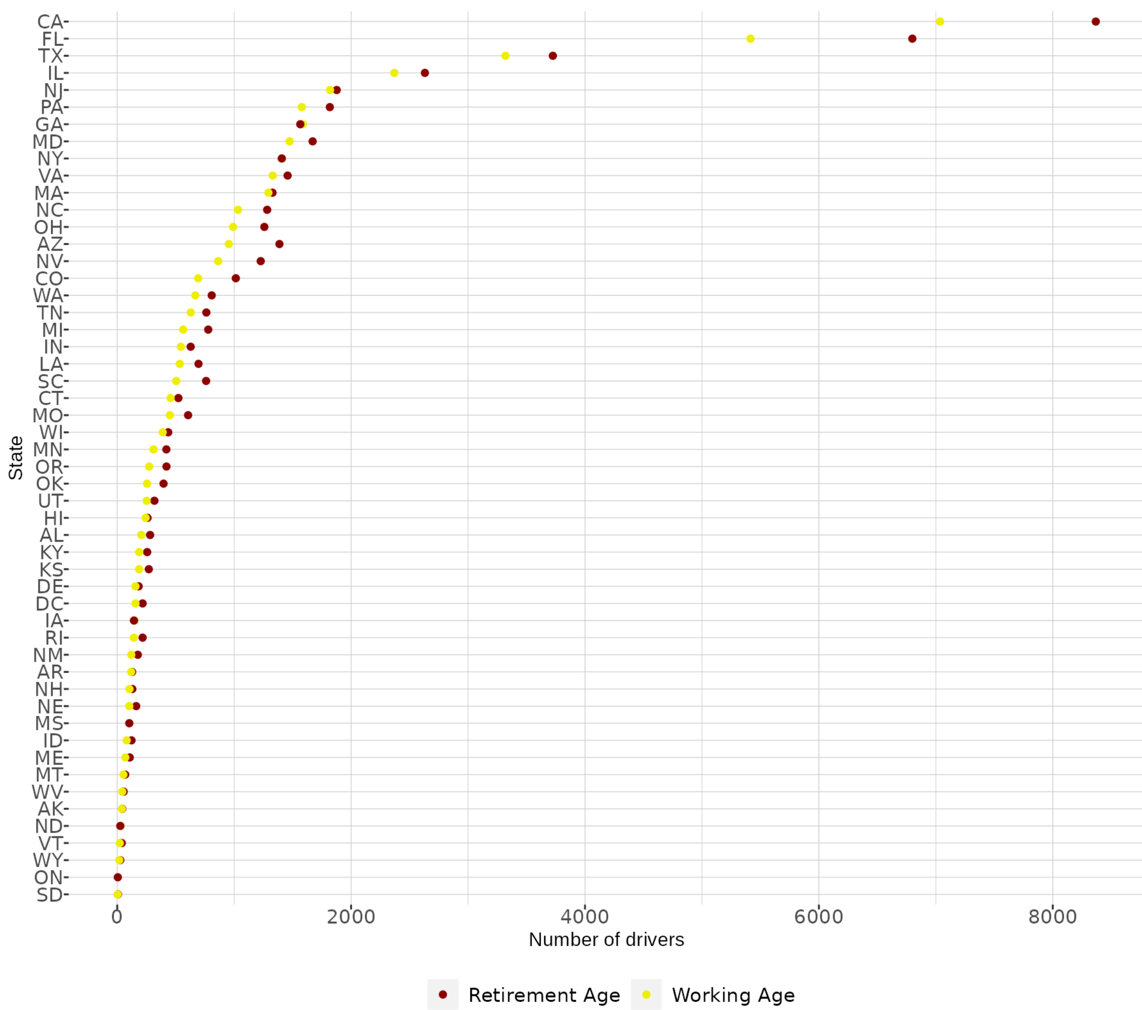
Figure A19: Total Driver Income by Source and by Age, ACS 2019-2019.



Data is from the the American Community Survey (ACS) for 2018 and 2019. We restrict the sample to survey respondents aged 62 and up, with occupation code corresponding to taxi drivers or shuttle drivers and chauffeurs. For each source of income category and total income, we compute the average amount, in dollars, of income received by age, grouping those aged 75 and up into one category. The stacked bar graph shows the average income breakdown across categories by age, in dollars.

B Supplemental Material

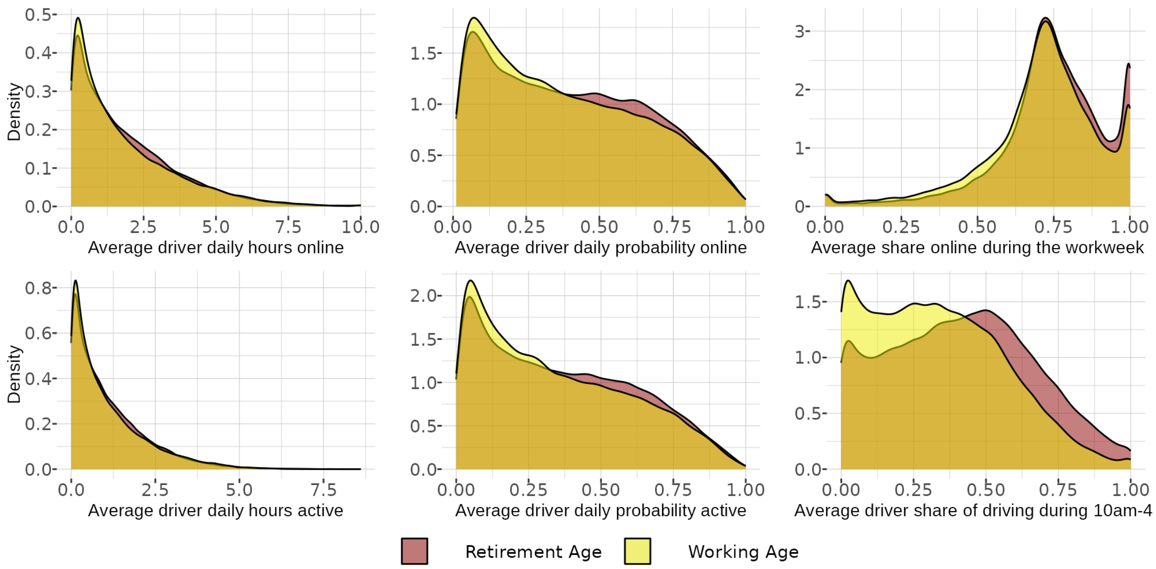
Figure A20: State of residence of drivers



B.1 Finite-sample inference

As an alternative to multiple hypothesis-corrected asymptotic p-value, we present below a large-scale randomization-inference procedure (Efron, 2013) in the spirit of Mu, Rubin, and Zou (2022). For each of the 90,961 drivers i in our sample, we employ the following algorithm: (1) Randomly draw a sample of 1,000 placebo drivers, indexed by j, from

Figure A21: Driving behaviors, Retirement Age vs Working Age

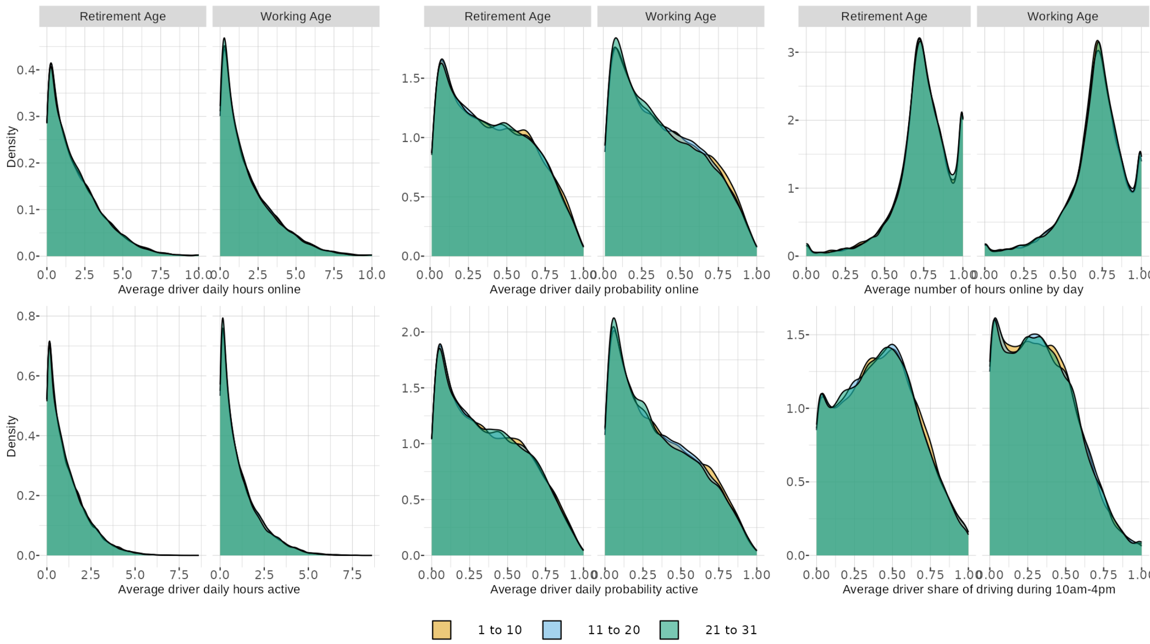


Distribution of driver-level average driving behaviors, by estimation sample.

different cohorts. For example, if Driver i was born between the 11th and the 20th day of a month, we randomly create a placebo sample of drivers with birthdays between the 1st and the 10th, or the 20th and the 31st. (2) Randomly assign each placebo driver to a cohort other their own. To illustrate, if Driver i was born between the 11th and the 20th and randomly chosen placebo Driver j was born between the 1st and the 10th, Driver j has a 50% chance of being assigned a placebo day of birth between the 11th and the 20th and a 50% chance of being assigned a placebo day of birth between the 21st and the 30th. (3) Compute the placebo seven-day Social Security–treatment window, D, for this placebo driver. (4) Estimate equation 6 for placebo driver j. (5) Repeat steps 2, 3, and 4 for each of the 999 other placebo drivers. (6) Compute, among the 1,000 placebo drivers, the randomization p-value as the number of estimated coefficients on D that are larger (in absolute value) than the real effect: $\hat{p}_i = \frac{\sum_{j=1}^{1000} I(\widehat{\tau}_j > \widehat{\tau}_i)}{1000}$. Here, $I(\cdot)$ is an indicator function equal to 1 if $(\widehat{\tau}_j > \widehat{\tau}_i)$—the null hypothesis of no effect.

By construction, we would expect a uniform distribution of randomization-inference

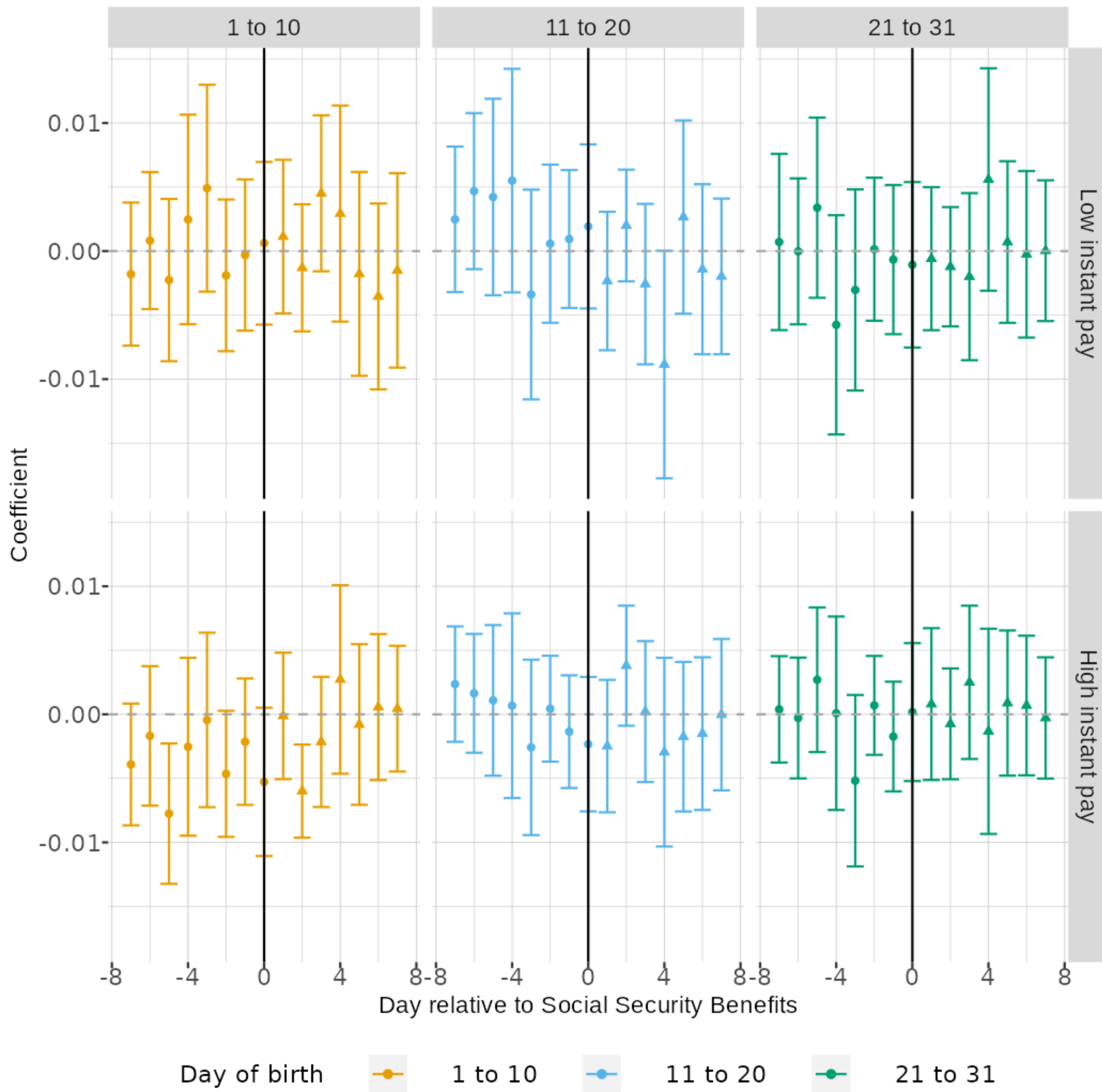
Figure A22: Driving behaviors, Retirement Age vs Working Age, by day of birth cohort



Distribution of driver-level average driving behaviors, by estimation sample and day of birth cohort.

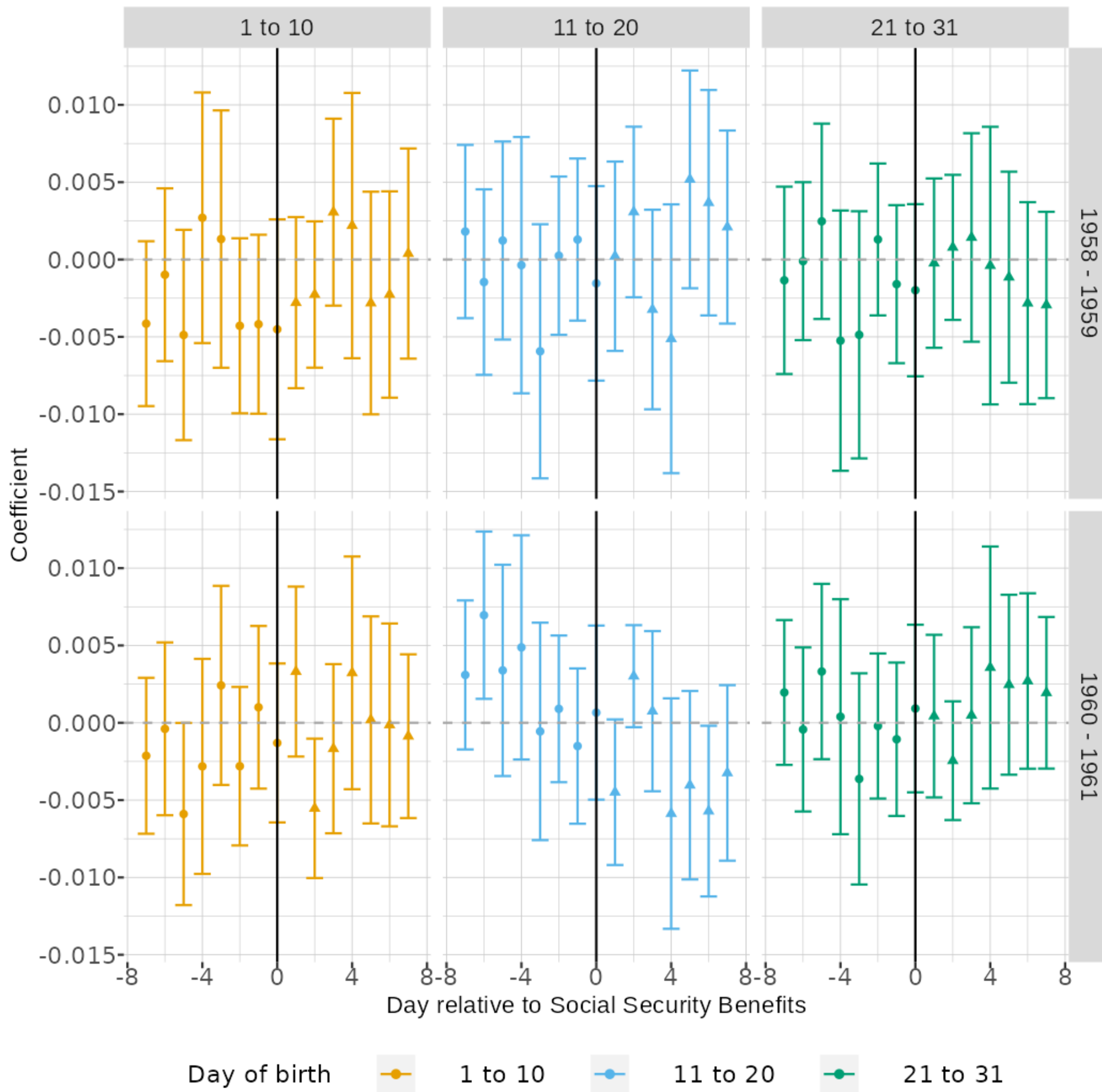
p-values. In particular, we would expect no differences between the p-values for positive and negative individual treatment effects. Figure A28 presents the histogram of the randomization p-values for each of the three cohorts for both samples. The p-values corresponding to positive and negative individual treatment effects are colored blue and red, respectively.

Figure A23: Working Age, instant pay heterogeneity, probability online



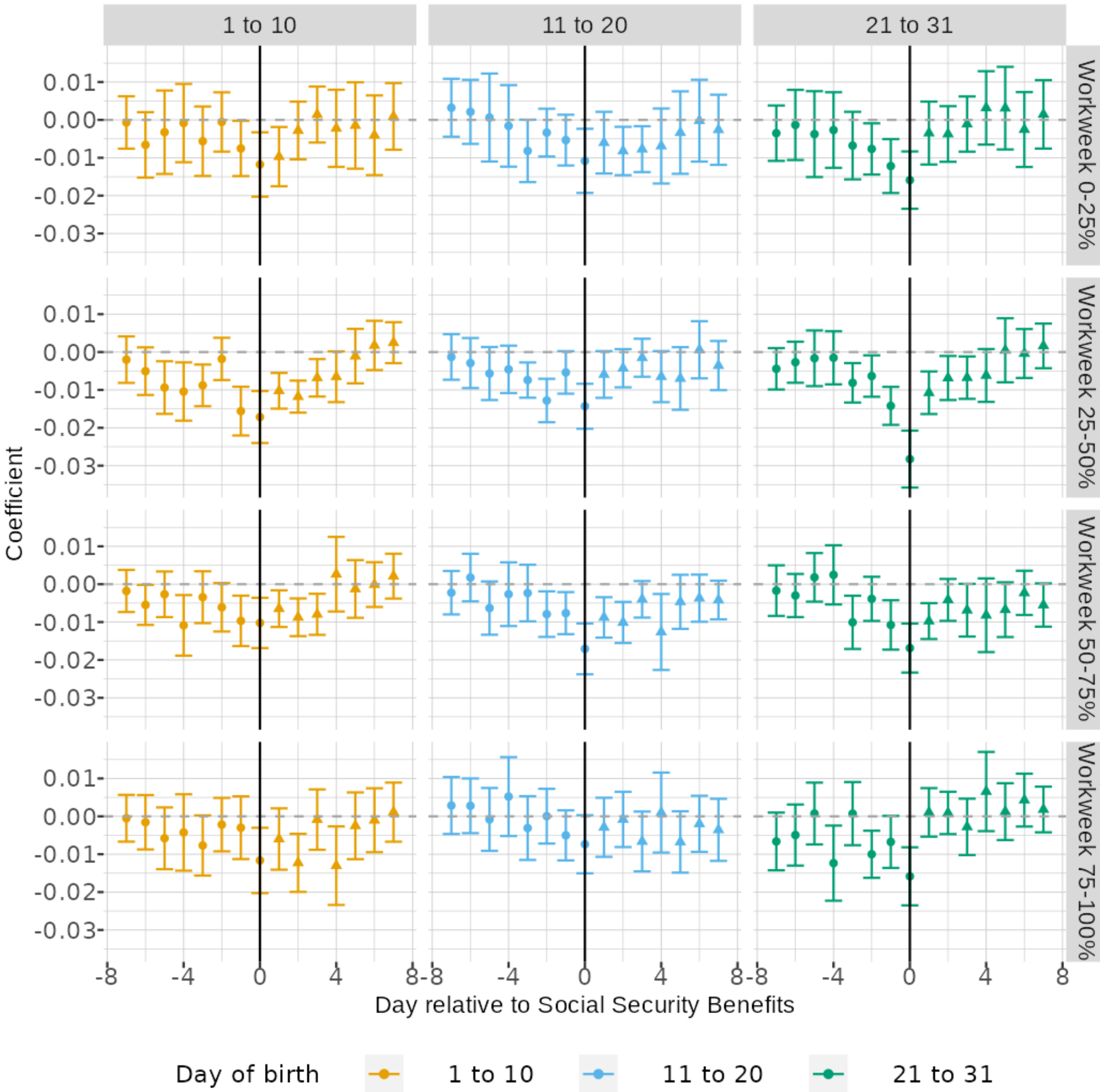
The figure depicts event-study estimates of the impact of days relative to benefit receipt on Probability Online, highlighting heterogeneity by Instant Page usage for the Working Age group. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels.

Figure A24: Working Age drivers, probability online, heterogeneity by age group



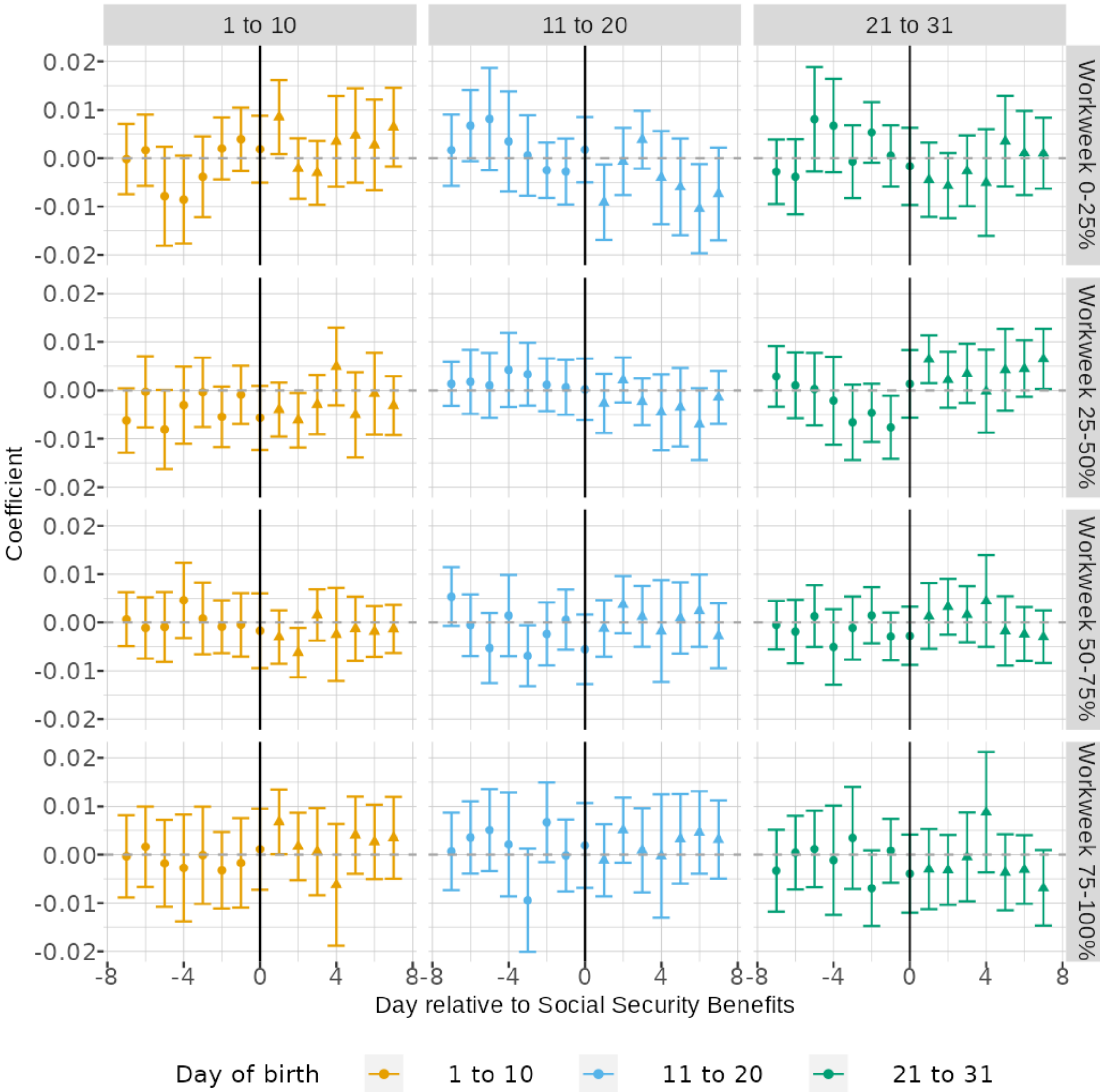
The figure depicts event-study estimates of the impact of days relative to benefit receipt on Probability Online, highlighting heterogeneity by age for the Working Age group. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels.

Figure A25: Retirement Age, workweek heterogeneity, hours online



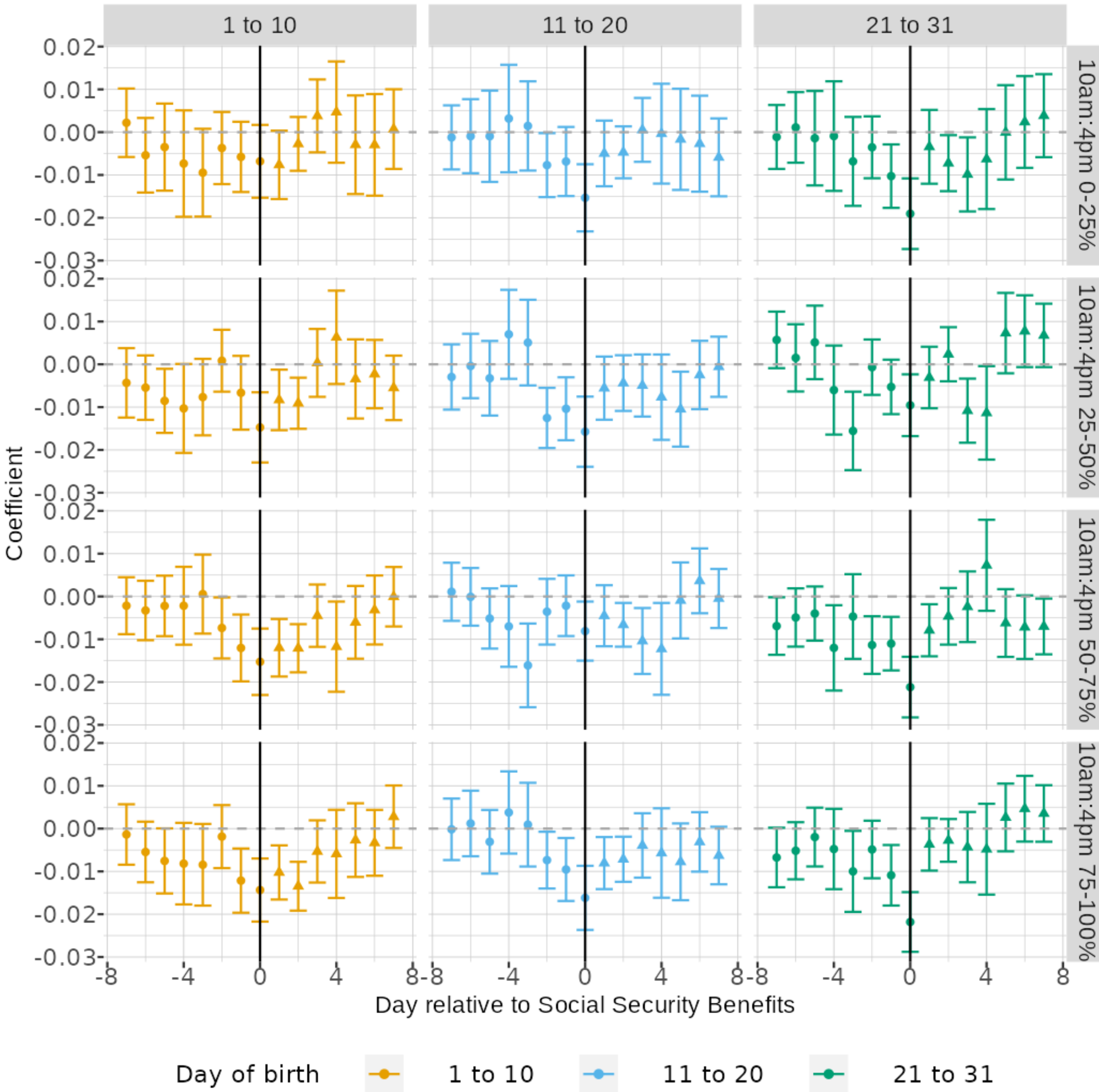
The figure depicts event-study estimates of the impact of days relative to benefit receipt on Hours Online, highlighting heterogeneity by workweek behavior for the Retirement Age sample. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels. To construct the quartiles of workweek behavior, we compute the share of time a driver is online during the workweek (as opposed to during the weekend) at the driver level, within the Retirement Age sample.

Figure A26: Working Age, workweek heterogeneity, hours online



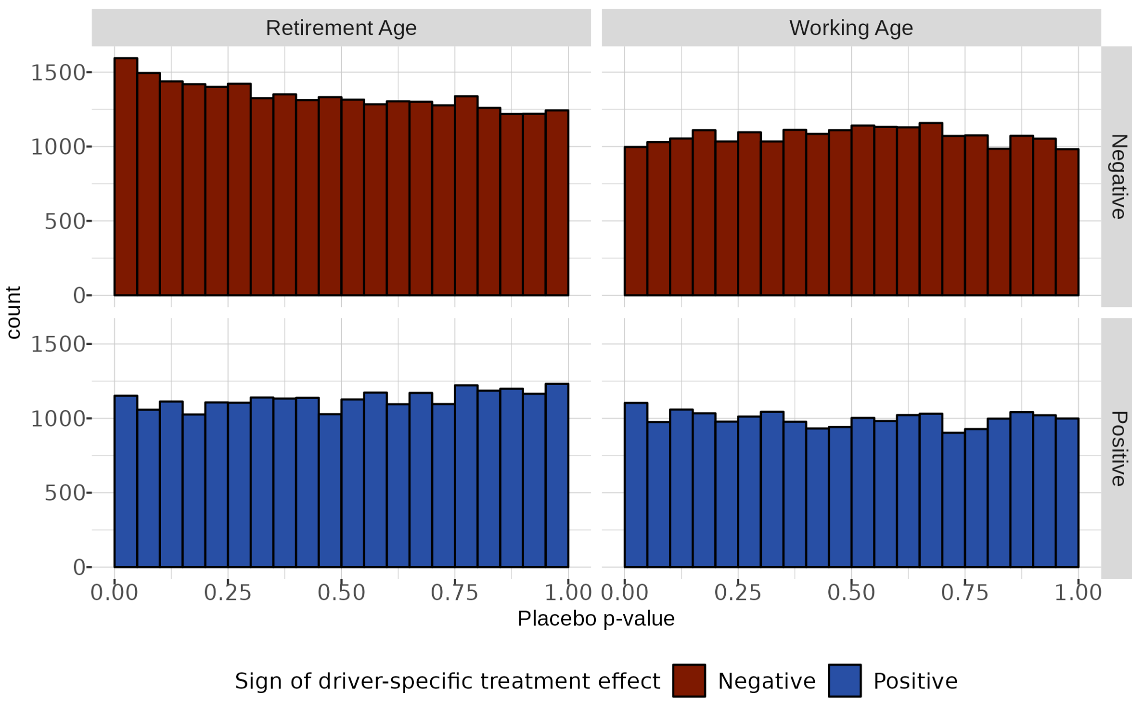
The figure depicts event-study estimates of the impact of days relative to benefit receipt on Hours Online, highlighting heterogeneity by workweek behavior for the Working Age sample. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels. To construct the quartiles of workweek behavior, we compute the share of time a driver is online during the workweek (as opposed to during the weekend) at the driver level, within the Working Age sample.

Figure A27: Retirement Age, workday heterogeneity, probability online



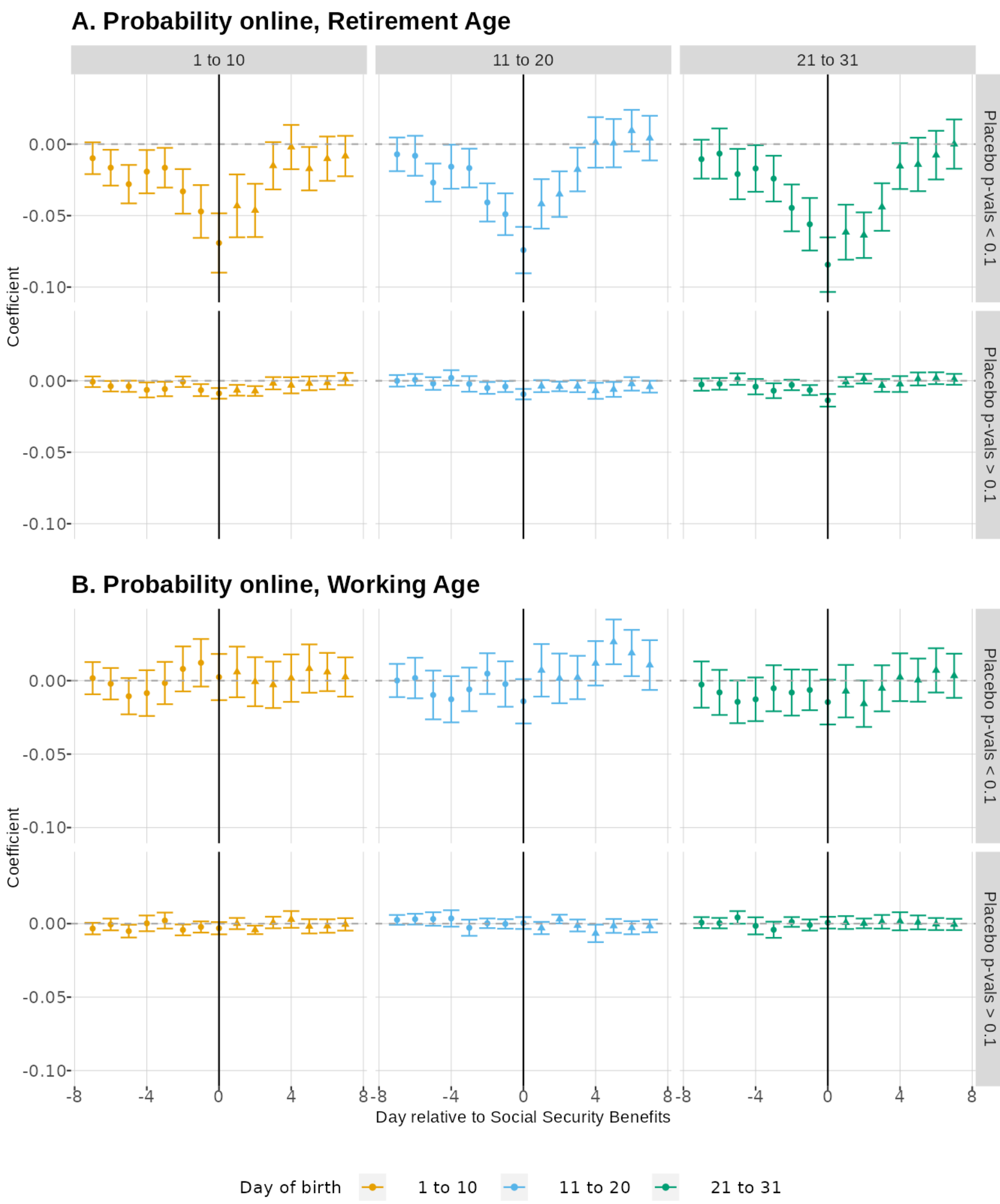
The figure depicts event-study estimates of the impact of days relative to benefit receipt on Probability Online, highlighting heterogeneity by workday behavior for the Retirement Age sample. The event studies are estimated separately for the before and after periods, resulting in six different regressions following equations 3 for each sample. The reference dummies are -8 for the before period and $+8$ for the after period. The control groups are within-sample drivers with a day of birth more than seven days away from their treatment date, following Figure 3. The error bars represent 95% confidence intervals with standard errors clustered at the year and individual levels. To construct the quartiles of workday behavior, we compute the share of time a driver is online during the 10am - 4pm window at the driver level, within the Retirement Age sample.

Figure A28: Histogram of the individual-treatment-effect randomization p-values



Individual-treatment-effect randomization-inference p-values using probability online as the outcome. For each driver-level treatment effect i estimated with regression 6, we randomly choose 1,000 drivers p in the same driver sample (Retirement Age or Working Age) and in the other two day-of-birth cohorts. We then randomly assign one of the two wrong treatment dates following a Bernoulli draw. For each driver i, the randomized p-value is computed as the share of the 1,000 placebo treatment effects p that are larger (in absolute value) than the estimated ITE for i. The p-values are plotted separately for positive (in blue) and negative (in red) treatment effects.

Figure A29: Paired-event-study estimates, heterogeneity by driver-level response, randomization p-values



Individual treatment effect randomization inference p-values using "hours online" as the outcome. For each driver-level treatment effect i estimated with regression 6, we randomly choose 1,000 drivers p in the same driver sample (Retirement Age or Working Age) and in the other two day of birth cohorts. We then randomly assign one of the two "wrong" treatment dates following a Bernouilli draw. For each driver i, the randomized p-value is computed as the share of the 1,000 placebo treatment effects p that are larger (in absolute value) than the estimated ITE for i.