Social Security and the Distribution of Effective Retirement Incentives

HongSeok Kim*
Iowa State University

March 26, 2026

Abstract

Social Security converts a common payroll-tax rate into different effective tax wedges across older workers through earnings-record recomputation, claiming rules, and the Retirement Earnings Test. We study whether that heterogeneity shapes retirement behavior using the 2011–2012 U.S. payroll-tax holiday (a temporary cut from 6.2% to 4.2%) and monthly SIPP data on workers ages 62–70. Already-claimed workers respond more than unclaimed workers, and among the unclaimed, responses rise with the rate at which current earnings raise future benefits. A post-62 dynamic model calibrated to pre-reform moments shows that the common 6.2% statutory rate maps into effective wedges ranging from negative values for unclaimed workers to above 20% for claimers facing the RET before the Full Retirement Age. Payroll-tax reform for seniors should be evaluated as a distribution of effective wedges, not as a uniform rate change.

Keywords: Social Security, payroll taxes, labor supply, retirement, policy incidence.
JEL codes: H55, J14, J22, H31, H24.

*I am deeply grateful to my advisor, Juan Carlos Cordoba, and my committee members, Rajesh Singh, Quinn Weninger, Sergio Lence, and Jian Li, for their thoughtful comments and sincere feedback.

1 Introduction

Debates over Social Security reform usually summarize payroll-tax policy with the statutory rate. Near retirement, that is the wrong object. A worker who has not yet claimed can still convert current earnings into higher future benefits through earnings-record recomputation, and delaying claim can raise benefits through the claim-age schedule. A worker who has already claimed loses the delay-claim channel, may retain only attenuated recomputation, and, before the Full Retirement Age, may also face the Retirement Earnings Test. Social Security therefore turns one common payroll-tax change into different effective labor wedges across seniors.

This paper asks how much heterogeneity Social Security creates in the effective return to work when the statutory OASDI payroll tax changes, and whether that heterogeneity materially changes retirement behavior. This is a policy-incidence question. In the solvency debate, changes in the Old-Age, Survivors, and Disability Insurance (OASDI) payroll-tax rate are among the main policy instruments routinely evaluated by the Office of the Chief Actuary (Social Security Administration, Office of the Chief Actuary, 2025; Goss, 2019). Yet policy discussion still treats the statutory rate as if it were the relevant incentive. For older workers, the relevant object is the effective wedge on current work after the benefit formula, claiming rules, and the Retirement Earnings Test are taken into account. A large literature shows that Social Security affects labor supply, but it has not directly answered how one common payroll-tax reform is transformed into many incentive changes near retirement.

We answer this question using the 2011–2012 U.S. payroll-tax holiday, which temporarily lowered the employee OASDI rate from 6.2 to 4.2 percent. The reform is useful because the statutory cut was common, but its effective size was not. Using monthly SIPP data on workers ages 62–70, we compare workers who entered the reform with different predetermined institutional positions: different baseline exposure to the employee payroll tax, different claiming status, and, among workers who had not yet claimed, different rates at which additional earnings translated into future benefits under the Social Security formula.

Three findings stand out. First, workers with larger baseline payroll-tax exposure change retirement timing more during the holiday. Second, workers who had already claimed benefits respond more strongly than otherwise similar workers who had not yet claimed. Third, among workers who had not yet claimed, responses vary systematically with the strength of work-linked marginal benefit recomputation evaluated at a fixed claim age. The common message is that, for seniors, retirement responses depend not only on the statutory payroll-

tax rate, but also on how tightly current contributions are linked to future benefits.

The contribution of the paper is to turn that idea into an empirical and quantitative account of policy incidence near retirement. Rather than asking whether Social Security matters for labor supply in general, we ask how the existing U.S. system maps one common payroll-tax reform into a distribution of effective wedges across seniors. That contribution has three parts. First, the empirical design isolates predetermined institutional sources of wedge heterogeneity under a common reform. Second, the theory organizes those comparisons around a single object, the effective wedge on current work, while separating work-linked recomputation from delay-claim incentives. Third, the calibrated model quantifies which institutional channels matter for which groups: the Retirement Earnings Test matters most for low-asset early claimers, while work-linked recomputation and delay-claim incentives matter most for seniors with room to postpone claim.

We interpret these facts with a parsimonious dynamic retirement model for seniors. The model is partial equilibrium and begins at age 62, taking earlier earnings histories and wealth accumulation as given. It is strongest as a conditional post-62 mechanism model, not as a full institutional or welfare-complete retirement model. The theory proceeds in two layers because the empirical design and the policy question impose different demands on the model. The first layer is an analytical block that conditions on observed claiming status at the start of the reform and maps directly onto the reduced-form evidence. The second layer embeds that logic in a calibrated dynamic program that begins at age 62 and lets initially unclaimed workers choose when to claim.

Policy counterfactuals cannot maintain that conditioning: a permanent payroll-tax cut, the elimination of the Retirement Earnings Test, the removal of post-62 earnings-record recomputation, or the removal of delay-claim credits would each change when workers claim, invalidating any analysis that holds claiming status fixed. Calibrated to pre-reform employment moments, the quantitative model can therefore ask the counterfactual questions the reduced form cannot: how much do wedge differences across claiming states, earnings positions, recomputation incentives, and delay-claim incentives matter for the incidence of payroll-tax reform on senior labor supply?

The main message of the paper is straightforward. For seniors, a payroll-tax change is not a common incentive change. Social Security maps the same statutory reform into different effective wedges across workers, and those wedge differences materially shape retirement behavior. The missing ingredient in current policy analysis is therefore not another average retirement elasticity. It is a quantitative account of how Social Security distributes the labor-

supply distortion from each reform across seniors.

Putting in the Literature

This paper connects to three strands of the literature. First, it contributes to work on benefit linkage and labor supply. Feldstein and Samwick (1992) documented computationally that the progressive benefit formula generates a wide cross-sectional distribution of effective marginal tax rates, but they did not estimate behavioral responses to a rate change. Liebman et al. (2009) and Liebman and Luttmer (2015) show that workers respond to Social Security incentives and to their understanding of those incentives, while Gelber et al. (2022) estimate employment responses at the earnings-test kink. More recently, French et al. (2022) show that changing the contribution-benefit link affects labor supply far from retirement age. Our paper complements that work by shifting the question from whether linkage matters to how linkage shapes the incidence of a common payroll-tax reform near retirement.

Second, the paper contributes to the literature on structural retirement models with Social Security incentives. A large literature models retirement as a dynamic decision shaped by Social Security rules, option values, health, and wealth (Stock and Wise, 1990; Rust and Phelan, 1997; French, 2005; French and Jones, 2011; Coile and Gruber, 2007; Coile, 2023; Blundell et al., 2016). Our paper uses the model for a different purpose. Rather than focusing on average retirement effects of Social Security, we use it to understand why the same payroll-tax reform leads different seniors to respond differently. The model is organized around the same effective-wedge object as the reduced-form evidence.

Third, the paper contributes to the literature on mortality heterogeneity and tax wedges. The broader lifecycle literature shows that mortality risk changes labor-supply and retirement incentives (Yaari, 1965; De Nardi et al., 2006, 2016), and work on subjective survival expectations shows that horizons help explain retirement behavior (Hurd and McGarry, 1995, 2002; Hurd et al., 2004). At the same time, macro labor-supply work often summarizes tax incentives with a single wedge (Prescott, 2004; Ohanian et al., 2008; Rogerson and Wallenius, 2009; Chetty et al., 2013; Laun and Wallenius, 2016). Our paper links these ideas in a narrow way: survival matters here because it changes the value of benefit accrual, and the relevant policy object is not one senior wedge but the distribution of effective wedges created by Social Security institutions.

Roadmap

The remainder of the paper proceeds as follows. Section 2 presents the dynamic retirement model for seniors. Section 3 tests the model to data by describing the institutional setting, the data, the reduced-form evidence, and the parameterization and calibration strategy. Section 4 presents the quantitative exercises. Section 5 concludes.

2 Model

This section develops a post-62 retirement model. The object of interest is not the statutory payroll-tax rate itself but the effective return to continued work once payroll taxes, earnings-record recomputation, claiming status, the Retirement Earnings Test, and survival are combined. Workers arrive at age 62 with predetermined earnings histories and asset positions; the paper studies how they adjust work and claiming thereafter.

We proceed in two steps. First, we write a benchmark dynamic program with endogenous retirement and claiming. Second, we derive a reduced extensive-margin representation that isolates the within-period work-versus-exit margin. The reduced form conditions on observed start-of-period claim status and delivers comparative statics for claiming status, recomputation, and survival. The quantitative analysis returns to the full benchmark.

2.1 Model Environment

At the start of each age t, a worker is characterized by claiming status $g_t \in \{u, c\}$, retirement status $r_t \in \{w, r\}$, an earnings-record state e_t, fixed joint type α, predetermined asset type x, and a transitory work-capacity shock z_t. If the worker has already claimed, the state also records the locked-in claim age c_t. The joint type α bundles permanent earnings capacity and health-survival class. Conditional on (z_t, α), labor income is $y_t(z_t, \alpha)$, survival is $\sigma_t(\alpha)$, and z_{t+1} evolves according to a transition law. The asset type does not evolve after age 62; it summarizes liquidity and therefore the cost of delaying claiming. Post-62 saving is shut down.

Choices. An active unclaimed worker has four options in each period: keep working and wait to claim, keep working and claim immediately, retire and wait, or retire and claim immediately. A worker who has already claimed chooses whether to keep working or retire. A retired but unclaimed worker chooses whether to remain retired and unclaimed or to claim while retired. Retirement is absorbing in the benchmark.

Payoffs. The primitive work payoff $u_t^W(z_t, \alpha)$ collects after-tax earnings net of work disutility before any claim-specific Social Security terms. The primitive retirement payoff $u_t^R(\alpha)$ collects the non-benefit flow from non-employment. Benefits are $B(c, e)$, where c is claim age and e is the earnings-record state. Current work changes future benefits through the earnings-record update $e_{t+1} = e_t + \phi(y_t)$. Delaying claim changes future benefits through a separate channel: $B(t+1, e_t)$ versus $B(t, e_t)$. The first is the value of working holding claim age fixed; the second is the value of waiting to claim holding the earnings record fixed. Claimed workers lose the delay channel but may retain some earnings-record recomputation and, before the Full Retirement Age, face the Retirement Earnings Test on earnings above the exempt amount. Remaining unclaimed carries a waiting cost $\chi(x)$, larger for low-asset types.

The asset type x captures the liquidity cost of delaying claiming: low-asset workers find it harder to wait and are more likely to claim early. The joint type α is the main state-space reduction. Instead of separate permanent earnings, health, and longevity indices, α summarizes all three. Higher-α workers earn more, have better survival, and therefore gain more from future recomputation. The transitory shock z_t captures short-run departures from that permanent type.

2.2 Benchmark Dynamic Program

Let V_t^{WU} denote the value of an active unclaimed worker, V_t^{WC} the value of an active claimed worker, V_t^{RU} the value of a retired unclaimed worker, and V_t^{RC} the value of a retired claimed worker. Let $B(c, e)$ denote the per-period benefit flow implied by claim age c and earnings-record state e. The active unclaimed state is

$$V_t^{WU}(z_t, e_t, \alpha, x) = \max\left\{V_t^{W,\text{wait}}, V_t^{W,\text{claim}}, V_t^{R,\text{wait}}, V_t^{R,\text{claim}}\right\}. \tag{1}$$

The four branches are

$$\begin{aligned}
V_t^{W,\text{wait}}(z_t, e_t, \alpha, x) &= u_t^W(z_t, \alpha) - \chi(x) + \beta\sigma_t(\alpha)\mathbb{E}_t\left[V_{t+1}^{WU}(z_{t+1}, e_t + \phi(y_t), \alpha, x)\right], \\
V_t^{W,\text{claim}}(z_t, e_t, \alpha, x) &= u_t^{WC}(z_t, t, e_t, \alpha)\,\beta\sigma_t(\alpha)\mathbb{E}_t\left[V_{t+1}^{WC}(z_{t+1}, t, e_t + \phi(y_t), \alpha)\right], \\
V_t^{R,\text{wait}}(e_t, \alpha, x) &= u_t^R(\alpha) - \chi(x) + \beta\sigma_t(\alpha)V_{t+1}^{RU}(e_t, \alpha, x), \\
V_t^{R,\text{claim}}(e_t, \alpha, x) &= u_t^R(\alpha) + B(t, e_t)\beta\sigma_t(\alpha)V_{t+1}^{RC}(t, e_t, \alpha).
\end{aligned}$$

Remaining unclaimed carries the waiting cost $\chi(x)$. Remaining employed while unclaimed updates the earnings record from e_t to $e_t + \phi(y_t)$. Waiting to claim preserves the delay channel: the worker moves from $B(t, e_t)$ to $B(t+1, e_t)$ whether she works or retires. Claiming

eliminates the delay channel but not necessarily the earnings-record channel, because later earnings can still alter benefits through recomputation.

Claimed state. Once a worker has claimed, the dynamic problem simplifies to a work-versus-retire choice:

$$V_t^{WC}(z_t, c, e_t, \alpha) = \max\Big\{ u_t^{WC}(z_t, c, e_t, \alpha) + \beta\sigma_t(\alpha)\mathbb{E}_t\big[V_{t+1}^{WC}(z_{t+1}, c, e_t + \phi(y_t), \alpha)\big] , \\ V_t^{RC}(c, e_t, \alpha)\Big\}. \tag{2}$$

The claimed-worker flow payoff is

$$u_t^{WC}(z_t, c, e_t, \alpha) \equiv u_t^{W}(z_t, \alpha) + B(c, e_t) - \theta_{\text{liq}} \max\{y_t(z_t, \alpha) - \bar{y}_t, 0\}\mathbf{1}\{a_t < \text{FRA}\}.$$

A claimed worker receives the work payoff, the benefit flow, and any RET penalty. The parameter θ_{liq} is not the statutory withholding rate. It is a net current-value wedge that compresses the withholding schedule and the later SSA benefit readjustment into a single object for the labor-supply decision.

Retired states. Retirement is absorbing. A retired unclaimed worker chooses whether to continue waiting or to claim while retired:

$$V_t^{RU}(e_t, \alpha, x) = \max\Big\{ u_t^{R}(\alpha) - \chi(x) + \beta\sigma_t(\alpha)V_{t+1}^{RU}(e_t, \alpha, x), \\ u_t^{R}(\alpha) + B(t, e_t) + \beta\sigma_t(\alpha)V_{t+1}^{RC}(t, e_t, \alpha)\Big\},$$

while a retired claimed worker simply receives benefits going forward:

$$V_t^{RC}(c, e_t, \alpha) = u_t^{R}(\alpha) + B(c, e_t) + \beta\sigma_t(\alpha)V_{t+1}^{RC}(c, e_t, \alpha).$$

Here $\mathbb{E}_t[\cdot]$ integrates over next period's transitory shock z_{t+1} conditional on current (z_t, α, r_t). Once the worker has claimed, the asset type x no longer matters because the waiting cost is zero.

Where work-linked benefits enter. For the labor-supply margin, the relevant object is the value of current work holding claim age fixed:

$$\Delta_t^{\text{work}}(c, e_t, y_t) \equiv B(c, e_t + \phi(y_t)) - B(c, e_t). \tag{3}$$

By contrast, the value of delaying claim while holding the earnings record fixed is

$$\Delta_t^{\text{delay}}(e_t) \equiv B(t+1, e_t) - B(t, e_t).$$

The object $\lambda_0(\alpha)$ is the derivative of Δ_t^{work} with respect to current earnings at a fixed reference claim age. The joint type α determines the marginal PIA bracket; the earnings-record state e_t determines whether current earnings enter the top-35 set. The exact object is time varying through both (c, e_t). In the reduced form below, we approximate $\lambda_u(t, \alpha) \approx \lambda_0(\alpha)$ and $0 \leq \lambda_c(t, \alpha) \leq \lambda_u(t, \alpha)$, where claimed workers may retain some recomputation. The delay object Δ_t^{delay} belongs in the claiming margin, not the labor wedge.

Why this benchmark is the right one. This benchmark contains the margins that matter for the paper's question. Claiming changes the effective return to work by eliminating the delay channel, attenuating recomputation, and before FRA exposing workers to the RET. Retirement is absorbing. Asset heterogeneity changes the value of waiting to claim. These ingredients are sufficient for the post-62 exercise without a full lifecycle savings model.

2.3 Derived Extensive-Margin Representation

The benchmark above is the full economic environment. For the theoretical results, we isolate the within-period work-versus-exit margin for a worker whose claiming status is observed at the start of the period. This is the margin the empirical design exploits.

Conditioning on start-of-period claim status $g \in \{u, c\}$ does not assume away selection into claiming. Earlier claiming choices shape group composition and therefore the group-specific survivor densities below. But conditional on g, the continuation threshold is pinned down by two objects: the continuation term $A_g(t, \alpha)$, which reflects Social Security rules on the work margin, and the dynamic continuation premium $\Delta_g(t, \tau, \alpha)$, which captures the option value of future work.

Three further reductions deliver a tractable stopping problem. Post-62 saving is shut down, so asset differences enter through reduced-form continuation values. All permanent heterogeneity is summarized by α. Retirement is absorbing. The residual transitory work-cost component is ν_t. The objects below are reduced-form summaries of the benchmark Bellman, not new primitives.

2.3.1 Mapping from the benchmark Bellman

Consider a worker of type α with claiming status $g \in \{c, u\}$. Time is discrete: $t \in \{0, 1, \ldots, N\}$, where $t = N$ is mandatory exit. The benchmark difference between work and retirement payoffs is collapsed into

$$w(\alpha)(1 - \tau) - \nu_t,$$

where $w(\alpha)$ is the type-specific earnings level and $\nu_t \geq 0$ is the residual work-cost shock. The retiree's claim-timing problem is solved out into $V^R(t)$. The only claim-status-specific object that remains is the effect of current work on future benefits and current RET exposure, $A_g(t, \alpha)$. The value of delaying claim is absorbed into $V^R(t)$ and does not enter the work wedge.

Define the work-specific benefit increment

$$\Delta_t^{\text{work}}(\alpha) \equiv B\Big(c^{\text{ref}}, e_t + \phi(w(\alpha))\Big) - B\Big(c^{\text{ref}}, e_t\Big), \tag{4}$$

where c^{ref} is a fixed reference claim age. The reduced form carries only this work-specific channel into $A_g(t, \alpha)$. The delay increment $B(t+1, e_t) - B(t, e_t)$ is not part of $\Delta_t^{\text{work}}(\alpha)$. Under the fixed-accrual approximation,

$$\Delta_t^{\text{work}}(\alpha) \approx \lambda_0(\alpha) w(\alpha).$$

2.3.2 Continuation payoff and normalized wedge

In the reduced form, α summarizes permanent earnings capacity and health-survival class, so survival from t to $t + 1$ is $\sigma_t(\alpha) \in (0, 1)$. The annuity factor at date t is

$$\mu_t(\alpha) \equiv \sum_{j=1}^{N-t} \beta^j \prod_{k=0}^{j-1} \sigma_{t+k}(\alpha), \tag{5}$$

with discount factor $\beta \in (0, 1)$. Higher-α workers have larger μ_t and therefore gain more from future benefits.

The continuation payoff. The claim-status-specific continuation term collects the present value of work-linked benefit changes and current RET exposure. For unclaimed workers,

$$A_u(t, \alpha) = \mu_t(\alpha) \Delta_t^{\text{work}}(\alpha) \approx w(\alpha) \lambda_0(\alpha) \mu_t(\alpha).$$

For claimed workers, some earnings-record recomputation may remain after claiming, so

$$A_c(t,\alpha) = w(\alpha)\lambda_c(\alpha)\mu_t(\alpha) - \theta_{\text{liq}} \max\{w(\alpha) - \bar{y}_t, 0\}\mathbf{1}\{a_t < \text{FRA}\}.$$

Using the reduced-form notation $\lambda_u(\alpha) = \lambda_0(\alpha)$ and $0 \leq \lambda_c(\alpha) \leq \lambda_u(\alpha)$, both cases can be summarized as

$$A_g(t,\alpha) \equiv \underbrace{w(\alpha)\lambda_g(\alpha)\mu_t(\alpha)}_{\text{work-linked recomputation value}} \underbrace{- \theta_{\text{liq}} \max\{w(\alpha) - \bar{y}_t, 0\}\mathbf{1}\{g = c,\ a_t < \text{FRA}\}}_{\text{RET current-value penalty for claimers before FRA}}. \tag{6}$$

$A_g(t,\alpha)$ is the only claim-status-specific term in the reduced Bellman. It captures the work-margin incentive, not the full value of Social Security.

Dividing by $w(\alpha)$ gives the normalized wedge:

$$\kappa_g(t,\alpha) \equiv \frac{A_g(t,\alpha)}{w(\alpha)} = \lambda_g(\alpha)\mu_t(\alpha) - \theta_{\text{liq}}\Big(1 - \frac{\bar{y}_t}{w(\alpha)}\Big)_+ \mathbf{1}\{g = c,\ a_t < \text{FRA}\}. \tag{7}$$

The effective payroll-tax wedge is then

$$\tau_g^{\text{eff}}(t,\alpha) \equiv \tau - \kappa_g(t,\alpha). \tag{8}$$

The reduced problem differs across claiming states only through κ_g. For an unclaimed worker, κ_u is large, so τ_u^{eff} can fall below the statutory rate or turn negative. For a claimed worker below FRA, the RET adds to the burden, pushing τ_c^{eff} above the statutory rate.

The difference in normalized wedges across claiming status is

$$\kappa_u(t,\alpha) - \kappa_c(t,\alpha) = \big[\lambda_u(\alpha) - \lambda_c(\alpha)\big]\mu_t(\alpha) + \theta_{\text{liq}}\Big(1 - \frac{\bar{y}_t}{w(\alpha)}\Big)_+ \mathbf{1}\{a_t < \text{FRA}\}. \tag{9}$$

Before FRA, the gap has two parts: a recomputation advantage for unclaimed workers and an RET penalty for claimed workers. After FRA, the RET drops out and the gap reduces to $\big[\lambda_u(\alpha) - \lambda_c(\alpha)\big]\mu_t(\alpha)$.

From individuals to the population. Let $m_t(s)$ be the mass of active workers at age t in state s. The cross-sectional distribution of effective wedges is

$$F_t^{\tau}(x) \equiv \frac{\sum_s m_t(s)\,\mathbf{1}\{\tau^{\text{eff}}(s) \leq x\}}{\sum_s m_t(s)}. \tag{10}$$

Even though the statutory rate is a single number, claiming rules, recomputation, and the RET spread the effective wedge across a wide range.

Flow payoff from continued work. The continuation branch of the reduced Bellman is

$$\Pi_g(t, \nu_t, \alpha) \equiv w(\alpha)(1-\tau) + A_g(t, \alpha) - \nu_t = w(\alpha)\big(1 - \tau + \kappa_g(t, \alpha)\big) - \nu_t, \tag{11}$$

The term $1 - \tau + \kappa_g$ is the effective share of each dollar that the worker keeps, so Π_g is the effective wage minus the cost of showing up. For notational convenience, write $w \equiv w(\alpha)$ and $\lambda_0 \equiv \lambda_0(\alpha)$ when no confusion arises.

Assumption 1. *Conditional on α, the primitive payoff and transition objects are common across groups except for $A_g(t, \alpha)$. Differences in the distribution of latent work costs may arise through past selection into employment and through earlier claiming choices. The comparative statics below are conditional on those start-of-period group distributions.*

Irreversible Retirement. In the reduced representation, once a worker exits at date t, the worker remains non-employed for all $s > t$. Let $V^R(t)$ denote the common retirement value after the retiree's claim-timing problem has already been optimized out. Then

$$V_c^R(t) = V_u^R(t) = V^R(t). \tag{12}$$

In the benchmark, retired claimed and retired unclaimed workers are distinct states. In the reduced problem, the retired claiming margin is solved out into $V^R(t)$ because it is not a current-work incentive. The marginal value of current work appears in Π_g through A_g, not again in V^R.

2.4 Continuation thresholds and monotone comparative statics

Let $V_g^W(t, \nu_t, \alpha; \tau)$ denote the value of an employed worker with claiming status g at date t. For $t \in \{0, \dots, N-1\}$,

$$V_g^W(t, \nu_t, \alpha; \tau) = \max\left\{ \underbrace{V^R(t)}_{\text{exit now}}, \quad \underbrace{\Pi_g(t, \nu_t, \alpha)}_{\text{current flow payoff}} + \underbrace{\beta \sigma_t(\alpha)\, \mathbb{E}_t\big[V_g^W(t{+}1, \nu_{t+1}, \alpha; \tau)\big]}_{\text{discounted continuation value}} \right\}, \tag{13}$$

with terminal condition

$$V_g^W(N, \nu_N, \alpha; \tau) = V^R(N).$$

The option value of continued work is

$$\mathcal{O}_g(t, \nu_t, \alpha; \tau) \equiv \Pi_g(t, \nu_t, \alpha) + \beta \sigma_t(\alpha)\, \mathbb{E}_t \left[V_g^W(t+1, \nu_{t+1}, \alpha; \tau) \right] - V^R(t).$$

The worker continues if and only if $\mathcal{O}_g(t, \nu_t, \alpha; \tau) \geq 0$. Because the current work-cost state enters additively, the continuation rule is threshold based:

$$\nu_t \leq \nu_g^*(t, \tau, \alpha) \equiv w(1-\tau) + A_g(t, \alpha) + \Delta_g(t, \tau, \alpha), \tag{14}$$

where

$$\Delta_g(t, \tau, \alpha) \equiv \beta \sigma_t(\alpha)\, \mathbb{E}_t \left[V_g^W(t+1, \nu_{t+1}, \alpha; \tau) \right] - V^R(t) \tag{15}$$

is the dynamic continuation premium, separating the one-period work incentive from the option value of future work.

Proposition 1. *Under Assumption 1, suppose $\sigma_t(\alpha)$ is weakly increasing in α for every t, and that both $A_g(t, \alpha)$ and the gap $A_u(t, \alpha) - A_c(t, \alpha)$ are weakly increasing in α, with*

$$A_u(t, \alpha) > A_c(t, \alpha) \qquad \textit{for every } (t, \alpha).$$

Then:

1. *For every $t \in \{0, \ldots, N-1\}$ and every α,*

$$\nu_u^*(t, \tau, \alpha) > \nu_c^*(t, \tau, \alpha). \tag{16}$$

   Unclaimed workers tolerate higher work disutility before exiting.

2. *For every $\alpha' \geq \alpha$ and each group $g \in \{c, u\}$,*

$$\nu_g^*(t, \tau, \alpha') \geq \nu_g^*(t, \tau, \alpha), \tag{17}$$

   and the threshold gap is weakly increasing in joint type:

$$\left[\nu_u^*(t, \tau, \alpha') - \nu_c^*(t, \tau, \alpha') \right] \geq \left[\nu_u^*(t, \tau, \alpha) - \nu_c^*(t, \tau, \alpha) \right]. \tag{18}$$

3. *For every $\tau' \leq \tau$ and each $g \in \{c, u\}$,*

$$\nu_g^*(t, \tau', \alpha) \geq \nu_g^*(t, \tau, \alpha). \tag{19}$$

Thus a payroll tax cut weakly raises the continuation threshold, making workers more willing to stay employed.

Proof. See Appendix A. □

The type α affects the threshold through $A_g(t,\alpha)$ and $\Delta_g(t,\tau,\alpha)$. Because $\lambda_u(\alpha) \geq \lambda_c(\alpha)$, better survival raises the unclaimed threshold more than the claimed threshold. This is the economic intuition behind the survival interaction below. Whenever $\Delta_g(t,\tau,\alpha)$ is differentiable in τ, the threshold sensitivity is

$$\frac{\partial \nu_g^*}{\partial \tau} = -w + \frac{\partial \Delta_g}{\partial \tau} \leq 0. \tag{20}$$

Because $\partial \nu_g^*/\partial \tau \leq 0$, a tax cut $d\tau < 0$ raises the continuation threshold by $|d\tau| \cdot w$ through the direct flow effect, plus an additional upward shift through the continuation premium.

2.5 Two-Channel Decomposition

Let $F_g(\cdot \mid t)$ denote the distribution of latent work costs among workers employed at the start of period t, with density $f_g(\cdot \mid t)$. This distribution may differ across groups because past decisions select different workers into continued employment. Let $C_t = 1$ denote continuation from t into $t+1$.

Holding fixed the start-of-period survivor distribution $F_g(\cdot \mid t)$, the period-t continuation probability is

$$P_g(C_t = 1 \mid \tau, \alpha) = F_g\big(\nu_g^*(t,\tau,\alpha) \mid t\big). \tag{21}$$

Differentiating (21), we get

$$\frac{\partial P_g}{\partial \tau} = f_g\big(\nu_g^*(t,\tau,\alpha) \mid t\big) \cdot \frac{\partial \nu_g^*}{\partial \tau}. \tag{22}$$

Proposition 2 (Two-channel decomposition of the extensive-margin differential)**.** *The differential response of continuation probabilities to a tax change is*

$$\frac{\partial P_c}{\partial \tau} - \frac{\partial P_u}{\partial \tau} = \underbrace{w\Big[f_u(\nu_u^*) - f_c(\nu_c^*)\Big]}_{\textit{density channel}} + \underbrace{f_c(\nu_c^*)\frac{\partial \Delta_c}{\partial \tau} - f_u(\nu_u^*)\frac{\partial \Delta_u}{\partial \tau}}_{\textit{continuation-value channel}}, \tag{23}$$

where (t,τ,α) arguments are suppressed.

Proof. Substitute (20) into (22) and rearrange. □

The density channel reflects where each group's threshold sits on the work-cost distribution: the group at a higher density gains more marginal continuers from a common threshold shift. The continuation-value channel reflects how much the tax cut raises each group's option value of future work.

Lemma 1. *For every g, t, and α,*

$$0 \leq -\frac{\partial \Delta_g(t,\tau,\alpha)}{\partial \tau} \leq w \sum_{j=1}^{N-t} \beta^j \prod_{k=0}^{j-1} \sigma_{t+k}(\alpha). \tag{24}$$

Hence a tax cut weakly raises the continuation premium (the option value of staying employed), and this gain is bounded above by the discounted, survival-weighted benefit of saving one wage unit of tax in each remaining period.

Proof. The lower bound follows from Proposition 1(iii): a lower tax weakly raises all future continuation values, so $\partial \Delta_g / \partial \tau \leq 0$. For the upper bound, compare the value effect of a marginal tax cut to the counterfactual path in which the worker continues working in every future period. The tax cut can raise the flow payoff by at most w in any future period, and future gains are discounted and survival weighted, which gives (24). □

For workers near retirement, $N - t$ is short, so the continuation-value channel is small relative to the density channel. The payroll tax holiday operates mainly through the immediate wage gain, not through changed option values.

Suppose $F_c(\cdot \mid t) = F_u(\cdot \mid t) \equiv F(\cdot \mid t)$ with unimodal density $f(\cdot \mid t)$ and mode $\bar{\nu}(t)$. Among workers still employed at ages 62+, most have low work costs — those with high costs have already exited. The mode therefore sits below both thresholds, placing ν_c^* and ν_u^* on the declining right tail. Because $\nu_c^* < \nu_u^*$ (Proposition 1(i)) and the density is decreasing,

$$f(\nu_c^* \mid t) > f(\nu_u^* \mid t). \tag{25}$$

Claimed workers have a smaller recomputation cushion, so their exit margin sits closer to where the remaining workers are concentrated. A tax cut therefore moves more claimed workers across the boundary.

Higher survival raises $\mu_t(\alpha)$ and shifts ν_u^* further right through $w\lambda_u\mu_t(\alpha)$. Because $\lambda_c < \lambda_u$, the claimed threshold moves less, widening the density gap $f(\nu_c^*) - f(\nu_u^*)$ and strengthening the differential response.

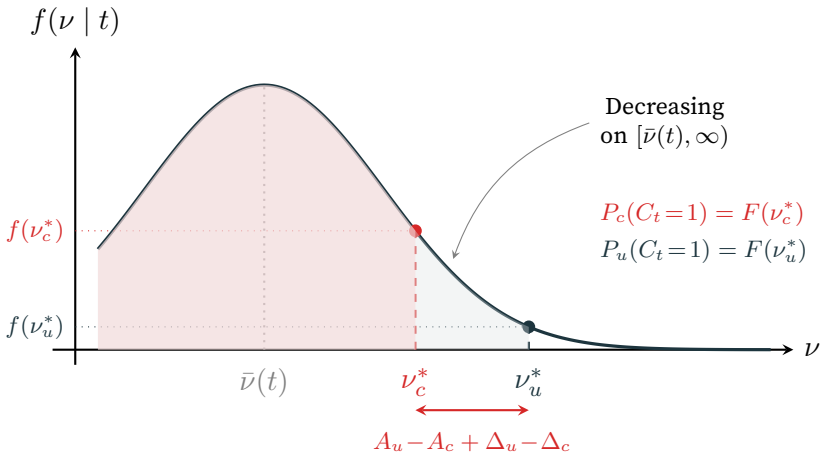


Figure 1: Continuation thresholds and the density condition for older workers. The survivor density $f(\nu \mid t)$ among workers still employed at ages 62+ peaks at a low mode $\bar{\nu}(t)$, reflecting selection: most remaining workers have low work costs. Both thresholds lie above the mode on the declining right tail. Because $\nu_c^* < \nu_u^*$ (Proposition 1(i)) and the density is decreasing, $f(\nu_c^*) > f(\nu_u^*)$: more claimed workers sit near the exit margin. A tax cut $d\tau < 0$ shifts both thresholds rightward by $w\,|d\tau|$, and the density channel implies that claimed workers gain more marginal continuers, producing a larger decline in their exit hazard. Shaded areas represent continuation probabilities conditional on employment at the start of period t.

2.5.1 Exit hazard and the empirical prediction

Define the period-t exit hazard, conditional on being employed and alive at the start of the period, as

$$h_g(t \mid \tau, \alpha) \;\equiv\; 1 - F_g\big(\nu_g^*(t, \tau, \alpha) \;\big|\; t\big) \;=\; 1 - P_g(C_t = 1 \mid \tau, \alpha). \tag{26}$$

Since a tax cut raises the continuation threshold (Proposition 1(iii)) and therefore increases the continuation probability, the exit hazard is weakly decreasing in the size of the cut for both groups. During the payroll tax holiday, both claimed and unclaimed workers face lower exit hazards than under the baseline tax rate.

Proposition 3. *A tax cut $d\tau < 0$ lowers both groups' exit hazards. Claimed workers' hazard falls more whenever the density channel dominates the continuation-value channel. Let*

$$B_t(\alpha) \equiv w \sum_{j=1}^{N-t} \beta^j \prod_{k=0}^{j-1} \sigma_{t+k}(\alpha)$$

denote the maximum sensitivity of the continuation premium to the tax (Lemma 1). If the density

gap exceeds this bound,

$$\underbrace{w\Big[f_c(\nu_c^* \mid t) - f_u(\nu_u^* \mid t)\Big]}_{\textit{density-channel advantage}} > \underbrace{f_u(\nu_u^* \mid t)\, B_t(\alpha)}_{\textit{max. continuation-value distortion}}, \tag{27}$$

then for any $d\tau < 0$*,*

$$\frac{\partial h_c}{\partial \tau} > \frac{\partial h_u}{\partial \tau} \geq 0 \qquad \Longrightarrow \qquad dh_c < dh_u \leq 0. \tag{28}$$

Both groups' exit hazards decline, but claimed workers' decline is larger. Because $B_t(\alpha)$ *shrinks as* $N - t$ *shortens, the sufficient condition is easy to satisfy near retirement.*

Proof. By definition, $\partial h_g/\partial \tau = -\partial P_g/\partial \tau$. From Lemma 1, $\partial \Delta_u/\partial \tau \geq -B_t(\alpha)$ and $-\partial \Delta_c/\partial \tau \geq 0$. Proposition 2 gives

$$\begin{aligned}\frac{\partial P_c}{\partial \tau} - \frac{\partial P_u}{\partial \tau} &\leq w\Big[f_u(\nu_u^*) - f_c(\nu_c^*)\Big] - f_u(\nu_u^*)\frac{\partial \Delta_u}{\partial \tau} \\ &\leq w\Big[f_u(\nu_u^*) - f_c(\nu_c^*)\Big] + f_u(\nu_u^*)\, B_t(\alpha).\end{aligned}$$

Under the density condition $f_c(\nu_c^*) > f_u(\nu_u^*)$, condition (27) makes the right-hand side strictly negative, so $\partial P_c/\partial \tau < \partial P_u/\partial \tau$. Hence $\partial h_c/\partial \tau > \partial h_u/\partial \tau$. Multiplying by $d\tau < 0$ reverses the inequality: $dh_c < dh_u$. Non-negativity of the derivatives follows from Proposition 1(iii), so $dh_g \leq 0$. □

This proposition translates the threshold model directly into the payroll tax holiday setting: during the 2011–2012 tax cut, both groups' exit hazards fall, but claimed workers see the larger decline because the tax cut moves more of them across the margin.

2.6 Employment stock dynamics and policy-sequence interpretation

Retirement is largely irreversible: workers who exit rarely return. Our model embeds this through absorbing exit, which implies that a temporary policy shock can have persistent effects on the employment stock even after the policy reverts.

Let $\ell_g(t \mid \tau, \alpha)$ denote the probability that a worker of type (g, α) is alive and employed at the start of period t, conditional on employment at $t = 0$. With absorbing exit and no re-entry,

$$\ell_g(t+1 \mid \tau, \alpha) = \ell_g(t \mid \tau, \alpha) \cdot P_g(C_t = 1 \mid \tau, \alpha). \tag{29}$$

Proposition 4 (Tax holiday generates propagating stock gaps)**.** *Fix (g, α) and consider a temporary tax cut at date t: the baseline path keeps the tax rate at τ in every period, while the holiday path sets $\tau_t = \tau - \delta$ with $\delta > 0$ at date t and reverts to τ thereafter. Assume further that after each period the latent work-cost draw is conditionally history-independent given current (g, α, t), so workers retained by the holiday face the same post-holiday continuation process as other employed workers of the same type.*

1. *(Impact.) The tax cut raises the continuation threshold, so fewer workers exit at date t. The holiday path therefore has strictly higher employment at $t+1$ than the baseline.*

2. *(Propagation.) Because exit is irreversible, the extra workers retained at date t remain employed in all subsequent periods (until they exit for other reasons). The employment gap evolves as*

$$\ell_g^{\mathrm{hol}}(t+s \mid \alpha) - \ell_g^{\mathrm{base}}(t+s \mid \alpha) = \Big[\ell_g^{\mathrm{hol}}(t+1 \mid \alpha) - \ell_g^{\mathrm{base}}(t+1 \mid \alpha)\Big] \prod_{r=1}^{s-1} P_g(C_{t+r} = 1 \mid \tau, \alpha) : \quad (30)$$

   the initial gain is carried forward and shrinks only through subsequent exit, never through reversal.

3. *(Reversion.) Once the tax reverts to τ, continuation probabilities are the same on both paths, so the stock gap stops growing but does not disappear. If the tax overshoots to $\tau+\epsilon$ upon restoration, a negative shock hits the larger workforce retained by the holiday.*

Proof. See Appendix A. □

This proposition maps directly onto the 2011–2012 payroll tax holiday. During the holiday, the 2 percentage point tax cut raises continuation probabilities and builds an employment stock gap relative to the no-holiday counterfactual. When the tax reverts at the start of 2013, the accumulated gap persists but no longer grows. The historical episode can therefore display asymmetric employment changes across the holiday and restoration phases, even when the tax-rate movements themselves are symmetric.

2.7 Interpretation and Link to the Quantitative Benchmark

Option values versus reservation wages. In a static model, a worker exits when the net wage falls below a reservation wage w^R; the value loss is proportional to the flow surplus $w(1-\tau) - w^R$, which is near zero for the marginal worker. Survival plays no role because the comparison is intratemporal. In the dynamic model with irreversible exit, the relevant

object is the option value $\mathcal{O}_g(t, \nu_t, \alpha)$, which includes both the value of future flexibility and the survival-weighted recomputation term $\lambda_u \mu_t(\alpha)$. A worker near the exit margin may have little current surplus yet forfeit a large continuation value upon exit.

Two workers can look similar today yet have different stakes. An unclaimed worker with high survival has a large option value; a claimed worker with low survival has a small one. A tax increase that pushes both out destroys more value for the first, even if their current reservation wages are similar. The annuity factor $\mu_t(\alpha)$ drives this wedge: it enters the option value but not the static reservation wage.

Irreversibility adds a further asymmetry. A tax increase permanently destroys continuation values of exiting workers. A tax cut only temporarily retains workers who will exit later anyway. Equal-sized tax changes therefore have asymmetric value effects (Proposition 4).[1]

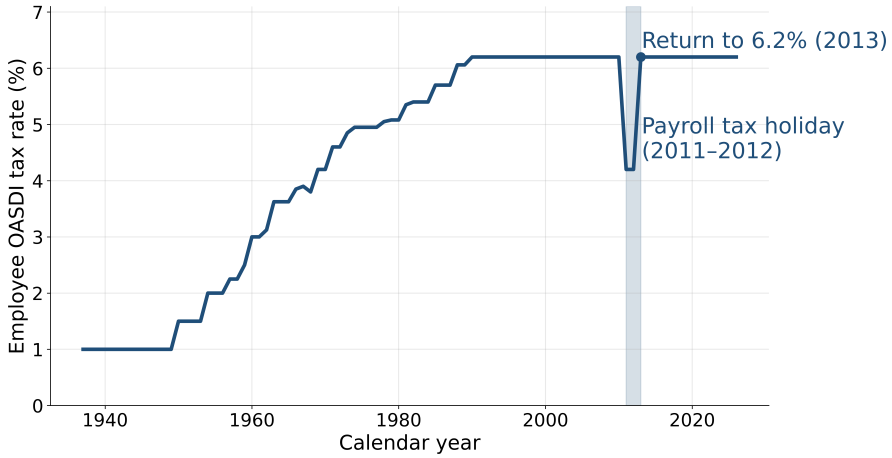
Connection to the option value of delay. The continuation premium $\Delta_g(t, \alpha)$ is analogous to the option value of delay in Stock and Wise (1990). Higher survival raises the expected number of periods over which favorable draws could arrive, increasing Δ_g. The recomputation channel $\lambda_g \mu_t(\alpha)$ provides an institutional reason for the option value to be high. The framework nests both the "waiting for a good draw" motive and the Social Security work-incentive motive.

Benchmark model versus derived representation. The benchmark is the full post-62 dynamic program with endogenous claiming, heterogeneous waiting costs, and absorbing exit. The extensive-margin system above is a conditional representation derived from it to isolate the work-margin channel.

The empirical design compares responses across workers with different observed claiming status and different $\lambda_0(\alpha)$. The derived representation mirrors this by conditioning on start-of-period claim status and mapping that variation into κ_u and κ_c. This does not treat claiming as exogenous: it isolates the work margin while taking group composition as given. The quantitative analysis returns to the full benchmark to verify that endogenous claiming does not overturn the conditional predictions.

[1] Because the reduced form shuts down post-62 savings, these are value-function comparisons, not a consumption-based welfare analysis.

Figure 2: Employee OASDI Payroll-Tax Rate Path, 2010–2013



Notes: The employee OASDI rate is 6.2% in 2010, 4.2% in 2011–2012, and 6.2% in 2013.

3 Testing the Model to Data

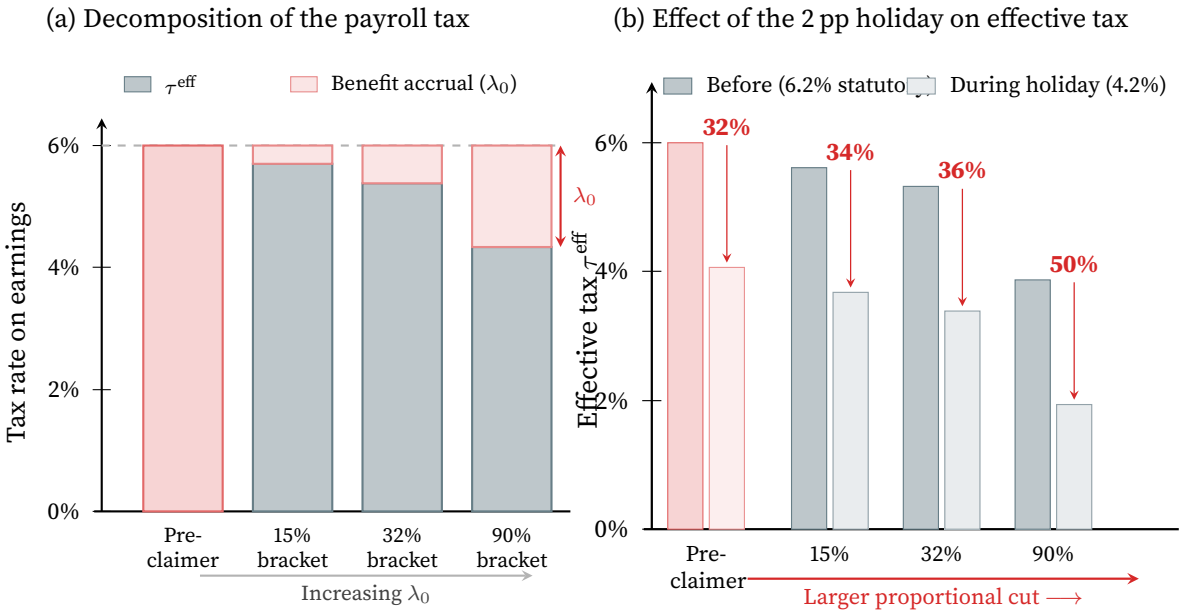
3.1 Institutional Background and Empirical Strategy

Payroll-Tax Policy Timing. The employee OASDI payroll-tax rate on covered earnings was 6.2% in 2010, fell to 4.2% in January 2011, stayed there through December 2012, and returned to 6.2% in January 2013. This creates two sharp time breaks: an expansionary shock at the start of 2011 and a reversal at the start of 2013. Because the reform was federal, all covered workers faced the same rate path at the same dates. Identification comes from differential responses across pre-policy worker types, not from geographic variation.

Claiming Status and Incentives. At ages 62–70, claiming status determines how a payroll-tax change translates into work incentives. For an unclaimed worker, part of the statutory tax is offset by future benefit recomputation, and the offset varies with position on the PIA formula. For a worker who has already claimed, the delay channel is gone, recomputation is weaker, and before FRA the RET may apply. The same 2 percentage-point cut therefore produces different proportional changes in the effective tax wedge across workers. Pre-policy claiming status is the central heterogeneity dimension in the empirical design.

Three SSA rules are relevant. First, benefits depend on the highest 35 years of covered earnings, so post-62 work can raise benefits through recomputation — even after claiming, when SSA refigures the record. Second, delaying claim raises benefits through the early-claim reduction and delayed retirement credit schedule. Third, the RET withholds benefits

Figure 3: Institutional Heterogeneity in the Effective Payroll Tax and the Holiday's Differential Impact



Notes: Illustrative example for an unclaimed worker at age 64. Panel (a): the statutory OASDI rate ($\tau = 6.2\%$) decomposes into an effective tax ($\tau^{\text{eff}} = \tau - \lambda_0$) and a work-linked recomputation offset (λ_0). The pre-claimer bar is shown with zero recomputation for visual simplicity; in the revised model, claimers may retain attenuated recomputation but lose the delay-claim channel. Panel (b): the 2 pp holiday reduces τ from 6.2% to 4.2% for all workers, but the proportional reduction in τ^{eff} is larger for high-λ_0 workers (50% for the 90% bracket vs. 32% for the stylized pre-claimer) because their effective tax is smaller to begin with. PIA brackets refer to the marginal replacement rate at the worker's 2010 AIME position: 90% below the first bend point, 32% between bend points, and 15% above the second. The figure is about the current-work margin; it does not include the separate value of delaying claim.

for claimed workers with earnings above the exempt amount before FRA; withheld benefits are later credited back through higher monthly payments (Social Security Administration, 2012, 2025a,b).

Horizon Heterogeneity. Workers with higher survival have a longer horizon over which future payoffs from continued work are realized. The same tax shock can therefore induce different responses even within a claiming group, depending on survival. In the empirical analysis, survival is a supporting heterogeneity dimension that lets the claiming-status response vary with horizon length.

Predetermined Heterogeneity in SIPP. Four sources of heterogeneity are fixed before the holiday. (i) Baseline payroll-tax exposure from 2010 covered earnings determines the dollar size of the cut. (ii) Pre-policy claiming status (PreClaimer_i) from the last observed month of

2010 separates workers facing a near-pure tax from those with benefit offsets. (iii) Baseline survival (S_i) is a transported proxy from 2010 covariates, held fixed. (iv) Among unclaimed workers, $\lambda_{0,i}$ is the marginal increase in future benefits from one more year of covered earnings, holding claim age fixed. All four are predetermined, so the holiday cannot change them.

3.1.1 Empirical Strategy

The empirical section uses three difference-in-differences comparisons built around the same national policy timing. Time-series variation comes from the holiday; cross-sectional variation comes from predetermined worker characteristics. The common treatment is

$$x_{it} = \text{Post}_t \times s_i,$$

where Post_t indicates the holiday period and s_i is a predetermined exposure measure. Year-month fixed effects absorb the common shock; person fixed effects absorb the level of s_i. What remains is the interaction: whether workers with larger baseline exposure change more during the holiday.

The main evidence compares 2011–2012 with 2010. The three designs are:

1. **Tax-exposure DiD.** s_i is baseline payroll-tax exposure (high vs. low). Does a larger dollar cut produce a larger change in retirement timing?

2. **Claiming-status DiD.** s_i is pre-policy claiming status. Do pre-claimers respond more than non-claimers? The specification includes $\text{Post}_t \times \text{Survival}_i$ and its interaction with PreClaimer_i so the main coefficient is evaluated at average survival.

3. **Continuous λ_0 DiD within unclaimed workers.** s_i is baseline $\lambda_{0,i}$, evaluated at a fixed claim age to exclude the delay channel. Does the response rise with the strength of recomputation?

This is a common-shock exposure design, closer to Finkelstein (2007) than to staggered-adoption designs: treatment timing is the same for everyone. The identifying assumption is conditional parallel trends: absent the holiday, the relevant groups would not have experienced different changes in retirement or employment over 2010–2012, after person and year-month fixed effects. Because all exposure measures are predetermined, reverse causality from the policy to treatment assignment is ruled out. Timing diagnostics and robustness checks are in Section B.

Table 1: SIPP Summary Statistics for Analysis Variables

Variable	Full sample	Pre-reform (2010–2010)	Policy window (2011–2012)
Age (years)	65.6 (2.6)	65.6 (2.6)	65.6 (2.6)
Employed indicator	0.353 (0.478)	0.352 (0.477)	0.353 (0.478)
Monthly earnings (TPEARN)	1,255 (3,305)	1,222 (3,259)	1,272 (3,328)
Annualized earnings (12×TPEARN)	15,058 (39,656)	14,666 (39,103)	15,259 (39,935)
Earnings / taxable maximum	0.140 (0.368)	0.137 (0.366)	0.141 (0.368)
Near-cap indicator (0.85–1.15 of max)	0.018 (0.131)	0.017 (0.130)	0.018 (0.132)
SS claiming indicator (ESSSELF=1)	0.702 (0.457)	0.710 (0.454)	0.698 (0.459)
SSI receipt indicator (ESTSSI=1)	0.204 (0.403)	0.209 (0.407)	0.201 (0.401)
Disability indicator (EDISABL=1)	0.284 (0.451)	0.281 (0.449)	0.285 (0.452)
Mortality proxy (10y, HRS transport)	0.132 (0.108)	0.131 (0.107)	0.133 (0.108)
Baseline mortality proxy (2010 person mean)	0.125 (0.099)	0.132 (0.102)	0.121 (0.097)
High-mortality group (top quartile, 2010)	0.265 (0.441)	0.279 (0.449)	0.258 (0.438)
Person-month observations	270,269	91,396	178,873
Unique individuals	12,036	9,364	10,587

Notes: Entries are mean (standard deviation). Pre-holiday is January–December 2010; holiday window is January 2011–December 2012. Employment indicator equals one if RMESR is 1 or 2. Near-cap indicator equals one if annualized earnings are between 85% and 115% of the year-specific OASDI taxable maximum.

3.2 Data

3.2.1 Data Source

The main data are from the 2008 panel of the Survey of Income and Program Participation (SIPP), using monthly core records from January 2010 through December 2012. This gives 12 pre-holiday months and 24 holiday months. The unit of observation is a person-month (i, t) for respondents aged 62–70. The main outcomes are monthly employment and retirement-onset hazard. The main heterogeneity measures are pre-policy claiming status, transported survival, and λ_0. Sample construction and variable definitions are in Appendix B.1 and B.2.

3.2.2 Descriptive Statistics

Table 1 reports summary statistics. The age profile and employment rate are stable across the pre-holiday and holiday windows (mean age 65.6 in both; employment 0.352 vs. 0.353), which is what a fixed-effects design around a common policy break requires. The sample contains 270,269 person-months from 12,036 individuals. The claiming indicator is high but not degenerate (0.710 in 2010), leaving meaningful variation. Near-cap observations are about 1.8% of the sample, so the estimates reflect broad late-career responses.

3.3 Motivating Evidence on Senior Labor Supply

Does the data move in the way our mechanism predicts? Before turning to the quantitative model, we need reduced-form evidence that the key channel is present. Our mechanism says that the response to the 2011–2012 payroll-tax holiday should depend on how large the tax change is and on how tightly payroll taxes are linked to future benefits on the work margin. We test this with common-shock exposure designs: the holiday is the common shock, and predetermined worker characteristics determine how strongly it affects effective work incentives. Because the model is organized around the extensive margin, we focus on retirement transitions and employment.

3.3.1 Data and Outcomes

All estimates use the monthly SIPP from January 2010 through December 2012. The holiday period is 2011–2012; the pre-period is 2010. Each specification takes the form $\text{Post}_t \times s_i$, where Post_t captures the timing of the tax cut and s_i is a predetermined exposure measure: baseline tax exposure, claiming status, or λ_0. The main outcome is the monthly retirement-onset hazard: the probability that a worker who was not retired in month $t-1$ enters retirement in month t. In one specification we use an employment indicator instead.

Before the main estimates, we document a descriptive fact. In both the HRS and the SIPP, older workers with higher survival expectations are more likely to be employed (Figure 4). This does not identify a causal effect, but it shows that survival is a meaningful source of heterogeneity in late-career labor supply.

3.3.2 Three Reduced-Form Patterns the Model Must Explain

We isolate three patterns, each using a different exposure measure s_i. The first uses baseline tax exposure. The second uses claiming status. The third uses the continuous recomputation rate λ_0 among unclaimed workers.

Pattern 1: Workers with larger baseline exposure show larger retirement changes around the holiday

The holiday lowered the payroll-tax rate by 2 percentage points for everyone, but the dollar size of the cut varied with covered earnings. Workers with higher baseline earnings received a larger tax reduction. If payroll taxes affect retirement timing, high-exposure workers should respond more.

Figure 4: Employment Gaps by Survival and Age Bin in HRS and SIPP



Notes: Each panel reports the high-minus-low survival coefficient from a within-year, within-age-bin regression for an employment indicator. Age bins are 60–64, 65–69, and 70–74. HRS high survival is defined by the within-cell median of subjective 10-year survival probability. SIPP high survival is defined by the within-cell median of transported baseline survival proxy (S_i). Error bars are 95% confidence intervals.

We split workers at the median of 2010 payroll-tax exposure and estimate:

$$Y_{it} = \alpha_i + \lambda_{ym} + \beta\,(\text{Post}_t \times \text{HighExposure}_i) + \Gamma_a + \varepsilon_{it}.$$

Here Y_{it} is the retirement-onset hazard, α_i are person fixed effects, and λ_{ym} are year-month fixed effects. The coefficient β asks whether the high-exposure group changed more after January 2011.

Table 2 shows it did. High-exposure workers experienced a 0.73 percentage-point larger increase in the monthly retirement-onset hazard during the holiday. Relative to a baseline hazard of about 2 percent, this is a large shift. We do not treat this as a standalone causal estimate — the two groups differ in baseline earnings and the pre-period is short — but it establishes that the holiday moved behavior and that the response was larger where the dollar cut was larger.

Pattern 2: The response is larger when payroll taxes are less tied to future benefits

A statutory tax cut is not the same as an effective tax cut for every worker. Workers who had already claimed face payroll taxes closer to a pure tax: the delay channel is gone, recomputa-

Table 2: Baseline Tax Exposure and Retirement-Onset Hazard Around the Holiday

	(1) Baseline	(2) + PreClaimer	(3) + Demographics	(4) Alt FE
Panel A: 2011–2012 vs 2010				
Post × HighExposure	0.0073***	0.0070***	0.0089***	0.0090***
SE	(0.0024)	(0.0025)	(0.0026)	(0.0026)
Post × PreClaimer		✓	✓	✓
Post × Baseline demographics (2010)			✓	✓
Age FE ($i(age)$)	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	
Year FE + Month FE				✓
Observations	77,142	77,142	77,142	77,142
Individuals	3,460	3,460	3,460	3,460
Panel B: 2013 vs 2011–2012				
Post × HighExposure	-0.0026	-0.0040	-0.0033	-0.0035
SE	(0.0036)	(0.0037)	(0.0040)	(0.0040)
Post × PreClaimer		✓	✓	✓
Post × Baseline demographics (2010)			✓	✓
Age FE ($i(age)$)	✓	✓	✓	✓
Individual FE	✓	✓	✓	✓
Year-month FE	✓	✓	✓	
Year FE + Month FE				✓
Observations	61,492	61,492	61,492	61,492
Individuals	2,860	2,860	2,860	2,860

Notes: Outcome is the monthly retirement-onset hazard. Post $= 1$ for 2011–2012. High exposure is the median split of 2010 payroll-tax bite. Ages 62–70. Standard errors clustered by person. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

tion is weaker, and some face the RET. Workers who had not yet claimed still offset part of the tax through benefit recomputation. If benefit linkage matters, pre-claimers should respond more.

We compare pre-claimers and non-claimers:

$$\begin{aligned} Y_{it} = \alpha_i + \lambda_{ym} &+ \beta_1 \left(\text{Post}_t \times \text{PreClaimer}_i\right) \\ &+ \beta_2 \left(\text{Post}_t \times \text{Survival}_i^c\right) \\ &+ \beta_3 \left(\text{Post}_t \times \text{PreClaimer}_i \times \text{Survival}_i^c\right) + \varepsilon_{it}. \end{aligned}$$

Here Y_{it} is employment, PreClaimer_i equals one for workers already claiming benefits, and Survival_i^c is centered baseline survival. Because survival is centered, β_1 is the pre-claimer effect at average survival.

Table 3 shows that pre-claimers responded more. Their employment changed by 7.6 percentage points more than non-claimers between 2010 and 2011–2012. Hours move in the same

Table 3: Claiming-Status Difference-in-Differences

	Employment indicator	Monthly hours
Post × Pre-claimer ($\hat{\beta}$)	0.076***	13.69***
	[0.064, 0.089]	[11.69, 15.70]
Survival interactions included	✓	✓
Person & year-month FE	✓	✓
Person-month observations	236,162	236,162
Unique individuals	8,499	8,499

Notes: Sample covers 2010 vs. 2011–2012. Specification includes Post× centered survival and its interaction with Pre-claimer; only the main coefficient is displayed. Because survival is centered, $\hat{\beta}_1$ is the pre-claimer effect at average survival. Ages 62–70. Brackets are 95% CIs clustered by person. *** $p < 0.01$.

direction. The response is larger when payroll taxes are less benefit-linked. This is a conditional comparison across workers in different initial claim states, not a structural estimate.

Pattern 3: Among unclaimed workers, the response varies with work-linked recomputation

Non-claimers are not all the same. The variable λ_0 measures the marginal increase in future annual benefits from one more year of covered earnings, holding claim age fixed at FRA. A higher λ_0 means a larger benefit offset, a smaller effective tax wedge, and therefore a larger proportional change in that wedge when the statutory rate falls by 2 percentage points. If the model is correct, high-λ_0 workers should respond more.

This moves from a binary to a continuous comparison. We restrict to unclaimed workers with observed earnings and estimate:

$$Y_{it} = \alpha_i + \lambda_{ym} + \beta \left(\text{Post}_t \times \lambda_{0,i}^{c}\right) + \text{Post}_t \times X_i'\gamma + \Gamma_a + \varepsilon_{it}.$$

Here Y_{it} is the monthly retirement-onset hazard, estimated on person-months at risk. α_i are person fixed effects, λ_{ym} are year-month fixed effects, and Γ_a are age fixed effects. $\lambda_{0,i}^{c}$ is centered recomputation: how far worker i's rate is above or below the sample mean. X_i includes centered survival and baseline demographics. Level effects of these time-invariant controls are absorbed by person fixed effects; what we identify is whether workers with different characteristics changed differently after the holiday.

The coefficient β is a continuous-treatment difference-in-differences. It asks whether the holiday effect grows with λ_0. If two unclaimed workers differ in λ_0 by 0.01, their response

Table 4: Benefit Accrual (λ_0) and Retirement-Onset Hazard Among Unclaimed Workers

	(1) Baseline	(2) + Demographics	(3) Alt FE
Post $\times$ λ_0 (centered)	0.8956***	0.6232**	0.6194**
SE	(0.2981)	(0.3098)	(0.3091)
95% CI	[0.3114, 1.4798]	[0.0160, 1.2303]	[0.0135, 1.2252]
Post $\times$ Survival (centered)	✓	✓	✓
Post $\times$ Demographics (2010)		✓	✓
Age FE	✓	✓	✓
Individual FE	✓	✓	✓
Year-month FE	✓	✓	
Year FE + Month FE			✓
Observations	44,321	44,321	44,321
Individuals	1,758	1,758	1,758

Notes: Sample is unclaimed workers aged 62–70 with observed earnings. Outcome is the monthly retirement-onset hazard. λ_0 is the derivative of annual benefits with respect to covered earnings, evaluated at 2010 AIME with claim age fixed at FRA. It captures work-linked recomputation, not the value of delaying claim. SEs clustered by person. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

should differ by $0.01 \times \beta$ probability units.

Table 4 shows a positive relationship. In the baseline, $\hat{\beta} = 0.90$ (SE 0.30, $p = 0.003$). The estimate is stable when demographics are added or when year-month fixed effects are replaced with separate year and month effects. Using the interquartile range of λ_0 (about 0.007), the estimate implies a 0.63 percentage-point larger increase in the monthly retirement-onset hazard for a worker at the 75th percentile versus the 25th. This is large relative to the baseline hazard. Robustness to alternative fixed claim ages is in the appendix.

A positive $\hat{\beta}$ does not mean high-λ_0 workers face a larger tax in levels. It means the holiday produced a larger proportional change in their effective wedge. Write the effective wedge as

$$\tau_i^{\text{eff}} = \tau^{\text{OASDI}} - \lambda_{0,i}.$$

A worker with $\lambda_0 = 0.022$ sees her wedge fall from 0.040 to 0.020, a 50% reduction. A worker with $\lambda_0 = 0.004$ sees hers fall from 0.058 to 0.038, a 34% reduction. The same statutory cut produces a larger effective-price change for the high-λ_0 worker.

This pattern is the reduced-form counterpart to the model's central state variable. It shows that the response varies continuously with the strength of benefit recomputation, not just across binary claiming groups.

3.3.3 Summary

These estimates establish that the mechanism is in the data before we ask the model to quantify it. The model will be disciplined by pre-reform employment moments, not by the holiday coefficients themselves. The three patterns provide benchmarks any credible model should match: larger exposure produces larger responses, the response is stronger when taxes are less benefit-linked, and the response varies with recomputation strength. The first pattern is suggestive. The second and third are tied to the mechanism and carry more weight.

Timing diagnostics, event studies, and robustness to alternative claim ages in the λ_0 construction are in Section B.

3.4 Parameterization

This section specifies the functional forms used to take the dynamic retirement model for seniors in Section 2 to data and maps the model's abstract objects to U.S. Social Security institutions. Section 3.4.1 parameterizes labor-supply costs; Section 3.4.2 describes the reduced-form retirement-value block; Section 3.4.3 introduces ex-ante asset heterogeneity across fixed asset types; and Sections 3.4.4–3.4.4 map the tax–benefit linkage to institutional rules. The key distinction is between the current-work margin and the claim-delay margin: only the former enters the effective wedge.

Our guiding principle is simple: each parameter should correspond to one transparent economic force. Some parameters govern how hard work feels as people age or as health worsens. Some govern how attractive retirement is. Others map the Social Security formula into the effective reward from one more year of work. This follows the broad approach of the structural retirement literature, which models retirement as the outcome of age-varying tastes, health, assets, and Social Security incentives (Rust and Phelan, 1997; French, 2005; French and Jones, 2011; Coile and Gruber, 2007; Blundell et al., 2016). The point of our parameterization is therefore not to maximize flexibility. It is to keep the model simple enough that each estimated object has a clear economic interpretation. Throughout this section, the notation α denotes the fixed joint productivity-health-survival support used in the quantitative model. There is no second post-62 health state in the benchmark Bellman: persistent health heterogeneity is already part of α, while short-run deterioration in work capacity is captured by the transitory shock that is later compressed into ν_t in the reduced-form theory.

3.4.1 Disutility Distribution

We parameterize the log-disutility mean as a kinked-linear function of age with a health-disutility term:

$$\ln \nu_t \;\sim\; \mathcal{N}\big(m(a,\alpha),\; \sigma^2\big), \qquad m(a,\alpha) \;=\; m_0 + m_1(a-62) + m_2 \max(0,\, a-65) + \psi\, r_\alpha, \quad (31)$$

where $\psi > 0$ is a common health-disutility scale and $r_\alpha \geq 0$ is the health-disutility rank associated with joint type α. Because the fixed type α jointly determines earnings and health, higher-α workers have both higher wages and better health, so r_α is decreasing in α: the lowest type has $r_\alpha = 1$ (highest health-related work disutility) and the highest type has r_α closest to zero. Flexible work costs that vary with age and health are standard in structural retirement models (Rust and Phelan, 1997; French, 2005; French and Jones, 2011). We extend by using a kinked-linear age profile after 62 and a type-specific health loading so that the parameters map directly into the density of workers near the retirement margin. The parameter m_0 is the baseline level of work disutility at age 62. The parameter m_1 is the baseline age slope in work disutility from age 62 onward: if m_1 is larger, work becomes harder more quickly as workers age. The parameter m_2 is an additional slope that turns on after age 65, so it captures the idea that the disutility of continued work may rise faster once workers are very close to the Full Retirement Age. The health term ψr_α has a similarly simple interpretation. The common scale ψ says how much health matters overall, while r_α says how far each joint type is from the healthiest benchmark.

The continuation premium $d_{g,t} \equiv V_g^W - V^R$ is denominated in dollars, but the disutility ν_t is a unitless draw from the log-normal distribution above. To make the two comparable, we divide the dollar premium by a normalization constant $\kappa = 10{,}000$ (Section 3.5.1) before evaluating the continuation probability. The rescaled quantity $d_{g,t}/\kappa$ is now on the same scale as the disutility draws: a worker continues if her realized disutility falls below $d_{g,t}/\kappa$. The value of κ acts as an exchange rate between dollars and disutility units—it determines how many dollars of continuation premium are needed to offset one unit of work disutility.

Under this specification, the density condition discussed in the model section is transparent: the log-normal density is unimodal with mode $\bar{\nu}(a,\alpha) = \exp(m(a,\alpha) - \sigma^2)$, increasing on $(0, \bar{\nu})$ and decreasing on $(\bar{\nu}, \infty)$. When both continuation thresholds lie above the mode, the density is declining, so $\nu_c^* < \nu_u^*$ implies $f(\nu_c^*) \geq f(\nu_u^*)$: claimed workers sit at a denser point.

3.4.2 Retirement Value

The benchmark quantitative model uses a reduced-form retirement value rather than a full post-retirement consumption–savings problem. This term captures the value of nonmarket time, home production, pension-like resources, and other non-labor components of retirement that are not modeled explicitly.

Formally, the retirement state carries an intercept ω_0 that shifts the value of exit relative to continued work. This keeps the benchmark model parsimonious and focuses attention on the margins that matter for the paper: the work disutility distribution, the value of delaying claiming, and the Social Security work-incentive schedule. It also makes clear that the model is designed for conditional labor-supply incidence after age 62, not for a full welfare analysis of household retirement and claiming.

3.4.3 Ex-Ante Asset Heterogeneity

Assets matter for retirement, but in the benchmark model they enter as ex-ante asset types rather than as a continuously evolving savings state. Workers enter age 62 as one of a small number of predetermined asset types. These types summarize liquidity and therefore how costly it is to postpone benefit receipt.

Formally, let $x \in \{L, M, H\}$ denote low-, middle-, and high-asset workers. The model maps these asset types into a waiting cost $\chi(x)$, with low-asset workers facing the highest cost of remaining unclaimed. This reduced-form asset structure is sufficient for the paper because the empirical question is how seniors with different balance-sheet positions respond to the payroll-tax holiday once they are already near retirement, not how they optimally save over the entire lifecycle.

3.4.4 Institutional Parameterization

The institutional mapping splits into three objects. First, the baseline marginal work-recomputation slope $\lambda_0(\alpha)$ is pinned down by the PIA formula through the earnings component of joint type α. Second, early-claim reductions and delayed retirement credits determine the benchmark claim-age schedule. Third, the Retirement Earnings Test enters separately as a current-value wedge for claimed workers. Mapping these institutional ingredients into the benchmark claim schedule and the reduced-form continuation term $A_g(t, \alpha)$ and normalized wedge $\kappa_g(t, \alpha)$ is standard in structural retirement models (Rust and Phelan, 1997; Coile and Gruber, 2007; Blundell et al., 2016). Our contribution is to organize that mapping around the effec-

tive wedge on current work, and then use it to connect the reduced-form evidence and the quantitative model. These institutional components matter because they tell us how much of today's payroll tax is really a tax and how much is payment for higher future benefits on the work margin.

Primary Insurance Amount and the Earnings Component of α The PIA formula converts a worker's Average Indexed Monthly Earnings (AIME) into a monthly benefit amount through a piecewise-linear schedule with three brackets. The bend points are indexed to the year a worker attains age 62. For workers attaining age 62 in 2010, the formula is (Social Security Administration, 2012)

$$\text{PIA}(\text{AIME}) \;=\; \begin{cases} 0.90 \times \text{AIME} & \text{if AIME} \le b_1, \\ 0.90 \times b_1 + 0.32 \times (\text{AIME} - b_1) & \text{if } b_1 < \text{AIME} \le b_2, \\ 0.90 \times b_1 + 0.32 \times (b_2 - b_1) + 0.15 \times (\text{AIME} - b_2) & \text{if AIME} > b_2, \end{cases} \tag{32}$$

where $b_1 = \$761$ and $b_2 = \$4{,}586$ are the 2010 bend points. The PIA formula factors—90%, 32%, and 15%—decline sharply with earnings, so higher-AIME workers receive a smaller fraction of their marginal earnings as benefits.

Write $e(\alpha) \in \{L, M, H\}$ for the earnings component of joint type α, where each value places the worker in one of the three PIA brackets. The baseline work-recomputation rate for a worker of type α is

$$\lambda_0(\alpha) \;=\; r(e(\alpha)) \times \frac{1}{12 \times 35}, \tag{33}$$

where $r(e(\alpha)) \in \{0.90, 0.32, 0.15\}$ is the marginal PIA replacement rate for the bracket corresponding to the earnings component of α, and $1/(12 \times 35)$ reflects the conversion from an additional dollar of annual earnings to the change in AIME (monthly average over 35 highest years).[2] Workers above the OASDI taxable maximum ($\$106{,}800$ in 2010–2011) pay no marginal tax and receive no marginal accrual, so $\lambda_0 = 0$ and $\tau^{\text{eff}} = 0$ for earnings above the cap.

The intuition is that the PIA formula determines the marginal reward from one more year of covered earnings holding claim age fixed. Low earners are on the steep part of the formula, so an extra dollar of earnings translates into a larger future benefit gain. High earners

[2]This expression assumes that the worker has at least 35 years of positive covered earnings and that current-year earnings are high enough to enter the top-35 set. If either condition fails—e.g., a worker with fewer than 35 positive years displaces a zero, or current earnings fall below all previous top-35 years—the marginal effect on AIME differs from $1/(12 \times 35)$. For workers aged 62 and above with long careers, both conditions are typically satisfied.

are on the flatter part, so the same extra dollar buys less additional Social Security. This is why the same payroll tax can imply different effective wedges across joint types through the earnings component of α. In the empirical baseline, we evaluate λ_0 at FRA, so (33) is a pure recomputation object and does not embed the separate value of delaying claim.

Delayed Retirement Credit Workers who delay claiming benefits past the Full Retirement Age (FRA, age 66 for the 1943–1954 birth cohorts) receive delayed retirement credits equal to $\frac{2}{3}$ of 1 percent per month of delay, or 8 percent per year, up to age 70. For a worker who claims at age $a_{\text{claim}} \geq$ FRA, the monthly benefit is

$$b(a_{\text{claim}}) \;=\; \text{PIA} \times \big[1 + 0.08 \times (a_{\text{claim}} - \text{FRA})\big], \tag{34}$$

where age is measured continuously in years. Under SSA rules the adjustment is applied monthly. Social Security rewards patience. A worker who waits until age 68 to claim—two years past her FRA of 66—receives a monthly check that is 16 percent larger than if she had claimed at 66, and that higher check continues for the rest of her life.

For workers who claim before FRA, benefits are reduced by $\frac{5}{9}$ of 1 percent per month for the first 36 months of early claiming and $\frac{5}{12}$ of 1 percent per month for any additional months. A worker with FRA 66 who claims at 63 therefore loses 6.67% per year for those three years, but claiming at 62 adds a further reduction at the lower 5%-per-year rate. The early reduction schedule is not symmetric with the delayed retirement credit. Together with the PIA level implied by the earnings record, these early-claim reductions and delayed-retirement credits determine the benchmark claim-age schedule used in Section 2.2. This schedule belongs to the claim-versus-wait margin. It is not part of λ_0, because a retire-and-wait worker can also move along the claim-age schedule without working.

Retirement Earnings Test The Retirement Earnings Test (RET) reduces current benefits for workers who have claimed before FRA and have earnings above an annual exempt amount. For beneficiaries who are under FRA for the entire year, SSA withholds $\$1$ of benefits for every $\$2$ of earnings above the exempt amount ($\$14{,}160$ in 2010; statutory slope $= 0.50$). In the year a worker attains FRA, the higher exempt amount of $\$37{,}680$ applies, benefits are withheld at $\$1$ for every $\$3$ of excess earnings (statutory slope $= 1/3$), and only earnings before the month of attaining FRA are tested. After the FRA month, the RET no longer applies. In the benchmark quantitative model, we summarize this rule as an effective current wedge on work for claimed workers below FRA.

This reduced-form wedge treats θ_{liq} as the contemporaneous current-value cost of the RET and abstracts from the later actuarial readjustment of benefits when withheld checks are credited back through SSA's adjustment formulas. That simplification is deliberate: the paper focuses on the immediate work incentive faced at the time of the labor-supply decision. We therefore interpret θ_{liq} as a net current-value or liquidity-adjusted RET wedge, not as the literal statutory withholding rule (Social Security Administration, 2025a,b).

This is the key reason the claimed-worker wedge can be so different from the unclaimed-worker wedge before FRA. Once a worker has claimed, continued work no longer creates delayed-claim credits. If earnings are high enough to trigger the RET, current work becomes less attractive because the worker faces both payroll taxes and a current-value benefit-withholding penalty, even though some later recomputation may remain. In the model section, this enters the claimed-worker continuation term $A_c(t,\alpha)$ in (6) and therefore the normalized wedge $\kappa_c(t,\alpha)$ in (7).

Effective Tax Rate How much does working actually cost a senior, once we account for all the institutional pieces above? Every worker pays the statutory payroll tax rate $\tau = 0.062$ (the employee OASDI rate). But the effective burden differs. A worker earns work-linked recomputation worth $\lambda_g(\alpha)\,\mu_t(\alpha)$ for each dollar of covered earnings, which offsets part of the tax—generously for low earners on the steep part of the PIA formula, and barely at all for high earners on the flat part. A claimed worker who is still below FRA and earning above the RET exempt amount faces the opposite problem: on top of the payroll tax, SSA may withhold part of the current benefit check, which we summarize with the current-value wedge θ_{liq}.

Putting these together, the effective tax rate is

$$\tau_g^{\text{eff}}(\alpha,t) \;=\; \underbrace{\tau}_{\text{statutory tax}} \;-\; \underbrace{\lambda_g(\alpha)\,\mu_t(\alpha)}_{\text{work-linked recomputation offset}} \;+\; \underbrace{\theta_{\text{liq}}\Big(1-\frac{\bar{y}_t}{w(\alpha)}\Big)_+ \mathbf{1}\{g=c,\; a_t<\text{FRA}\}}_{\text{RET current-value wedge for claimed workers}}.$$

During the 2011–2012 payroll tax holiday, τ fell from 0.062 to 0.042. In our calibration, all three earnings-support points embedded in α ($\$21{,}000$, $\$30{,}000$, $\$40{,}500$) exceed the RET exempt amount $\bar{y} = \$14{,}160$, so every working claimer below FRA faces the RET current-value wedge.

This equation ties the whole paper together. Congress cut the statutory rate by the same 2 percentage points for everyone, but the effective tax change was not the same. A worker whose joint type places her on the low-earnings part of the PIA schedule already had a low effective tax because of her generous recomputation offset, so the 2pp cut is a smaller proportional change in her effective wedge. A worker whose joint type places her on the high-

earnings part of the schedule gets almost no recomputation offset, so her effective tax is close to the statutory rate and the same 2pp cut is proportionally larger. A claimed worker below FRA faces the statutory tax plus the RET current-value wedge and only attenuated recomputation—her effective tax is the highest of the three, and the 2pp cut provides the same dollar relief but operates on a very different baseline. These differences in the effective wedge across joint types and claiming groups generate the heterogeneous treatment effects that the model matches quantitatively.

3.5 Calibration and Validation

We calibrate a dynamic retirement model for seniors to 2010 SIPP employment data—the year immediately before the payroll tax holiday—treating it as the steady state of the benchmark economy. The model begins at age 62 and tracks individuals through retirement. This is not a full lifecycle model. We do not ask how households built their wealth over several decades, how spouses coordinate claiming, or how post-62 savings respond. Instead, we take the key late-life state variables as given at age 62 and study how they shape work and retirement decisions after that point.

One of those predetermined state variables is assets. We model asset heterogeneity as ex-ante heterogeneity rather than as a post-62 savings problem. This is a reasonable simplification for our setting because most wealth accumulation has already occurred by the early 60s, while our empirical question is about a short-run tax change hitting workers who are already close to retirement. Assets in the benchmark are therefore supports for liquidity and waiting costs, not a post-62 choice variable.

We estimate the parameters needed to fit retirement behavior. Policy parameters are fixed at their 2010 statutory values. The remaining non-estimated parameters are either normalization or reduced-form supports introduced because the model does not contain a savings decision or earnings process. Social Security rules therefore follow 2010 law: claiming is permitted from age 62 to 70, the Full Retirement Age is 66, delayed retirement credits accrue after FRA, and the Retirement Earnings Test applies before FRA.

The benchmark economy features a 6.2 percent employee payroll tax. The counterfactual lowers this rate to 4.2 percent, holding everything else fixed. For workers who have not yet claimed Social Security, an additional year of covered earnings generates a future benefit recomputation credit. Claimed workers may retain attenuated recomputation, but they lose the delay-claim channel and may face the RET current-value wedge before FRA. We build these current-work incentives into the dynamic programming problem.

3.5.1 Targeted Moments and Estimation

The model has 11 parameters that we need to estimate: how much workers dislike working at different ages, how health affects the cost of continued employment, and how fixed joint productivity-health types shape labor-market behavior after age 62. We choose these parameters so that the model's predicted employment patterns match what we observe in the data before the payroll tax holiday.

The data targets are 13 pre-reform employment moments. Nine of them are age-specific employment rates from age 62 to 70, which discipline how quickly workers exit as they age. The remaining four are employment rates for each of the four self-reported health groups (pooled across ages), which discipline how much health matters for the retirement decision and therefore the health component of the joint type α. These moments map to the permanent health-survival part of α, not to a separate post-62 health transition state. Policy parameters (tax rates, benefit rules, the RET) and the heterogeneity grids (joint productivity-health types, asset types) are not estimated. We fix these directly from institutional rules or discretization choices described in Section 3.4, keeping the estimation focused on the parameters that the data can identify.

Formally, let m^{data} be the vector of data moments and $m(\theta)$ the model-simulated counterpart. We estimate

$$\hat{\theta} = \arg\min_{\theta} \left[m(\theta) - m^{\text{data}}\right]' W \left[m(\theta) - m^{\text{data}}\right],$$

using a diagonal weight matrix W. The age block receives twice the weight of the health block. Within the age block, we place more weight on ages 62–65, where the bulk of retirement transitions occur, and less on ages 66–70, where employment rates are lower and measured less precisely.

Parameters Table 13 reports the fixed parameters. We group them into three classes: parameters fixed by convention or normalization, parameters fixed by 2010 Social Security law, and parameters that we discipline directly from the data before the SMM. Table 14 reports the 11 behavioral parameters estimated by SMM.

Model Fit Table 15 reports fit to the 13 targeted moments and Figure 5 shows the model fit. The model reproduces the steep drop in employment from age 62 through 64 and still tracks the data reasonably well at age 70. It overpredicts employment modestly from age 65 onward, so retirement remains somewhat too slow around the Full Retirement Age. Across

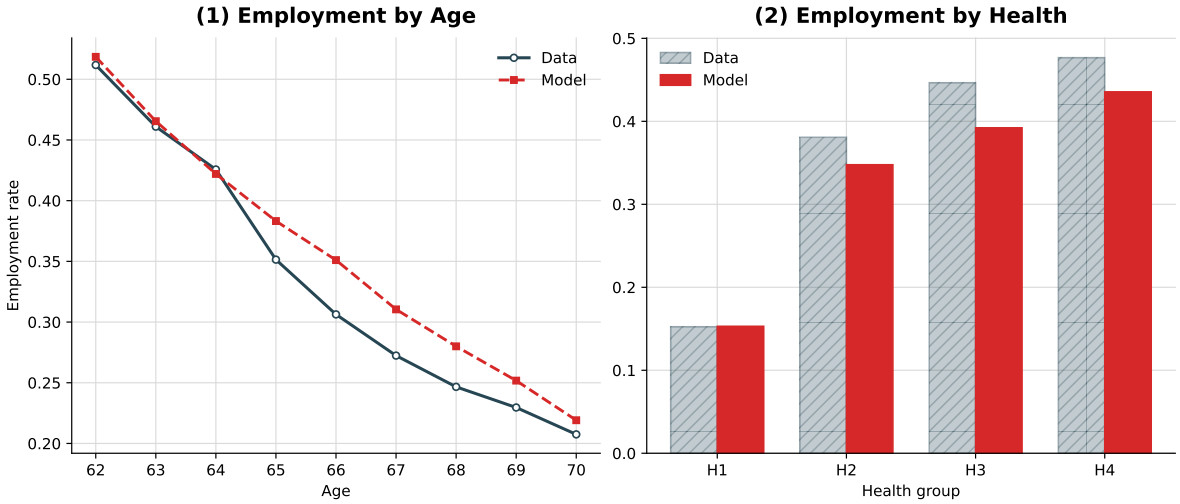


Figure 5: Benchmark fit to targeted moments. Left panel: employment by age, 62–70. Right panel: employment by self-reported health group.

health groups, the model matches the worst-health group $H1$ almost exactly and understates employment for $H2$–$H4$, with the largest miss in $H3$.

Overall, the calibrated model captures the main features of pre-reform senior employment: the rapid retirement between 62 and 65, the modest additional decline after the Full Retirement Age, and the large gap between healthy and unhealthy workers.

3.5.2 Distribution of Effective Wedges

Before turning to employment responses, it is useful to see how much the effective tax varies across workers. We do this with a simple exercise: hold the statutory payroll tax fixed at 6.2 percent for everyone, simulate the model forward, and for each worker still employed at a given age, compute her effective tax rate using (8). The result is a distribution of work-specific effective wedges across the workforce at each age. Because we hold the statutory rate fixed, all variation in this distribution comes from Social Security's institutional rules interacting with claiming status, work-linked recomputation, the fixed joint productivity-health type α, and the current work-capacity shock, rather than from the tax holiday itself.

Figure 6 shows two facts. First, the distribution is wide and asymmetric. Second, the dispersion is tightly tied to institutional status. Unclaimed workers ages 62–65 face lower wedges because work-linked recomputation offsets part of the statutory tax. Claimed workers of the same ages face the highest wedges because the delay-claim channel is gone and the RET adds a current-value penalty before FRA. After FRA, claimed workers move back toward the statu-

tory wedge, while unclaimed workers continue to face lower effective wedges because one more year of work still buys higher future benefits through recomputation.

The point of the figure is not that workers face different statutory tax rates. They do not. The point is that the same statutory rate maps into very different effective prices of work once Social Security's rules are taken seriously. That distribution of effective wedges is the object that the reduced-form evidence and the counterfactual exercises are both trying to discipline.

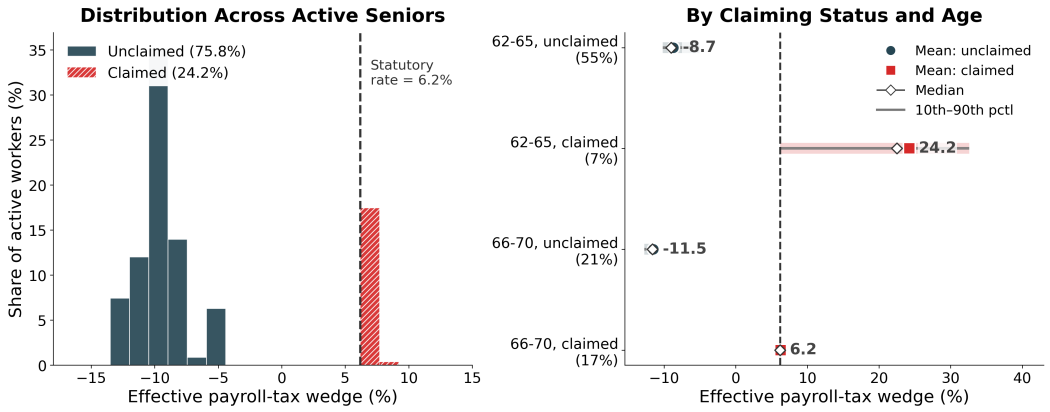


Figure 6: Model-implied distribution of effective payroll-tax wedges. Panel A plots the weighted histogram of work-specific effective wedges among active workers ages 62–70 under a constant 6.2 percent employee payroll tax. Panel B decomposes the distribution by claiming status and age block; dots are means, hollow markers are medians, and horizontal bars show the 10th to 90th percentile range. The statutory payroll tax rate is shown by the dashed vertical line.

3.5.3 Results

Who responds to the tax cut, and who does not? Table 16 breaks down the employment response across age, health, assets, and earnings. The corrected calibration points to a modest aggregate response: the permanent 2pp payroll-tax cut raises employment for ages 62–70 by about 0.3 percentage points. The reform still matters most for workers who are close to the retirement margin, but the heterogeneity is much more compressed than the wide dispersion in effective wedges might suggest.

Age. Employment rises by 0.1 percentage points at ages 62–65 and by 0.5 points at ages 66–70. The response is therefore somewhat larger at older ages, where more workers are already near the exit margin.

Health. There is no strong monotone health gradient in the corrected calibration. The worst-health group ($H1$) barely responds at all (0.1 pp), while $H2$ and $H4$ each rise by about

0.4 points and $H3$ by about 0.2 points. Poor health still reduces responsiveness, but most of the remaining dispersion across health groups is modest.

Assets and earnings. Asset differences are also limited: low- and middle-asset workers respond by about 0.3 percentage points, compared with 0.2 points for high-asset workers. Along the earnings dimension, the response is hump-shaped rather than monotone: it is about 0.3 points at \$30k and \$60k, rises to 0.5–0.7 points around \$90k–\$110k, and then falls back to 0.3 points at \$150k. Workers near the taxable maximum experience the largest effective change in the payroll tax, while workers far above the cap see the cut diluted because only part of earnings is taxable.

What drives these patterns? To find out, Table 17 decomposes the response by shutting off heterogeneity channels. In each panel, we start from a stripped-down model in which every group has a common baseline response, and then add channels back to see which ones pull the groups apart.

Assets (Panel A). Without asset-dependent channels, every group would respond by about 0.4 percentage points. Asset heterogeneity mostly compresses that common response rather than amplifying it. The main force is the waiting-cost channel, which lowers the response for every asset group and especially for high-asset households; the wage-sorting channel is quantitatively small. On net, asset heterogeneity narrows responses into a range of roughly 0.2–0.3 percentage points.

Health (Panel B). If we shut off all health-specific channels, every health group would respond by the same 0.52 percentage points. Health then mostly operates by offsetting that common response. For the sickest group, lower survival and poorer work capacity almost wipe out the effect, leaving a response of only 0.06 percentage points. For the remaining groups, longevity and work-capacity forces partly offset one another, so the net response stays between about 0.24 and 0.41 percentage points.

Interpretation. The two decompositions tell a more muted but still informative story. Assets and health matter primarily by shifting how much of the common incentive survives once workers are sorted across waiting costs, survival, and work capacity. The aggregate employment effect remains small not because effective wedges are homogeneous—they are not—but because many seniors are either already claimed, already retired, or not close enough to the extensive-margin threshold for a 2pp payroll-tax cut to move behavior much.

The broader lesson is that the aggregate employment effect of a tax reform depends not on the average incentive it creates, but on where in the joint distribution of wealth, earnings, and health the marginal workers sit. Two policies with identical fiscal cost can produce very different responses if they reach different parts of this distribution.

4 Quantitative

4.1 Counterfactual Exercises

We use the calibrated model to study five policy experiments. The first is a permanent cut in the payroll tax rate from 6.2 to 4.2 percent. This is deliberately stronger than the historical two-year holiday; interpret it as a comparative-static experiment that isolates the incentive channel. The second removes the Retirement Earnings Test (RET) current-value wedge while holding all other policy rules fixed. The third shuts down the marginal link between post-62 earnings and future benefits while holding the separate claim-age schedule fixed. The fourth removes the claim-age adjustment schedule while holding work-linked recomputation fixed. The fifth aggregates overlapping cohorts and imposes the actual 2011–2012 holiday as an unexpected calendar-year shock. Because assets are predetermined at age 62 in our model, differences across asset groups below should be read as differences across seniors who start retirement with different balance sheets, not as the result of a new savings response after the reform.

4.1.1 Substitution Versus Income Effects

In the model, cutting the payroll tax raises employment for ages 62–70 from 0.3583 to 0.3612, or by about 0.29 percentage points. The basic question is why this happens. Is it because seniors simply have more money, or because working becomes more rewarding?

To answer that question, we compare the tax cut to two simpler cases. First, we replace the tax cut with a lump-sum transfer that costs the government the same amount. This transfer lowers employment very slightly, from 0.3583 to 0.3581. In other words, giving seniors extra income by itself does not make them work more; if anything, it makes them retire a bit earlier.

Second, we require the reform to be budget balanced. We let the payroll-tax cut be financed by a small proportional cut in Social Security benefits. The model implies that the required benefit cut is only about 0.77 percent. Even with this financing, employment is still 0.3611, which is almost the same as in the original tax-cut experiment.

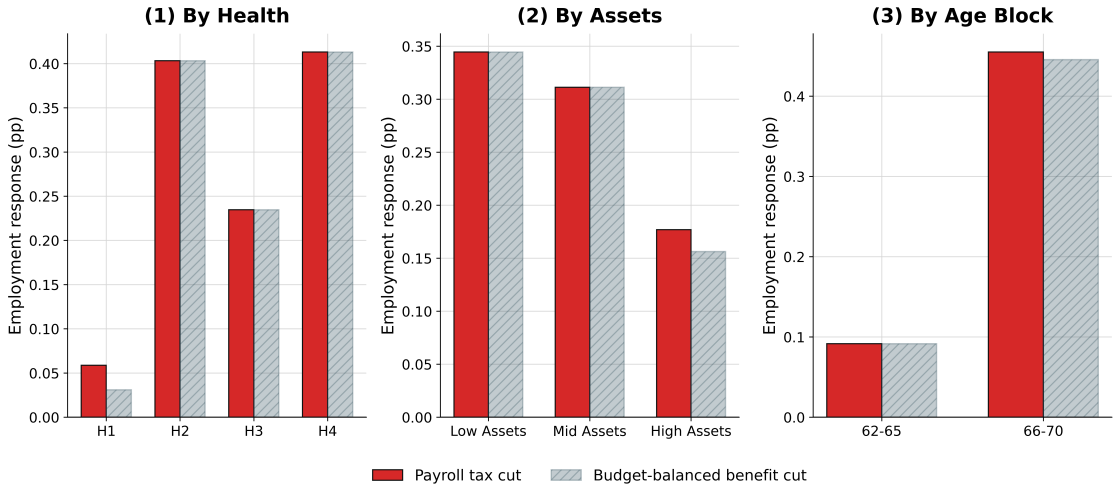


Figure 7: Payroll-tax cut versus budget-balanced reform. Bars report employment responses by health, assets, and age block. The budget-balanced reform finances the tax cut with a proportional benefit reduction chosen to match the benchmark fiscal balance over ages 62–90.

The lesson is straightforward. The payroll-tax reform matters mainly because it changes the payoff from working one more year. That is a substitution effect. A pure income transfer does not generate the same response, and adding a small benefit cut barely changes the result. Figure 7 confirms this pattern across health, asset, and age groups: the tax-cut and budget-balanced bars are nearly identical in every subgroup. So the main channel in the model is the incentive to work, not the extra cash itself, even though the corrected calibration implies that the overall employment effect is modest.

4.1.2 Removing the Retirement Earnings Test

The next three exercises isolate distinct channels through which Social Security shapes senior labor supply. We begin with the Retirement Earnings Test. Before the Full Retirement Age, a worker who has already claimed benefits loses current benefit income when earnings rise above the exempt amount. On paper, this is a timing device: withheld benefits are mostly returned later through a higher monthly payment. For seniors with enough savings and a long life expectancy, the repayment is close to actuarially fair. But for early claimers with little savings, a dollar returned years later is worth much less than a dollar lost today. For them, the RET creates a sizable current-value wedge that penalizes combining work with benefit receipt before FRA.

We run this exercise for two reasons. First, it is directly policy-relevant. Congress elimi-

nated the RET for workers above the FRA in the Senior Citizens' Freedom to Work Act of 2000, but kept it in place for workers between 62 and the FRA—precisely the age range in our model. Whether to extend the repeal to pre-FRA workers has been raised in subsequent legislative proposals. Second, it helps interpret our main mechanism. In the data and in the model, claimed workers respond differently from unclaimed workers. That gap can arise for two reasons: claimed workers no longer earn delayed retirement credits from postponing claiming, and before FRA they may also face the RET. Shutting down the RET isolates the second force.

In the counterfactual, we set the RET current-value wedge to zero and leave all other parameters unchanged. It asks what happens if seniors who already claimed can keep their full current-value benefit while continuing to work before FRA. The result is still large. Table 5 reports the main moments. Employment at ages 62–65 rises from 49.7 to 54.8 percent, a 5.1 percentage-point increase. The effect is smaller after FRA, where the RET no longer applies mechanically: employment at ages 66–70 rises from 31.1 to 31.9 percent. Claiming also rises before FRA, from 80.1 to 84.7 percent. The reason is intuitive. Once the RET penalty is removed, some seniors no longer have to choose between claiming for current income and continuing to work.

The sharpest pattern is across assets. Pre-FRA employment rises by 8.2 percentage points for the low-asset group, 5.5 points for the middle-asset group, and only 0.2 points for the high-asset group. Low-asset seniors value current benefit income the most, so the RET is especially distortionary for them. Removing it lets them claim for liquidity reasons without leaving work at the same time.

Two findings stand out. First, the RET current-value wedge is quantitatively large: removing it raises pre-FRA employment by 5.1 percentage points, far larger than the 0.3 percentage-point effect of the permanent payroll-tax cut. Second, the response is concentrated among low- and middle-asset seniors, the groups for whom current benefit income matters most. This implies that the RET does not simply shift the timing of benefit receipt in practice—its current-value component actively discourages work among seniors who need current income the most.

4.1.3 Delinking Benefits from Additional Earnings

The delinking counterfactual works through a different channel. It applies to workers who have not yet claimed. For them, working longer can raise future Social Security benefits through earnings-record recomputation, so continued work brings both current wages and higher future benefits. If we remove that link, working longer becomes less attractive, and

Table 5: Removing the Retirement Earnings Test: Employment and Claiming Responses

	Employment rate (%)			Claiming rate (%)		
	Benchmark	No RET	Diff.	Benchmark	No RET	Diff.
Panel A. Overall						
Ages 62–65	49.7	54.8	5.1	80.1	84.7	4.6
Ages 66–70	31.1	31.9	0.7	92.9	92.9	0.0
Ages 62–70	39.6	42.3	2.7	87.1	89.2	2.1
Panel B. Pre-FRA employment by asset group (ages 62–65)						
	Employment rate (%)					
	Benchmark	No RET	Diff.			
low_assets	43.8	52.0	8.2			
mid_assets	48.4	54.0	5.5			
high_assets	60.0	60.2	0.2			

Notes: Panel A reports benchmark and no-RET employment and claiming rates. Panel B reports benchmark and no-RET employment rates by ex-ante asset group for ages 62–65. The no-RET counterfactual sets the benchmark current-value RET wedge to zero while holding all other calibrated parameters fixed. Differences are in percentage points.

claiming begins earlier. This is not the same mechanism as the RET. The RET changes the return to working *after* claiming. Earnings-record recomputation changes the return to working *before* claiming.

The RET exercise isolates one part of the claimed-worker wedge. A different question is what happens if the Social Security system stops rewarding extra work after age 62 through higher future benefits. This is the most direct counterfactual for our mechanism because the paper's central current-work object is benefit linkage through recomputation.

It is important to be precise about what we change. We do *not* give everyone the same Social Security benefit. That would be a much broader redistribution reform. Instead, we hold benefit levels fixed and remove only the extra benefit credit from one more year of work after age 62, holding the claim-age schedule fixed. In other words, we ask what happens if older workers no longer earn additional future benefits by staying on the job even though delayed claiming still changes benefits through the ARF-DRC schedule.

The answer is clear. Table 6 shows that delinking lowers employment and raises claiming. Employment at ages 62–65 falls from 49.7 to 46.9 percent, a 2.8 percentage-point decline. Over ages 62–70, employment falls by 2.6 percentage points, while claiming rises by 5.9 percentage points. The largest employment decline is concentrated around age 66, where it reaches about 5.9 percentage points. The basic intuition is simple: when extra work no longer raises future benefits through recomputation, the payoff to continued work falls, so some seniors claim

earlier and leave work sooner.

The heterogeneity is also informative. The pre-FRA employment decline is 1.4 percentage points for the lowest-asset group, 1.5 points for the middle-asset group, and 6.9 points for the highest-asset group. This does not mean high-asset seniors are worse off in a welfare sense. It means they are the group with the most room to adjust. In the benchmark, low-asset seniors already tend to claim early, so there is less additional margin to lose. By contrast, higher-asset seniors are the ones who use the option to delay claiming and keep working. Once that extra reward from work is removed, they move noticeably toward earlier claiming and earlier exit. Across health groups the pattern is less sharply ordered, so the dominant heterogeneity in this counterfactual is financial rather than health-based.

This is the key takeaway. In the model, work-linked recomputation sustains late-life work by rewarding continued earnings among seniors who still have room to delay claiming. Removing that linkage pushes claiming earlier and lowers employment. The companion claim-age schedule remains in place, so this counterfactual isolates the work margin rather than shutting down every channel through which waiting can raise benefits.

Table 6: Removing Work-Linked Earnings-Record Recomputation After Age 62

	Employment rate (%)			Claiming rate (%)		
	Benchmark	Delinked	Diff.	Benchmark	Delinked	Diff.
Panel A. Overall						
Ages 62–65	49.7	46.9	-2.8	80.1	88.1	8.0
Ages 66–70	31.1	28.7	-2.5	92.9	97.0	4.1
Ages 62–70	39.6	37.0	-2.6	87.1	92.9	5.9

Panel B. Pre-FRA employment by asset group (ages 62–65)

	Employment rate (%)		
	Benchmark	Delinked	Diff.
low_assets	43.8	42.4	-1.4
mid_assets	48.4	47.0	-1.5
high_assets	60.0	53.1	-6.9

Notes: The counterfactual sets the work-linked earnings-record recomputation term to zero while holding the claim-age schedule, benefit levels, and all other calibrated parameters fixed. Differences are in percentage points.

4.1.4 Removing Delay-Claim Credits

The previous experiment held the claim-age schedule fixed and shut down only the work-linked recomputation channel. We now do the reverse. This counterfactual sets the claim-

age adjustment schedule to zero while leaving earnings-record recomputation in place. It therefore removes the payoff to delaying claim without erasing the separate payoff to current work through a stronger earnings record.

Table 7 shows that this claim-delay margin also matters, but in a different way. Employment at ages 62–65 falls from 49.7 to 47.8 percent, a 1.9 percentage-point decline, and employment over ages 62–70 falls by 1.0 percentage point. Claiming responds much more strongly than employment: the claim share over ages 62–70 rises from 87.1 to 100 percent. In other words, once delay credits disappear, nearly everyone claims as soon as it is feasible, but work does not collapse because the recomputation channel is still operating.

The heterogeneity is again concentrated in assets. Pre-FRA employment is essentially unchanged for the low- and middle-asset groups, but it falls by 7.6 percentage points for the high-asset group. This is exactly the group that uses delay claiming most intensively in the benchmark. The contrast with Table 6 is useful. Removing recomputation weakens the return to current work directly, while removing delay credits primarily changes the timing value of waiting to claim. The two channels both matter, but the model assigns the larger employment effect to work-linked recomputation and the larger claiming effect to delay credits.

Table 7: Removing Claim-Age Adjustments After Age 62

	Employment rate (%)			Claiming rate (%)		
	Benchmark	No delay credits	Diff.	Benchmark	No delay credits	Diff.
Panel A. Overall						
Ages 62–65	49.7	47.8	-1.9	80.1	100.0	19.9
Ages 66–70	31.1	30.9	-0.3	92.9	100.0	7.1
Ages 62–70	39.6	38.6	-1.0	87.1	100.0	12.9

Panel B. Pre-FRA employment by asset group (ages 62–65)

	Employment rate (%)		
	Benchmark	No delay credits	Diff.
low_assets	43.8	43.8	0.0
mid_assets	48.4	48.4	-0.0
high_assets	60.0	52.4	-7.6

Notes: The counterfactual sets the claim-age adjustment scale to zero while holding work-linked earnings-record recomputation and all other calibrated parameters fixed. Differences are in percentage points.

4.1.5 Unexpected Holiday Shock as MIT Shock

The natural counterpart to the policy episode is an unexpected payroll-tax shock. We therefore build a calendar-year sequence with overlapping cohorts. Each year, a new equal-mass

cohort enters the model at age 62. Cohorts already in the model follow the no-holiday benchmark through 2010 and are hit by the holiday unexpectedly at the start of 2011. Cohorts entering in 2011 and 2012 face the lower tax at entry. We then aggregate outcomes over ages 62–70 within each calendar year, matching the reduced-form sample more closely than the single-cohort exercise.

We model the 2011 payroll-tax holiday as an MIT shock. Before 2011, agents do not anticipate the reform and all cohorts evolve under the no-holiday benchmark with a constant 6.2 percent employee payroll tax. At the start of 2011, the tax path switches unexpectedly to 4.2 percent in 2011 and 2012 and then returns to 6.2 percent from 2013 onward; once the shock arrives, agents correctly foresee this temporary path. For cohorts already in the model in 2011, we take the full benchmark state distribution at the end of 2010, including employment status, claiming status, health, assets, preferences, and idiosyncratic shocks, and use it as the initial condition for the post-shock continuation problem. Cohorts that enter at age 62 in 2011 or 2012 face the lower tax upon entry. We then aggregate outcomes across overlapping cohorts ages 62 to 70 within each calendar year. This construction removes anticipatory responses before 2011 and makes the model counterpart line up with the calendar-year policy timing in the data.

Computational algorithm. The exercise is a direct nonlinear transition calculation. For each cohort-entry year, we proceed in five steps. First, we solve the benchmark dynamic program by backward induction under the no-holiday path, with a constant 6.2 percent employee payroll tax. Second, we solve the shocked dynamic program under the temporary holiday path for that cohort's calendar years: 4.2 percent in 2011 and 2012 and 6.2 percent otherwise. Third, we simulate the benchmark economy forward. If the cohort is already alive in 2011, we stop the benchmark simulation at the end of 2010 and record the full cross-sectional state, including active unclaimed, active claimed, retired unclaimed, and retired claimed masses by health, asset, preference, and wage-shock states. Fourth, we use that benchmark state as the initial condition for the shocked continuation problem starting in 2011. Cohorts that enter in 2011 or later simply enter directly under the shocked policy path. Fifth, we aggregate benchmark and shocked outcomes across equal-mass overlapping cohorts and compute calendar-year employment for ages 62–70. This algorithm makes the MIT timing explicit: the reform is unanticipated before 2011, but once it arrives agents understand the full temporary path.

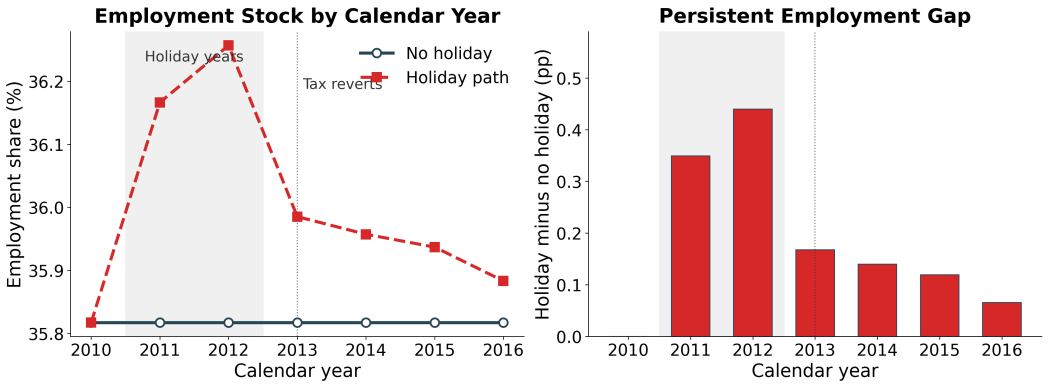


Figure 8: Unexpected 2011–2012 holiday with overlapping cohorts. Each calendar year aggregates equal-mass cohorts ages 62–70. Panel (1) compares the employment path under no holiday and under the actual holiday with reversion in 2013. Panel (2) plots the resulting employment gap. Because retirement is absorbing, the gap remains positive after the tax reverts.

Results. Figure 8 displays the result. Panel (1) plots the employment stock under the no-holiday benchmark (flat at about 35.8 percent) against the holiday path. During the holiday years, employment rises only modestly—from 35.8 to 35.9 percent in 2011 and remaining slightly above baseline in 2012—as new age-62 entrants face a lower tax at the main retirement margin. Panel (2) shows the corresponding gap: 0.12 percentage points in 2011 and 0.07 points in 2012.

The key feature is what happens after the tax reverts in 2013. The gap falls to about 0.02 percentage points in 2013 and is essentially gone by 2016, though it remains slightly positive throughout. This is still consistent with the stock-gap persistence from Section 2, but quantitatively the persistence is small in the calibrated annual model. Retirement is irreversible: once a senior exits, she does not return. The holiday therefore leaves behind a slightly larger employment stock after reversion, even though the resulting effect is limited.

5 Conclusion

This paper asks how much Social Security transforms a common payroll-tax reform into different effective work incentives for older workers, and whether those differences matter for retirement behavior. For seniors, the statutory payroll-tax rate is not the right policy object. What matters is the effective wedge on current work after earnings-record recomputation, claim status, and the Retirement Earnings Test are taken into account. Once these institutional features are incorporated, a common payroll-tax reform no longer implies a common incentive change.

Using the 2011–2012 payroll-tax holiday and monthly SIPP data on workers ages 62–70, we show that retirement responses vary systematically with baseline tax exposure, claiming status, and the strength of work-linked marginal benefit recomputation evaluated at a fixed claim age. A conditional post-62 dynamic mechanism model organized around the same effective-wedge object quantifies the policy margins behind this heterogeneity: a permanent payroll-tax cut works mainly through substitution, removing the RET current-value wedge raises employment sharply among low-asset early claimers, shutting down work-linked recomputation lowers employment among seniors with room to delay, and because retirement is absorbing, employment gains from the temporary holiday persist after the tax reverts.

This paper's contribution is therefore not simply to show that Social Security matters for retirement, nor to show that stronger contribution-benefit linkage affects labor supply. The contribution is to show that, for seniors, the existing U.S. Social Security system makes the incidence of a common payroll-tax reform fundamentally heterogeneous by converting one statutory tax change into a distribution of effective wedges. That shift in perspective matters for both empirical work and policy analysis. A single senior labor-supply elasticity is not enough when the relevant object is a distribution of wedges. For older workers, payroll-tax policy cannot be evaluated from the statutory rate alone. It must be evaluated through the institutional mapping from the statutory rate to the effective return to current work. Because the model is deliberately partial equilibrium and conditional on post-62 states, we read this contribution as a mechanism result about labor-supply incidence, not as a full welfare analysis of retirement and claiming.

A Omitted Proofs

Assumption 2 (Regularity for monotone comparative statics)**.** *The conditional law of ν_{t+1} under $\mathbb{E}_t[\cdot]$ may depend on the current date t, but it is otherwise independent of past work-cost histories and of (g, α, τ). In addition, $V^R(t) \geq 0$ for every t.*

Proof. Throughout the proof, write

$$x_+ \equiv \max\{x, 0\}.$$

By (13), (15), and (14), the continuation branch in the Bellman equation can be written as

$$w(1-\tau) + A_g(t,\alpha) - \nu_t + \beta\sigma_t(\alpha)\mathbb{E}_t\big[V_g^W(t+1, \nu_{t+1}, \alpha; \tau)\big] = V^R(t) + \nu_g^*(t, \tau, \alpha) - \nu_t.$$

Hence, for every $t \in \{0, \ldots, N-1\}$,

$$V_g^W(t, \nu_t, \alpha; \tau) = V^R(t) + \big(\nu_g^*(t, \tau, \alpha) - \nu_t\big)_+. \tag{35}$$

We use (35) repeatedly.

Part (i). Define the threshold gap

$$G_t(\alpha; \tau) \equiv \nu_u^*(t, \tau, \alpha) - \nu_c^*(t, \tau, \alpha).$$

At $t = N-1$, the terminal condition $V_u^W(N, \nu_N, \alpha; \tau) = V_c^W(N, \nu_N, \alpha; \tau) = V^R(N)$ implies

$$G_{N-1}(\alpha; \tau) = A_u(N-1, \alpha) - A_c(N-1, \alpha).$$

By the maintained work-margin ordering condition in Proposition 1,

$$A_u(t, \alpha) - A_c(t, \alpha) > 0,$$

so $G_{N-1}(\alpha; \tau) > 0$.

Now suppose $G_{t+1}(\alpha; \tau) > 0$ for some $t \leq N-2$. By (35), for every realization of ν_{t+1},

$$V_u^W(t+1, \nu_{t+1}, \alpha; \tau) - V_c^W(t+1, \nu_{t+1}, \alpha; \tau) = \big(\nu_u^*(t+1, \tau, \alpha) - \nu_{t+1}\big)_+ - \big(\nu_c^*(t+1, \tau, \alpha) - \nu_{t+1}\big)_+ \geq 0,$$

because $\nu_u^*(t+1,\tau,\alpha) > \nu_c^*(t+1,\tau,\alpha)$. Therefore,

$$\mathbb{E}_t\big[V_u^W(t+1,\nu_{t+1},\alpha;\tau) - V_c^W(t+1,\nu_{t+1},\alpha;\tau)\big] \geq 0.$$

Using (14),

$$G_t(\alpha;\tau) = \big[A_u(t,\alpha) - A_c(t,\alpha)\big] + \beta\sigma_t(\alpha)\mathbb{E}_t\big[V_u^W(t+1,\nu_{t+1},\alpha;\tau) - V_c^W(t+1,\nu_{t+1},\alpha;\tau)\big] > 0.$$

Backward induction yields $G_t(\alpha;\tau) > 0$ for every $t \in \{0,\ldots,N-1\}$, which is (16). Finally, (35) implies the pointwise weak value ordering

$$V_u^W(t,\nu_t,\alpha;\tau) \geq V_c^W(t,\nu_t,\alpha;\tau) \qquad \text{for all } (t,\nu_t,\alpha,\tau).$$

Thus unclaimed workers tolerate strictly higher work disutility before exiting.

Part (ii): monotonicity of each threshold in productivity type. Fix $g \in \{c,u\}$ and $\alpha' \geq \alpha$. We prove by backward induction that

$$\nu_g^*(t,\tau,\alpha') \geq \nu_g^*(t,\tau,\alpha) \qquad \text{for every } t.$$

By the maintained assumptions in Proposition 1, each $A_g(t,\alpha)$ is weakly increasing in α for every g.

At $t = N-1$,

$$\nu_g^*(N-1,\tau,\alpha') - \nu_g^*(N-1,\tau,\alpha) = \big[A_g(N-1,\alpha') - A_g(N-1,\alpha)\big] + \beta\big[\sigma_{N-1}(\alpha') - \sigma_{N-1}(\alpha)\big]V^R(N) \geq 0,$$

where the inequality uses $A_g(N-1,\alpha') \geq A_g(N-1,\alpha)$ and Assumption 2.

Now suppose

$$\nu_g^*(t+1,\tau,\alpha') \geq \nu_g^*(t+1,\tau,\alpha)$$

for some $t \leq N-2$. By (35), for every realization of ν_{t+1},

$$V_g^W(t+1,\nu_{t+1},\alpha';\tau) - V_g^W(t+1,\nu_{t+1},\alpha;\tau) = \big(\nu_g^*(t+1,\tau,\alpha') - \nu_{t+1}\big)_+ - \big(\nu_g^*(t+1,\tau,\alpha) - \nu_{t+1}\big)_+ \geq 0.$$

Since the law of ν_{t+1} is independent of α by Assumption 2,

$$\Psi_g(t,\alpha';\tau) \equiv \mathbb{E}_t\big[V_g^W(t+1,\nu_{t+1},\alpha';\tau)\big] \geq \mathbb{E}_t\big[V_g^W(t+1,\nu_{t+1},\alpha;\tau)\big] \equiv \Psi_g(t,\alpha;\tau).$$

Moreover, (35) and Assumption 2 imply $\Psi_g(t,\alpha';\tau) \geq V^R(t+1) \geq 0$. Therefore,

$$\sigma_t(\alpha')\Psi_g(t,\alpha';\tau) - \sigma_t(\alpha)\Psi_g(t,\alpha;\tau) = \big[\sigma_t(\alpha') - \sigma_t(\alpha)\big]\Psi_g(t,\alpha';\tau) + \sigma_t(\alpha)\big[\Psi_g(t,\alpha';\tau) - \Psi_g(t,\alpha;\tau)\big] \geq 0.$$

Combining this with $A_g(t,\alpha') \geq A_g(t,\alpha)$ in (14) gives

$$\nu_g^*(t,\tau,\alpha') \geq \nu_g^*(t,\tau,\alpha).$$

This proves (17).

Part (ii): monotonicity of the threshold gap in productivity type. We next show that $G_t(\alpha;\tau)$ is weakly increasing in α. At $t = N-1$,

$$G_{N-1}(\alpha;\tau) = A_u(N-1,\alpha) - A_c(N-1,\alpha),$$

which is weakly increasing in α by the maintained assumption in Proposition 1.

Now fix $t \leq N-2$ and assume that at date $t+1$ both thresholds are weakly increasing in α and the gap $G_{t+1}(\alpha;\tau)$ is weakly increasing in α. For a fixed realization of ν_{t+1}, define

$$D_{t+1}(\nu_{t+1},\alpha;\tau) \equiv V_u^W(t+1,\nu_{t+1},\alpha;\tau) - V_c^W(t+1,\nu_{t+1},\alpha;\tau).$$

By (35),

$$D_{t+1}(\nu_{t+1},\alpha;\tau) = \big(\nu_u^*(t+1,\tau,\alpha) - \nu_{t+1}\big)_+ - \big(\nu_c^*(t+1,\tau,\alpha) - \nu_{t+1}\big)_+.$$

Using part (i), this can be rewritten as

$$D_{t+1}(\nu_{t+1},\alpha;\tau) = \begin{cases} G_{t+1}(\alpha;\tau), & \nu_{t+1} < \nu_c^*(t+1,\tau,\alpha), \\ \nu_u^*(t+1,\tau,\alpha) - \nu_{t+1}, & \nu_c^*(t+1,\tau,\alpha) \leq \nu_{t+1} < \nu_u^*(t+1,\tau,\alpha), \\ 0, & \nu_{t+1} \geq \nu_u^*(t+1,\tau,\alpha). \end{cases}$$

Because $\nu_u^*(t+1,\tau,\alpha)$ and $G_{t+1}(\alpha;\tau)$ are weakly increasing in α, a direct check of the three regions shows that $D_{t+1}(\nu_{t+1},\alpha;\tau)$ is weakly increasing in α for every fixed ν_{t+1}. Since $D_{t+1}(\nu_{t+1},\alpha;\tau) \geq 0$ by part (i), Assumption 2 gives

$$\mathbb{E}_t\big[D_{t+1}(\nu_{t+1},\alpha';\tau)\big] \geq \mathbb{E}_t\big[D_{t+1}(\nu_{t+1},\alpha;\tau)\big] \geq 0.$$

Therefore,

$$\begin{aligned} G_t(\alpha';\tau) - G_t(\alpha;\tau) &= \big[A_u(t,\alpha') - A_c(t,\alpha')\big] - \big[A_u(t,\alpha) - A_c(t,\alpha)\big] \\ &\quad + \beta\Big(\sigma_t(\alpha')\mathbb{E}_t\big[D_{t+1}(\nu_{t+1},\alpha';\tau)\big] - \sigma_t(\alpha)\mathbb{E}_t\big[D_{t+1}(\nu_{t+1},\alpha;\tau)\big]\Big) \\ &\geq 0, \end{aligned}$$

where the last inequality uses the maintained monotonicity of $A_u(t,\alpha) - A_c(t,\alpha)$, the monotonicity of D_{t+1} just established, and $\sigma_t(\alpha') \geq \sigma_t(\alpha)$. This proves (18).

Part (iii). Fix $g \in \{c,u\}$ and $\tau' \geq \tau$. We prove by backward induction that

$$\nu_g^*(t,\tau',\alpha) \leq \nu_g^*(t,\tau,\alpha) \qquad \text{for every } t.$$

At $t = N-1$,

$$\nu_g^*(N-1,\tau',\alpha) - \nu_g^*(N-1,\tau,\alpha) = -w(\tau'-\tau) \leq 0.$$

Now suppose $\nu_g^*(t+1,\tau',\alpha) \leq \nu_g^*(t+1,\tau,\alpha)$ for some $t \leq N-2$. By (35), for every realization of ν_{t+1},

$$V_g^W(t+1,\nu_{t+1},\alpha;\tau') \leq V_g^W(t+1,\nu_{t+1},\alpha;\tau).$$

Since the law of ν_{t+1} is independent of τ by Assumption 2,

$$\Delta_g(t,\tau',\alpha) = \beta\sigma_t(\alpha)\mathbb{E}_t\big[V_g^W(t+1,\nu_{t+1},\alpha;\tau')\big] - V^R(t) \leq \beta\sigma_t(\alpha)\mathbb{E}_t\big[V_g^W(t+1,\nu_{t+1},\alpha;\tau)\big] - V^R(t) = \Delta_g(t,\tau,\alpha).$$

Substituting into (14) gives

$$\nu_g^*(t,\tau',\alpha) - \nu_g^*(t,\tau,\alpha) = -w(\tau'-\tau) + \Delta_g(t,\tau',\alpha) - \Delta_g(t,\tau,\alpha) \leq 0.$$

Backward induction proves (19). □

Proof of Proposition 4.

Proof. We prove each part in turn under the maintained assumption that, after each period, the latent work-cost draw is conditionally history-independent and therefore follows the same post-holiday law on the holiday and baseline paths once (g,α,t) are fixed.

Part (i): By Proposition 1(iii), the tax cut raises the continuation threshold: $\nu_g^*(t,\tau-\delta,\alpha) \geq \nu_g^*(t,\tau,\alpha)$. Since a worker continues if and only if $\nu_t \leq \nu_g^*$, the continuation probability satis-

fies

$$P_g(C_t = 1 \mid \tau - \delta, \alpha) \;=\; F_g\big(\nu_g^*(t, \tau - \delta, \alpha) \mid t\big) \;\geq\; F_g\big(\nu_g^*(t, \tau, \alpha) \mid t\big) \;=\; P_g(C_t = 1 \mid \tau, \alpha),$$

with strict inequality whenever $f_g > 0$ at the threshold. The two paths coincide up to date t, so $\ell_g^{\text{hol}}(t \mid \alpha) = \ell_g^{\text{base}}(t \mid \alpha)$. Applying the stock recursion (29),

$$\begin{aligned} \ell_g^{\text{hol}}(t{+}1 \mid \alpha) &= \ell_g^{\text{hol}}(t \mid \alpha) \;\cdot\; P_g(C_t{=}1 \mid \tau{-}\delta, \alpha) \\ &> \ell_g^{\text{base}}(t \mid \alpha) \;\cdot\; P_g(C_t{=}1 \mid \tau, \alpha) \\ &= \ell_g^{\text{base}}(t{+}1 \mid \alpha). \end{aligned} \tag{36}$$

Part (ii): After date t the tax reverts to τ on both paths. By Assumption 2, workers retained by the holiday face the same subsequent draw process as other employed workers of the same type, so the continuation probabilities coincide for every $r \geq 1$. Write $\mathcal{P}_s \equiv \prod_{r=1}^{s-1} P_g(C_{t+r} = 1 \mid \tau, \alpha)$ for the common post-holiday survival factor. Iterating (29) from $t{+}1$ to $t{+}s$ on each path,

$$\begin{aligned} \ell_g^{\text{hol}}(t{+}s \mid \alpha) &= \ell_g^{\text{hol}}(t{+}1 \mid \alpha)\, \mathcal{P}_s, \\ \ell_g^{\text{base}}(t{+}s \mid \alpha) &= \ell_g^{\text{base}}(t{+}1 \mid \alpha)\, \mathcal{P}_s. \end{aligned}$$

Subtracting,

$$\ell_g^{\text{hol}}(t{+}s \mid \alpha) - \ell_g^{\text{base}}(t{+}s \mid \alpha) \;=\; \underbrace{\Big[\ell_g^{\text{hol}}(t{+}1 \mid \alpha) - \ell_g^{\text{base}}(t{+}1 \mid \alpha)\Big]}_{>0 \text{ by } (36)} \cdot\, \mathcal{P}_s, \tag{37}$$

which is (30). Each factor in $\mathcal{P}_s$ lies in $(0, 1]$, so the gap is strictly positive for all $s \geq 1$ but weakly shrinking. Crucially, the gap decays only through subsequent exit, never through reversal, because workers retained by the holiday cannot be selectively removed once the tax reverts.

Part (iii): When the tax reverts to τ at date $t{+}1$, the continuation probabilities are identical on both paths, so no new gap accrues. The stock gap from (37) persists at level $[\ell_g^{\text{hol}}(t{+}1 \mid \alpha) - \ell_g^{\text{base}}(t{+}1 \mid \alpha)] \cdot \mathcal{P}_s$, shrinking only through attrition.

If instead the tax is raised to $\tau{+}\epsilon$ upon restoration, the continuation probability at $t{+}1$ falls below its baseline level $P_g(C_{t+1} = 1 \mid \tau, \alpha)$. This negative shock acts on the larger holiday-path stock $\ell_g^{\text{hol}}(t{+}1 \mid \alpha)$. The force behind the overshoot is therefore the larger stock exposed to the restoration, not a different post-holiday law of future work-cost draws. □

B SIPP Robustness and Supplemental Evidence

This appendix collects supplementary sample-construction details, variable definitions, diagnostics, and robustness checks referenced in Section 3.3.

B.1 Sample Construction

We begin from all SIPP monthly records and restrict the analysis to respondents ages 62–70, the age range where both labor-supply adjustment and Social Security claiming behavior are economically relevant for our setting. For the main reduced-form estimates, we keep calendar years 2010–2012. We require non-missing labor outcomes and valid claiming assignment in each person-month used for estimation. In specifications that interact treatment with survival heterogeneity, we further require a non-missing baseline transported survival measure constructed from 2010 information, so heterogeneity is fixed before policy exposure.

The resulting panel is unbalanced at the month level because respondents can rotate out of interview months or have missing items in specific months. The broad age-62–70 sample over January 2010 through December 2012 contains 270,269 person-months from 12,036 individuals. The estimating sample is smaller in particular specifications when we impose additional baseline-earnings, claiming-status, survival, or risk-set restrictions.

B.2 Key Variables

Outcomes. Our primary outcome is monthly employment status. We set

$$\text{Employed}_{it} = \mathbf{1}\{\text{employed}_{it}\},$$

so the indicator equals one for respondents reporting that they are employed (either at work or with a job) and zero otherwise.

Our secondary (corroborative) labor-supply outcome is monthly hours. Let h_{it}^{w} denote mapped weekly hours from the weekly-hours response category, and let n_{it} denote reported weeks worked in the month. We define

$$\text{Hours}_{it} = \begin{cases} 0 & \text{if not employed,} \\ h_{it}^{w} \times n_{it} & \text{if employed,} \end{cases}$$

where h_{it}^{w} maps the reported weekly-hours category to a weekly-hour point.[3] For descriptive

[3] In the SIPP 2008 recode, the weekly-hours categories are: 0 = none, 1 = all weeks 35+ hours, 2 = all weeks

dynamics we also use a 3-month moving average of monthly hours. Because hours are recovered from categorical bins and are more exposed to seam-related measurement noise, we treat hours estimates as corroborative rather than primary.

Income measures and units. Monthly total income is measured by `TPMSUM1`, and monthly Social Security income is measured by `THSOCSEC`. Both are recorded in nominal U.S. dollars for the SIPP reference month. In interpretation, `THSOCSEC` is a household-level Social Security income aggregate rather than a pure own-benefit flow, so non-claimers can still show positive values when another household member (for example, a spouse) receives benefits.

Policy timing indicators. Policy exposure is encoded with calendar-time indicators. For the payroll-tax holiday period,

$$\text{Post}_t = \mathbf{1}\{\text{year} \in \{2011, 2012\}\}.$$

Thus the main estimation window is January 2010 through December 2012, with $\text{Post}_t = 0$ in 2010 and $\text{Post}_t = 1$ in 2011–2012.

Pre-policy claiming status. PreClaimer_i equals one if respondent i reports receiving own Social Security retirement benefits in the latest observed month of 2010, and zero otherwise. This status is fixed before policy exposure and aligns with the January 2010 start of the monthly analysis window.

Survival heterogeneity proxy in SIPP. SIPP does not include linked mortality outcomes over the policy window, so we construct S_i via external transport from HRS in three steps. The construction follows two related literatures: mortality-risk indexing in older-adult/HRS settings (Lee et al., 2006; Cruz et al., 2013) and transport of source-sample prediction models to target populations with covariates but no observed outcomes (Steingrimsson et al., 2023).

First, in RAND HRS we estimate a wave-to-wave mortality model on ages 59–72 (waves 1–14), keeping respondents interviewed at wave w with known alive/dead status at $w + 1$. With

1–34 hours, 3–4 = a mix of 35+ and 1–34 hours, 5 = some weeks 35+ hours and otherwise 0, and 6 = some weeks 1–34 hours and otherwise 0. We convert these categories to continuous weekly-hour points as $0 \mapsto 0$, $\{1, 5\} \mapsto 40$, $\{2, 6\} \mapsto 17.5$, and $\{3, 4\} \mapsto 28.75$. The value 17.5 is the midpoint of the 1–34 range, and 28.75 is the midpoint between 17.5 and 40 for mixed-hour categories. Monthly hours are then this mapped weekly value times the reported number of weeks worked in the month.

$D_{i,w+1} = 1$ if next-wave status is death, we estimate a weighted probit:

$$\Pr(D_{i,w+1} = 1 \mid X_{iw}) = \Phi\big(\alpha + \beta_1 \text{Age}_{iw} + \beta_2 \text{Age}_{iw}^2 + \beta_3 \text{Female}_i + \beta_4 \text{Disabled}_{iw} + \beta_5 \text{SSI}_{iw}\big)\,.$$

All coefficients are estimated in HRS and then fixed before SIPP analysis.

Second, we map harmonized SIPP covariates (age, sex, disability status, and SSI receipt) into X_{it} and compute month-level transported mortality over the model horizon (2 years):

$$\widehat{m}_{it}^{(2)} = \Phi(\widehat{\alpha} + X_{it}'\widehat{\beta}).$$

We convert to a 10-year horizon using a constant-hazard mapping:

$$\widehat{h}_{it} = -\frac{\log(1 - \widehat{m}_{it}^{(2)})}{2}, \qquad \widehat{m}_{it}^{(10)} = 1 - \exp(-10\widehat{h}_{it}), \qquad \widehat{s}_{it}^{(10)} = 1 - \widehat{m}_{it}^{(10)}.$$

Third, to keep heterogeneity assignment predetermined, we collapse to person-level baseline risk using 2010 months only:

$$\overline{m}_{i,2010}^{(10)} = \frac{1}{N_{i,2010}} \sum_{t \in 2010} \widehat{m}_{it}^{(10)}, \qquad S_i \equiv \overline{s}_{i,2010}^{(10)} = 1 - \overline{m}_{i,2010}^{(10)}.$$

This baseline S_i is then merged to all analysis months and centered in the regressions. In grouped analyses, high/low mortality groups are defined by the top/bottom quartiles of $\overline{m}_{i,2010}^{(10)}$ ($\geq q_{0.75}$ and $\leq q_{0.25}$, respectively).

Relative to the common HRS approach of using respondents' own subjective survival expectations measured in the same sample (Hurd and McGarry, 1995, 2002; Hurd et al., 2004), our design differs in two ways: (i) mortality risk is generated out-of-sample (estimated in HRS, applied in SIPP), and (ii) because the target SIPP panel does not observe mortality over the policy window, we do not re-estimate or re-calibrate on SIPP outcomes and instead fix heterogeneity at the 2010 baseline.

B.3 Conditional Parallel-Trends Diagnostics

The main SIPP analysis uses three common-shock exposure designs. In each case, the holiday starts at the same calendar date for everyone, and identification comes from predetermined differences in exposure. Figures 9 to 11 report the corresponding timing diagnostics. For the tax-exposure design, the cleanest diagnostic is the monthly coefficient event study from the regression itself. For the claiming-status and λ_0 designs, the appendix shows treatment and

comparison group paths, with the λ_0 companion aggregated to quarterly frequency to make the underlying trend easier to see. Because the estimation window begins in January 2010, the pre-period is short, so these figures are diagnostics rather than formal proof. They are meant to show whether the relevant groups were moving similarly before the holiday and how those paths diverged after the reform began in January 2011.

The first and third designs show broadly similar pre-period paths. The claiming-status comparison is less flat in 2010, so that design should be read together with the rest of the evidence rather than in isolation. That is one reason the paper emphasizes the full pattern across the three designs, especially the continuous-λ_0 result, instead of leaning on any single coefficient.

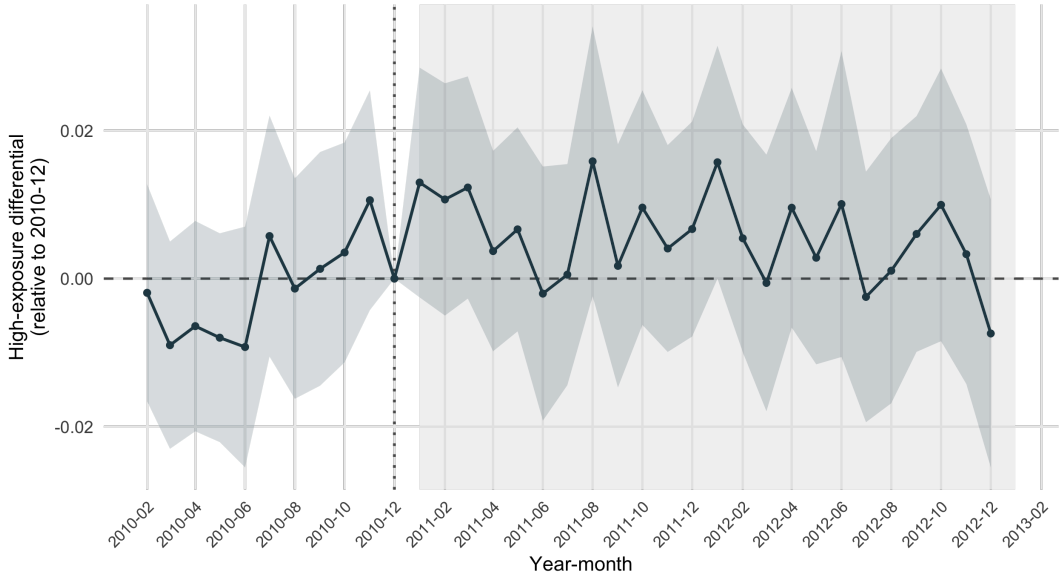


Figure 9: Monthly Event-Study Coefficients: Tax-Exposure Design

Notes: The figure plots the month-specific coefficients on $1\{t = m\} \times \text{HighExposure}_i$ from the Fact 1 event-study specification. The regression includes individual fixed effects, year-month fixed effects, age fixed effects, and month-by-pre-claimer interactions. After estimation, the full series is explicitly normalized so that December 2010 equals zero, and every other month is plotted relative to that benchmark. The vertical line marks January 2011, the first month of the payroll-tax holiday.

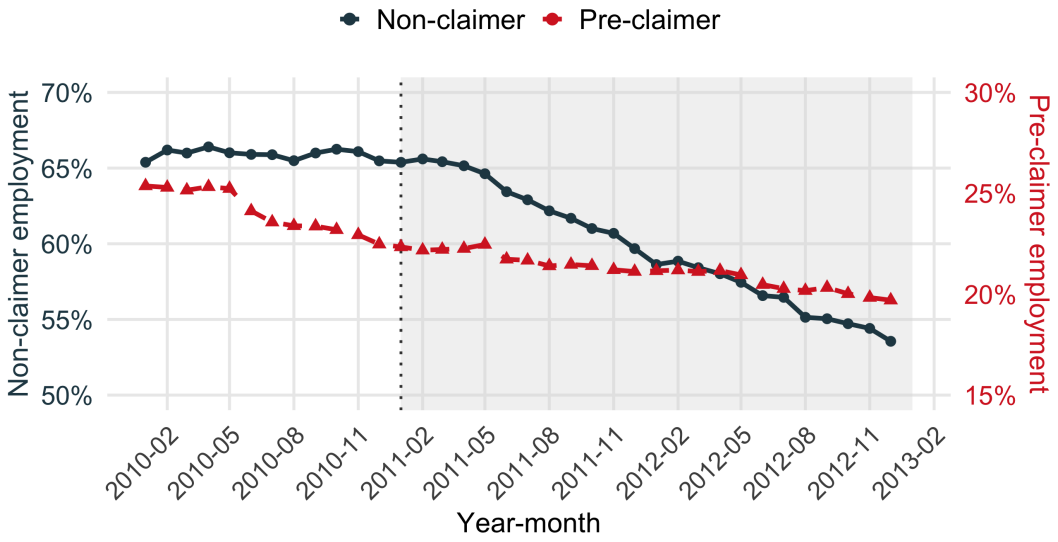


Figure 10: Monthly Employment Trends by Pre-policy Claiming Status

Notes: The figure plots monthly employment rates for pre-claimers and non-claimers in the sample used for Fact 2. The vertical line marks January 2011, the first month of the payroll-tax holiday. These are descriptive monthly paths; the main regression adds person fixed effects, year-month fixed effects, and the survival interactions discussed in the text.

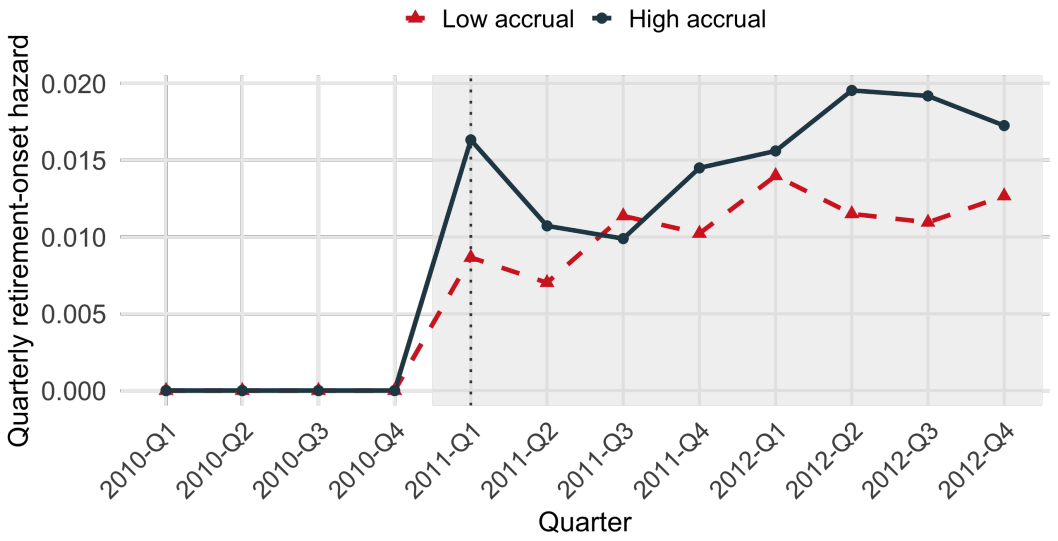


Figure 11: Quarterly Retirement-Onset Hazard: High-λ_0 vs. Low-λ_0 Among Unclaimed Workers

Notes: The figure plots quarterly retirement-onset hazards for unclaimed workers above and below the median baseline λ_0. The vertical line marks 2011:Q1, the first quarter of the payroll-tax holiday. The main specification in Fact 3 uses continuous λ_0, so this median split is a descriptive companion rather than the estimating equation itself.

Table 8: Companion Claiming-Onset and Retirement Hazards Around the Holic

Outcome	Comparison	Control hazard (%)	Treated hazard (%)	Diff
Claiming onset hazard	Episode A: 2011-2012 vs 2010	3.02	3.00	-0.
Claiming onset hazard	Episode B: 2013 vs 2012	3.06	2.45	-0.
Retirement hazard (employment exit)	Episode A: 2011-2012 vs 2010	2.48	2.20	-0.
Retirement hazard (employment exit)	Episode B: 2013 vs 2012	2.21	2.09	-0.

Notes: Episode A compares 2011–2012 with 2010; Episode B compares 2013 with 2012. The key object for the revised interpretation moves little relative to the retirement hazard, which supports reading the main holiday estimates as labor-supply responses cond as a pure claiming-margin effect.

Table 9: λ_0 Robustness to Alternative Fixed Claim Ages

	(1) FRA-fixed	(2) Age+1	(3) Modal 67
Post $\times$ λ_0 (centered)	0.6232**	0.4041	0.3452
SE	(0.3098)	(0.2969)	(0.2410)
95% CI	[0.0160, 1.2303]	[-0.1778, 0.9859]	[-0.1272, 0.8177]
Observations	44,321	44,321	44,321
Individuals	1,758	1,758	1,758
Spec: S2 (+ demographics), individual + year-month FE			
Sample: Unclaimed workers with observed earnings; baseline λ_0 is FRA-fixed.			

Notes: The baseline specification evaluates λ_0 at FRA, so it is a pure earnings-record recomputation slope. The alternatives instead fix claim age at $a + 1$ or at 67. The sign and magnitude are similar across definitions, showing that the continuous-λ_0 result does not rely on a particular fixed-claim-age normalization.

B.4 Aggregate Retirement Claiming: SSA Administrative Evidence

Table 10 reports the number of retired-worker awards (new claims for Social Security retired-worker benefits) from SSA administrative records (Annual Statistical Supplement, Table 6.B5) for years 2005–2017. The table shows annual totals (in thousands) and the percentage of awards at age 62, separately by sex.

Table 10: Retired-Worker Awards by Year, 2005–2017 (SSA Table 6.B5)

	Awards (thousands)			Average age at award			% claiming at 62	
Year	Men	Women	Total	Men	Women	All	Men	Women
2005	1,061	939	2,000	63.6	63.6	63.6	50.2	54.3
2006	1,057	942	1,999	63.8	63.8	63.8	45.6	50.2
2007	1,075	961	2,036	64.0	63.9	64.0	42.4	47.6
2008	1,200	1,079	2,279	64.0	63.9	64.0	42.5	48.3
2009	1,452	1,288	2,740	64.0	63.9	64.0	44.3	49.8
2010	1,387	1,248	2,635	64.1	63.9	64.0	42.6	47.9
2011	1,340	1,238	2,578	64.0	63.8	63.9	41.4	46.5
2012	1,419	1,316	2,735	64.2	64.0	64.1	37.2	42.4
2013	1,447	1,347	2,794	64.3	64.2	64.3	34.8	40.2
2014	1,433	1,339	2,772	64.4	64.2	64.3	34.5	39.7
2015	1,475	1,364	2,839	64.5	64.4	64.5	31.9	37.4
2016	1,519	1,392	2,911	64.6	64.5	64.6	30.4	35.5
2017	1,531	1,444	2,975	64.7	64.6	64.7	29.0	32.9

Source: Social Security Administration, Annual Statistical Supplement, Table 6.B5 (Master Beneficiary Record, 100% data). Blue-shaded rows mark the payroll tax holiday period (2011–2012). Gray-shaded row marks the 2009 recession spike. "Awards" are new retired-worker benefit claims by calendar year of award action. "Average age" and "% claiming at 62" are computed from the age distribution reported in the same table. The "All" average age column is the unweighted mean of the sex-specific averages.

B.5 Inference and Sample-Composition Robustness

C Earnings Proxy Validation

The calibration in Section 3.4.4 assigns workers to PIA brackets using current or recent earnings observed in the SIPP, under the assumption that current earnings approximate a worker's position in the Social Security benefit formula. A potential concern is that current earnings may not accurately reflect lifetime earnings position (AIME): a high-wealth senior working part-time could have high lifetime earnings but low current earnings, leading to misclassification of the marginal replacement rate $r(e)$ and, consequently, the benefit accrual rate $\lambda_0(e)$.

This appendix uses administrative Social Security benefit data linked to HRS respondents to assess the validity of this earnings proxy.

Data. The RAND HRS Social Security Wealth file (`SSWEALTH`) contains SSA-computed present discounted values of projected benefit streams at ages 62, FRA, and 70, derived from administrative earnings records. We extract the wave-10 (year 2010) benefit wealth at FRA (`R10SSWRNR`), which we convert to an approximate monthly PIA using age-specific actuarial annuity factors.[4] From the RAND HRS core file, we extract wave-10 earnings (`r10iearn`), work status, hours, and household wealth.

We merge these files on `hhidpn` and restrict to a *broad sample* of working adults aged 60–69 with valid earnings and valid SSA benefit wealth ($N = 1{,}186$), and a *near-cap sample* further restricted to earnings within 0.8–1.2 times the 2010 taxable maximum of $\$106{,}800$ ($N = 99$).

Bracket classification. We classify each worker into a marginal PIA bracket by two methods. The *SSW-implied bracket* converts SSA benefit wealth at FRA to approximate monthly PIA and maps it to the 90%, 32%, or 15% bracket using the 2010 bend points ($b_1 = \$761$, $b_2 = \$4{,}586$) and the corresponding PIA kink thresholds ($\$684.90$ and $\$1{,}908.90$ per month). The *earnings-implied bracket* approximates monthly AIME as current annual earnings divided by 12 and maps it directly to the same bend points.

Results. Table 11 is easiest to read as five validation questions about the earnings proxy. Panel A asks whether current earnings and SSA-based lifetime benefit measures rank workers similarly. Panel B asks whether current earnings places workers in the same marginal PIA bracket as SSA data. Panel C asks a weaker but policy-relevant question: even if the exact 15%

[4] We use a baseline annuity factor of 12.7 at FRA (age 66, 2.9% real discount rate, SSA period life table for males), adjusted for each respondent's age at the time of interview: for ages below FRA, the factor is discounted by $(1+r)^{-(66-a)}$; for ages above FRA, the factor is reduced by 0.85 per year past FRA to approximate the shorter remaining benefit horizon. The annuity factor ranges from 10.7 (age 60) to 12.7 (age 66) in our sample.

versus 32% bracket is noisy near the taxable cap, are near-cap workers still concentrated near the top of the lifetime-earnings distribution? Panel D asks whether any mismatch is mainly a part-time-work problem. Panel E asks whether the mismatched cases look like unusually wealthy part-timers.

Table 11: Can Current Earnings Recover Lifetime Benefit Position? HRS Validation

Validation question	Statistic	Value
Panel A. Do current earnings and SSA benefit wealth rank workers similarly?		
Broad sample ($N = 1{,}186$)	Spearman ρ (earnings vs. SSA wealth)	0.516
	Pearson r	0.498
Near-cap sample ($N = 99$)	Spearman ρ	0.059
Panel B. Does current earnings place workers in the correct PIA bracket?		
Broad sample	Overall accuracy	64.1%
	15% bracket correct (of 385 classified)	47.8%
	32% bracket correct (of 676 classified)	80.5%
Near-cap sample	Fraction truly in 15% bracket (from SSA data)	58.6%
Panel C. Are near-cap workers still in the top of the SSA distribution?		
Near-cap workers above the 80th percentile of SSA wealth		51.5%
Near-cap workers above the 60th percentile of SSA wealth		69.7%
Panel D. Is mismatch mainly a part-time-work problem?		
Full-time (≥ 35 hrs/wk, $N = 838$)	Bracket accuracy	65.5%
Part-time (< 35 hrs/wk, $N = 317$)	Bracket accuracy	62.2%
Panel E. Are misclassified workers unusually wealthy?		
Correctly classified ($N = 760$)	Median household wealth	\$181,500
Misclassified ($N = 426$)	Median household wealth	\$202,310

Notes: Sample is HRS wave-10 (2010) male respondents aged 60–69 who are currently working with valid earnings and valid SSA benefit wealth from the `SSWEALTH` file. Earnings, SSA benefit wealth, and household wealth are winsorized at the 5th and 95th percentiles. SSA benefit wealth at FRA is converted to approximate monthly PIA using age-specific actuarial annuity factors (range 10.7–12.7). Bracket classification uses 2010 PIA bend points ($b_1 = \$761$, $b_2 = \$4{,}586$). The near-cap sample restricts to earnings within 0.8–1.2 times the taxable maximum ($\$106{,}800$). Panel C uses percentile ranks of SSA benefit wealth, which do not depend on the annuity factor conversion. The low correlation within the near-cap sample in Panel A is therefore not the main object, because earnings are intentionally compressed in that subsample. Full-time is defined as ≥ 35 hours per week at the primary job. Household wealth is total non-housing financial wealth (`h10atotb`).

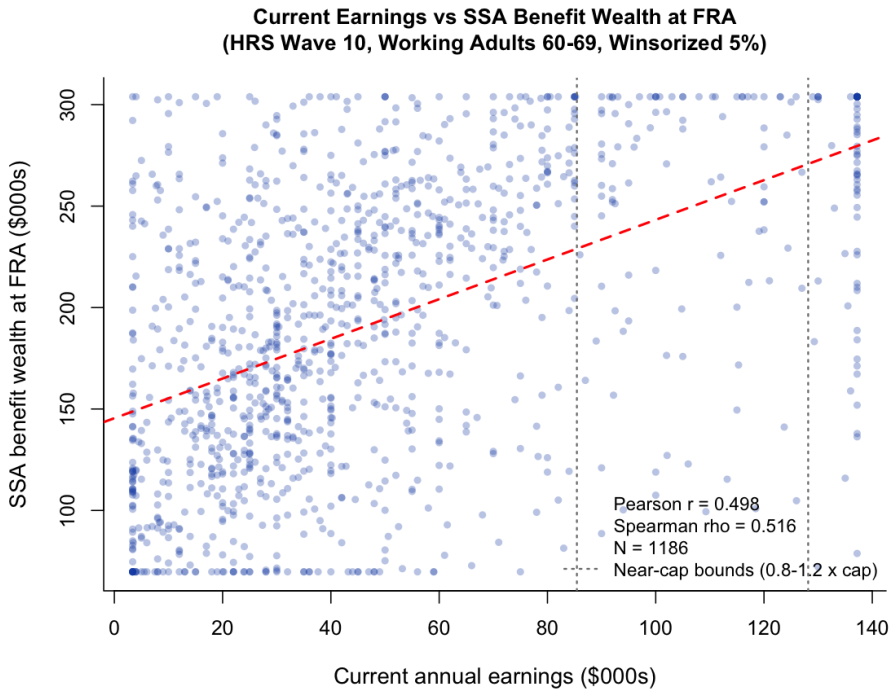


Figure 12: Current Earnings vs. SSA Benefit Wealth at FRA (HRS Wave 10)

Notes: Each point is a male HRS respondent aged 60–69 who is currently working, with both valid current earnings and valid SSA-linked benefit wealth at FRA from the `SSWEALTH` file ($N = 1{,}186$). Both variables are winsorized at the 5th and 95th percentiles. The dashed red line is the OLS fit. Vertical gray lines mark the near-cap region (0.8–1.2 times the 2010 taxable maximum of \$106,800). Spearman $\rho = 0.516$; Pearson $r = 0.498$.

Discussion. Three findings bear on the credibility of the earnings-based bracket assignment.

First, in the broad sample, current earnings and SSA benefit wealth move together in a meaningful way. The Spearman rank correlation is 0.516 and the Pearson correlation is 0.498, so current earnings preserve a substantial part of the ranking in lifetime benefit position.

Second, the near-cap rows should be read differently. Because current earnings are mechanically compressed near the taxable cap, the raw correlation in that subsample is not very informative. The relevant check is whether near-cap workers still look like high lifetime earners in the SSA data. They do: 51.5% are above the 80th percentile of SSA wealth and 69.7% are above the 60th percentile. So the proxy succeeds at the margin that matters for the paper: workers earning near the cap are overwhelmingly in the upper part of the lifetime earnings distribution, even if the exact 15% versus 32% split is noisy.

Third, the part-time concern is modest. Bracket accuracy among part-time workers (62.2%)

is only slightly lower than among full-time workers (65.5%), and misclassified workers have only modestly higher median household wealth ($202,310 vs. $181,500). That is weakly consistent with the "wealthy part-timer" story, but not by enough to overturn the main proxy.

Overall, the validation supports using current earnings as a proxy for PIA bracket position, particularly in the near-cap sample that drives the third empirical result. The rank correlation is moderate, the near-cap percentile concentration is strong, and the full-time/part-time differential is small.

D Calibration Tables and Figures

This appendix collects all tables and figures referenced in Section 3.5.

Table 12: Targeted Moments

Block	Moments	Count	Block weight
Age profile	Employment by age, 62–70	9	2.0
Health profile	Employment by health group, ages 62–70	4	1.0

Notes: Within the age block, ages 62–65 receive higher weight than ages 66–70.

Table 13: Fixed Parameters

Symbol	Parameter	Value
Panel A: Standard or Normalized Parameters		
β	Discount factor	0.98
b	Asset grid	$(0.0,\ 1.5,\ 5.0)$
ω_b	Asset weights	$(0.35,\ 0.40,\ 0.25)$
ν	Work-preference support	$(-1,\ 0,\ 1)$
ω_ν	Work-preference weights	$(0.25,\ 0.50,\ 0.25)$
τ_h, τ_b	Health and asset wage tilts	0.60, 0.45
ζ_z, ρ_z	Bad-shock wage scale and persistence	0.60, 0.75
s_x	Bad-shock exit scale	0.35
$\kappa_{62:65}$	Early work-cost profile	$(0.15,\ 0.12,\ 0.08,\ 0.05)$
Panel B: Policy Parameters		
a^*	Full Retirement Age	66
δ_D	Delayed retirement credit (per year)	0.08
θ, θ^{FRA}	RET withholding rates	0.50, 0.333
$\bar{y}, \bar{y}^{FRA}$	RET exempt earnings	\$14,160, \$37,680
$\bar{w}$	Taxable maximum	\$106,800
τ, τ'	Benchmark / counterfactual tax rates	0.062, 0.042
Panel C: Author-Constructed from Data		
ζ_w	Earnings support in joint type α	$(0.70,\ 1.00,\ 1.35)$
ω_w	Earnings-support weights in α	$(0.35,\ 0.45,\ 0.20)$
$r^z_{H1:H4}$	Health-profile support in α	$(1.000,\ 0.162,\ 0.110,\ 0.060)$

Notes: Panel A lists values fixed prior to SMM because they are either standard annual normalizations or reduced-form supports that are not separately identified by the targeted moments. We set $\beta = 0.98$ in line with annual retirement calibrations in French (2005); French and Jones (2011); Gustman and Steinmeier (2005). The asset grid is a reduced-form liquidity support because the simplified model has no savings choice and the current estimation sample does not contain a comparable wealth measure. The early work-cost profile is fixed to capture the steep decline in employment at ages 62–65 without adding more free parameters. Panel B reports statutory values from 2010 law. Panel C reports parameters we construct directly from the data. Together, these supports define the fixed joint productivity-health type α used in the theory section: α is shorthand for the same joint support already used in the quantitative model. The earnings support is a coarse three-point approximation to the 2010 SIPP earnings distribution of employed seniors: weighted annual earnings have mean \$41.9k, median \$27.2k, and 75th percentile \$52.6k, so the support $(\$21k, \$30k, \$40.5k)$ spans the center of the distribution while keeping the state space small. The health profile is anchored to the quartile-specific survival ordering from the HRS-based survival mapping. No numerical recalibration is implied by the relabeling to α.

Table 14: SMM-Estimated Parameters

Symbol	Parameter	Estimate
ψ_0	Retirement-utility intercept	0.9244
ψ_1	Age slope of work disutility	0.0402
ψ_2	Post-65 work-disutility increment	0.0064
κ_0	Fixed cost of work	0.5529
χ	Cost of remaining unclaimed	0.6359
q	Bad-shock arrival probability	0.1288
p_f	Bad-shock exit intensity	0.6447
κ_h	Health-work disutility scale in α	0.3047
r_{H2}	Health-component loading, group 2	0.0775
r_{H3}	Health-component loading, group 3	0.0536
r_{H4}	Health-component loading, best health	0.0394

Notes: The worst-health rank is normalized to one, so only (r_{H2}, r_{H3}, r_{H4}) are estimated. In the α notation, these parameters govern the health component of the fixed joint type. These parameters are jointly estimated by SMM from the 13 targeted employment moments.

Table 15: Fit to Targeted Moments

Employment by Age				*Employment by Health Group, 62–70*			
Age	Data	Model	Diff.	Group	Data	Model	Diff.
62	0.512	0.518	+0.007	$H1$	0.152	0.153	+0.001
63	0.461	0.465	+0.004	$H2$	0.381	0.348	−0.033
64	0.426	0.422	−0.004	$H3$	0.446	0.392	−0.054
65	0.351	0.383	+0.032	$H4$	0.477	0.436	−0.041
66	0.306	0.351	+0.045			*RMSE*	*0.038*
67	0.272	0.310	+0.038				
68	0.247	0.280	+0.033				
69	0.230	0.252	+0.022				
70	0.207	0.219	+0.012				
		RMSE	*0.026*				

Notes: The SMM criterion targets all 13 moments. The final weighted objective value is 0.00247. $H1$ denotes the worst health and $H4$ denotes the best health case.

Table 16: Permanent Payroll-Tax Cut: Employment Responses by Group

Margin	Group	Benchmark	Counterfactual	Change
Overall	Ages 62–70	0.358	0.361	$+0.003$
Age	62–65	0.447	0.448	$+0.001$
Age	66–70	0.282	0.287	$+0.005$
Health	$H1$ (worst)	0.153	0.154	$+0.001$
Health	$H2$	0.348	0.352	$+0.004$
Health	$H3$	0.392	0.395	$+0.002$
Health	$H4$ (best)	0.436	0.440	$+0.004$
Assets	Low	0.302	0.306	$+0.003$
Assets	Middle	0.339	0.342	$+0.003$
Assets	High	0.467	0.469	$+0.002$
Earnings	\$30k	0.356	0.359	$+0.003$
Earnings	\$60k	0.348	0.351	$+0.003$
Earnings	\$90k	0.339	0.344	$+0.005$
Earnings	\$110k	0.330	0.337	$+0.007$
Earnings	\$150k	0.327	0.330	$+0.003$

Notes: The overall row uses the benchmark age distribution over ages 62–70. Earnings rows come from auxiliary type-by-type counterfactuals that vary only the earnings component of the joint type α, holding all other calibrated parameters fixed.

Table 17: Channel Decomposition of the Employment Response

Panel A: By asset group					
Group	Total	Baseline	Waiting-cost	Wage-sorting	Interaction
Low	0.344	0.396	−0.048	−0.024	+0.020
Middle	0.311	0.396	−0.085	0.000	0.000
High	0.177	0.396	−0.151	−0.004	−0.064
Panel B: By health group					
Group	Total	Baseline	Longevity	Work-capacity	Wage-sorting
$H1$	0.059	0.516	−0.415	−0.022	−0.020
$H2$	0.403	0.516	−0.342	+0.196	+0.034
$H3$	0.235	0.516	+0.017	−0.318	+0.019
$H4$	0.413	0.516	+0.522	−0.692	+0.067

Notes: Entries are percentage-point changes in pooled employment, ages 62–70. In each panel, "Baseline" is the response when all panel-specific channels are shut off. Panel A channels: "Waiting-cost" is the extra contribution of asset-dependent claiming costs; "Wage-sorting" is the extra contribution of the asset tilt in the earnings component of joint-type composition; "Interaction" is the residual. Panel B channels: "Longevity" adds back health-specific survival differences; "Work-capacity" adds back health-related work disutility and bad-shock risk; "Wage-sorting" adds back the health tilt in the earnings component of joint-type composition.

References

Blundell, R., French, E., and Tetlow, G. (2016). Retirement incentives and labor supply. In Piggott, J. and Woodland, A., editors, *Handbook of the Economics of Population Aging*, volume 1, pages 457–566. Elsevier.

Chetty, R., Guren, A., Manoli, D., and Weber, A. (2013). Does indivisible labor explain the difference between micro and macro elasticities? a meta-analysis of extensive margin elasticities. *NBER Macroeconomics Annual*, 27:1–56.

Coile, C. and Gruber, J. (2007). Future social security entitlements and the retirement decision. *The Review of Economics and Statistics*, 89(2):234–246.

Coile, C. C. (2023). Changing retirement incentives and retirement in the US. In *Social Security Programs and Retirement around the World: The Effects of Reforms on Retirement Behavior*, chapter 12. University of Chicago Press.

Cruz, M., Covinsky, K., Widera, E. W., Stijacic-Cenzer, I., and Lee, S. J. (2013). Predicting 10-year mortality for older adults. *JAMA*, 309(9):874–876.

De Nardi, M., Fella, G., and Paz-Pardo, G. (2016). The implications of richer earnings dynamics for consumption and wealth. NBER Working Paper 21917, National Bureau of Economic Research.

De Nardi, M., French, E., and Jones, J. B. (2006). Differential mortality, uncertain medical expenses, and the saving of elderly singles. NBER Working Paper 12554, National Bureau of Economic Research.

Feldstein, M. and Samwick, A. A. (1992). Social security rules and marginal tax rates. *National Tax Journal*, 45(1):1–22.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *The Quarterly Journal of Economics*, 122(1):1–37.

French, E. (2005). The effects of health, wealth, and wages on labour supply and retirement behaviour. *The Review of Economic Studies*, 72(2):395–427.

French, E. and Jones, J. B. (2011). The effects of health insurance and self-insurance on retirement behavior. *Econometrica*, 79(3):693–732.

French, E., Lindner, A. S., O'Dea, C., and Zawisza, T. A. (2022). Labor supply and the pension-contribution link. NBER Working Paper 30184, National Bureau of Economic Research.

Gelber, A. M., Jones, D., Sacks, D. W., and Song, J. (2022). The employment effects of the social security earnings test. *Journal of Human Resources*, 57(2):341–371.

Goss, S. C. (2019). Testimony by stephen c. goss, chief actuary, social security administration, before the house committee on ways and means. SSA testimony. July 25, 2019.

Gustman, A. L. and Steinmeier, T. L. (2005). The social security early entitlement age in a structural model of retirement and wealth. *Journal of Public Economics*, 89(2–3):441–463.

Hurd, M. D. and McGarry, K. (1995). Evaluation of the subjective probabilities of survival in the health and retirement study. *Journal of Human Resources*, 30:S268–S292.

Hurd, M. D. and McGarry, K. (2002). The predictive validity of subjective probabilities of survival. *The Economic Journal*, 112(482):966–985.

Hurd, M. D., Smith, J. P., and Zissimopoulos, J. M. (2004). The effects of subjective survival on retirement and social security claiming. *Journal of Applied Econometrics*, 19(6):761–775.

Laun, T. and Wallenius, J. (2016). Social insurance and retirement: A cross-country perspective. *Review of Economic Dynamics*, 22:72–92.

Lee, S. J., Lindquist, K., Segal, M. R., and Covinsky, K. E. (2006). Development and validation of a prognostic index for 4-year mortality in older adults. *JAMA*, 295(7):801–808.

Liebman, J. B. and Luttmer, E. F. P. (2015). Would people behave differently if they better understood social security? evidence from a field experiment. *American Economic Journal: Economic Policy*, 7(1):275–299.

Liebman, J. B., Luttmer, E. F. P., and Seif, D. G. (2009). Labor supply responses to marginal social security benefits: Evidence from discontinuities. *Journal of Public Economics*, 93(11–12):1208–1223.

Ohanian, L., Raffo, A., and Rogerson, R. (2008). Long-term changes in labor supply and taxes: Evidence from OECD countries, 1956–2004. *Journal of Monetary Economics*, 55(8):1353–1362.

Prescott, E. C. (2004). Why do Americans work so much more than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review*, 28(1):2–13.

Rogerson, R. and Wallenius, J. (2009). Micro and macro elasticities in a life cycle model with taxes. *Journal of Economic Theory*, 144(6):2277–2292.

Rust, J. and Phelan, C. (1997). How social security and medicare affect retirement behavior in a world of incomplete markets. *Econometrica*, 65(4):781–831.

Social Security Administration (2012). Annual statistical supplement, 2012: Appendix d. computing a retired-worker benefit. SSA Annual Statistical Supplement.

Social Security Administration (2025a). Benefits planner: Retirement | receiving benefits while working. SSA Benefits Planner.

Social Security Administration (2025b). Program explainer: Retirement earnings test. SSA program explainer.

Social Security Administration, Office of the Chief Actuary (2025). Summary of provisions that would change the social security program: Provisions affecting payroll taxes. SSA OCACT policy options page. Estimates based on the 2025 Trustees Report.

Steingrimsson, J. A., Gatsonis, C., Li, B., and Dahabreh, I. J. (2023). Transporting a prediction model for use in a new target population. *American Journal of Epidemiology*, 192(2):296–304.

Stock, J. H. and Wise, D. A. (1990). Pensions, the option value of work, and retirement. *Econometrica*, 58(5):1151–1180.

Yaari, M. E. (1965). Uncertain lifetime, life insurance, and the theory of the consumer. *The Review of Economic Studies*, 32(2):137–150.