The Intergenerational Welfare Cost of Social Security Insolvency

HongSeok Kim†

Job Market Paper

Latest version available here.

September 2026

This paper studies the intergenerational welfare effects of a 22-percent reduction in Social Security payments, following assumed OASI reserve depletion at the end of 2033. Everyone will face the same cut, but at different ages. I use a life-cycle model in which saving, work, investment in human capital, and the claiming date are endogenous, and compare cohorts that meet the cut in retirement, in mid-career, and at the start of working life. The cohort that loses the most dollars is not the cohort that loses the most welfare. What a cohort loses depends on how much time it has to self-insure and on what self-insurance costs at its age. Subjective onset hazards are mapped from a financial-risk index and disciplined by HRS expectations. The welfare ranking depends on the expected duration of reduced payments. It governs who bears the largest welfare cost: brief benefit shortfalls weigh most heavily on cohorts near retirement, while persistent shortfalls shift the burden toward younger generations.

Keywords: Social Security, reserve depletion, retirement, intergenerational incidence

JEL Classification: H55, J26, E21, D15.

I am deeply grateful to my advisor, Juan Carlos Córdoba, for invaluable advice, guidance, and encouragement. I am grateful to my dissertation committee, Rajesh Singh, Quinn Weninger, Sergio Lence, and Jian Li, for their continued support. I also thank Kathleen McGarry, John Winters, Agnitra Roy Choudhury, and Xiaobo Yu for thoughtful comments and feedback, as well as participants at the 2026 North American Summer Meeting of the Econometric Society, the 2026 WEAI Annual Conference, and the Iowa State Micro/Macro Reading Group.
† Department of Economics, Iowa State University, Email: hongseok@iastate.edu

1. Introduction

Social Security trust-fund depletion would impose the same proportional benefit reduction on households at different stages of life. I consider depletion at the end of 2033, following the timing projection in U.S. Social Security Administration (2025), with payments falling to 78 percent of scheduled amounts in 2034. A retired household has limited opportunities to replace lost income through work, but has already received part of its lifetime benefits. A young worker has more opportunities to adjust, but more future benefits at risk. Neither observation alone determines who loses more. That comparison requires specifying how long reduced payments last and valuing the choices households make in response. This paper examines how these two considerations determine welfare losses across birth cohorts.

Accounting is a common approach to evaluate who loses the most when the Social Security insolvency happens. It asks how many present-value discounted dollars of scheduled benefits will each cohort fail to receive. This is a useful benchmark, but it is insufficient. No one can buy an insurance against a benefit cut, but individuals anticipating insolvency can save more, work longer, invest in human-capital, or claim later. A worker of age 58 has few years left before retirement, but the opportunity cost of self-insurance is high. On the contrary, a worker of age 28 has a whole career to self-insure but may lose a larger share of scheduled benefits. Hence, understanding who bears the burden of Social Security insolvency cannot be answered from an accounting exercise which implicitly assume a fixed-behavior.

This is a quantitative question, and to answer it, I use a life-cycle model in which saving, labor supply, human-capital investment and timing of claiming benefits as endogenous choices. I simulate different birth cohorts confronting the same insolvency event but at different ages in a life cycle model. Two forces are at work. The earlier the benefit cut comes in a cohort's life, the larger the share of its retirement it spends on reduced benefits. An earlier event also leaves more working years in which to adjust earnings, saving, and claiming. Households in the anticipated environment choose under their subjective hazard profiles from age 18, so the welfare consequences depend on the entire life-cycle allocation. Which force dominates is not obvious ex ante. I find that the cohort that loses the most dollars as in accounting exercise is not necessarily the cohort who loses the most welfare.

Model. I extend the life-cycle model of Fan, Seshadri, and Taber (2024), in which households choose consumption, saving, binary labor supply, Ben–Porath human-capital investment, and when to claim Social Security. The extension introduces three aggregate payment states: full payment before insolvency, reduced payment after reserve depletion, and restored full payment. Households take this payment process as given and adjust their choices in response. An important life-cycle friction is a risky return to human-capital investment. Employment and investment determine current earnings and future earning capacity, while covered earnings and claiming age determine scheduled retirement benefits. Payment risk changes the return to covered earnings and delayed claiming, while the prospect of lower retirement income also creates incentives to accumulate resources. The net response is determined within the model. All decisions are determined jointly throughout the life cycle. Therefore, the model captures how households change both their lifetime resources and their retirement income in response to Social Security insolvency under subjective beliefs.

I simulate five birth cohorts that enter the model at age 18 in different calendar years. Each cohort encounters the same aggregate Social Security insolvency at a different age. In the anticipated environment, a common calendar-time series of index-mapped subjective hazards, disciplined by HRS expectations, generates a different age-specific risk profile for each cohort. These hazards enter the dynamic programming problem from model entry, allowing future payment risk to affect choices before insolvency arrives. I hold the realized reserve depletion year at the end of 2033 and the subsequent paid-benefit fraction at 0.78, while changing the expected time until the Congress makes a reform and restores full payment.

To distinguish the consequences of the shortfall from those of preparing for it, I compare three environments. In the full-payment environment as a benchmark, households perceive no insolvency or onset risk and receive full scheduled benefits. In the surprise-shortfall environment, they follow the benchmark's decision rules through 2033 and re-adjust when reduced payments begin in 2034. In the anticipated-shortfall environment, they incorporate subjective beliefs about Social Security insolvency and restoration into their choices from the beginning of the life cycle. The surprise and anticipated environments face identical realized paths of paid-benefit fraction, 0.78, although actual benefits can differ because their covered earnings and claiming age can differ.

These comparisons separate the surprise-shortfall effect, which includes adjustment after payments fall, from the anticipation effect, which captures the consequences of preparing beforehand. Anticipatory saving, employment, and investment can support consumption during the shortfall, but requires reallocation of consumption, leisure, or current earnings throughout the life cycle. Whether self-insurance or preparation improves the welfare depends on these trade-offs, the resources it produces, and the household's age in the life cycle. I evaluate each environment using lifetime utility and express each comparison in consumption-equivalent units.

Quantitative results. The main analysis asks how the expected duration of reduced payments changes which generation bears the largest welfare cost. The policy lever that I employ to measure welfare is a duration risk until a full payment resumes after insolvency. Each finite expected duration represents a distribution of possible shortfall lengths, rather than a predetermined restoration date. The permanent-shortfall case is the limiting case in which full payments never resume.

Expected duration changes both the size and the distribution of welfare costs. With an expected shortfall of five years, the 1965 cohort has the largest model-implied welfare cost, approximately 0.34 percent of lifetime consumption. This cohort is already eligible for retirement benefits when payments fall, whereas younger cohorts are more likely to reach retirement after full payments have resumed. As expected duration increases, reduced payments extend further into younger cohort's retirement years. At an expected duration of twenty years, the 1975 cohort has the largest welfare-cost point estimate, at 0.649 percent. Under permanent reduced payments, the welfare costs are approximately 1.26 percent of lifetime consumption for each 1975, 1982, and 2005 cohorts, compared with 0.89 percent for the 1965 cohort and 0.36 percent for the 1957 cohort.

Two mechanisms explain this shift. First, longer shortfalls increase benefit exposure, especially for cohorts whose retirement lies farther in the future. Second, households adjust to perceived payment risk through saving, employment, human-capital investment, and claiming. These choices can support resources after insolvency but entails costs in the life cycle. The anticipation comparison measures their net welfare consequences relative to an unexpected shortfall. Its sign varies across cohorts and duration scenarios, so total welfare costs need not increase monotonically with expected duration.

The comparison with unpaid benefits in dollar shows why an accounting measure alone cannot establish welfare incidence. The accounting calculation does not capture the value of transferring resources throughout the life cycle. The model-implied welfare measure evaluates consumption, leisure, and bequests under the choices households make in response to payment risk. The difference between accounting and welfare losses therefore compares valuation concepts; it does not identify the fraction of lost income recovered through self-insurance.

The results describe distributional incidence conditional on the specified payment processes. Full-payment restoration is represented by a transition probability rather than a particular legislative reform. The taxes and spending changes required to finance restoration are not specified which is out of scope of this paper. The welfare analysis in this paper identifies how the persistence of a payment shortfall after Social Security insolvency changes intergenerational incidence.

Empirics. The preceding welfare comparisons depend critically on how households perceive future Social Security payment-risk or insolvency. I discipline these subjective beliefs using a calendar-time financial-risk index and Health and Retirement Study (HRS) survey expectations. The empirical analysis also examines the relationship between perceived benefit risk and intended work, providing descriptive motivation for the model's anticipatory choices.

The empirical analysis motivates anticipatory household behavior and constructs the subjective onset hazards used in the model. I first examine how perceived Social Security benefit risk relates to work expectations. I then combine public information about the program finances with HRS survey beliefs to obtain a common calendar-time hazard series, which enters each cohort's dynamic programming problem at corresponding ages.

In the 2006 Health and Retirement Study (HRS), respondents ages 50–61 in the sample assign an average probability of 60.8 percent that policy changes within ten years will reduce their own future Social Security benefits. A one standard deviation higher perceived risk is associated with a 2.3 percentage-point higher stated probability of working full time after age 62. In other words, individuals who expect higher probability of benefit cut plans to work longer in the late-age. Stated work plans also predict subsequent employment. Among respondents observed again in 2010, a 10 percentage point higher stated work probability is associated with a 3.7 percentage point higher probability of full time employment, conditional on prior employment and household characteristics. I use these descriptive evidence to motivate allowing households to self-insure for future payment risk.

The model requires beliefs about payment risk at each calendar year by different generations. Survey observations provide these beliefs only at a few dates, leaving much of the relevant time path unobserved. Assigning a single survey probability to every year would be incorrect as information set about Social Security finances changes over time. I therefore construct a calendar-time measure of public financial conditions and use the survey observations to discipline its mapping into household-level perceived insolvency risk. This approach allows me to generate annual belief profile while allowing different cohorts to encounter changes in perceived risk at different age.

To construct subjective onset hazards, I build a monthly financial-risk index from Social Security Trustees projections, Congressional Budget Office reports, labor market conditions, revisions to official forecasts, and news-based economic uncertainty (Baker, Bloom, and Davis 2016). Once I construct the headline index, I map its annual index level into a conditional annual hazard. The mapping parameters are disciplined by HRS survey average for the perceived probability that Social Security becomes less generous within ten years. Assuming a constant annual hazard over the horizon, the fitted annual hazard averages 12.2 percent during 2010–2019 and reaches 14.1 percent in the beginning of 2026. I use these estimated calendar-time hazards in the structural analysis.

Related Literature. This paper mainly contributes to literature on Social Security policy risk and the intergenerational effects of pension reform. Shoven and Slavov (2006) quantify political risk in Social Security, and Luttmer and Samwick (2018) measure the welfare cost of perceived uncertainty about future benefits. Caliendo, Gorry, and Slavov (2019) study household decisions and welfare under uncertainty about the timing and structure of reform. They show that optimal saving can substantially reduce the welfare cost of uncertainty about the timing and structure of Social Security reform. Kitao (2014) compares the distributional consequences of alternative financing reforms, while Kitao (2018) studies policy uncertainty and the intergenerational costs of delaying reform in Japan. She studies how households adjust saving and labor supply as reform uncertainty resolves, and quantifies the resulting welfare consequences across generations in general equilibrium. I ask a complementary question compared to the above. The first contribution is to quantify how the expected persistence of a payment shortfall due to insolvency changes welfare incidence across birth cohorts. None of these papers have human-capital investment which is a crucial component for self-insurance in life-cycle.

Second, this paper studies how adjustment opportunities shape the welfare consequences of anticipating a common payment shortfall. The structural framework of Fan, Seshadri, and Taber (2024) highlights the interaction between human-capital investment and labor supply in evaluating Social Security changes. I extend that framework with annual subjective onset hazards constructed from public financial information and disciplined by HRS expectations. The same calendar hazard enters each cohort's decision problem at a different age. Comparing anticipated and unexpected shortfalls under identical realized payment-regime histories measures whether anticipatory choices cushion or amplify welfare losses, and how that balance changes with life-cycle position and expected shortfall duration. Saving, employment, human-capital investment, and claiming are jointly endogenous, so the analysis values both the resources produced by preparation and the consumption, leisure, and current earnings it costs.

Roadmap. Section 2 presents the life-cycle model and its payment regimes. Section 3.1 documents perceived personal benefit risk and intended work. Section 3.2 constructs the Social Security financial-risk index and maps public information into a survey-based onset hazard. Section 4 defines the counterfactual environments and welfare comparisons. Section 5 traces welfare incidence along the expected-duration frontier, and Section 6 concludes.

2. Life-Cycle Model

I extend the life-cycle model of Fan, Seshadri, and Taber (2024) to study how anticipated Social Security payment shortfalls affect different birth cohorts. The novel feature of this paper is to introduce an aggregate payment regime into a household problem with saving, full-time employment, human-capital investment, and claiming choices. These choices determine both the resources available when payments fall and the consumption and leisure reallocated.

2.1. Model overview and timing

Time is annual. Households enter at age $\underline{a} = 18$ and live through terminal age $T = 80$. I use t to index decision periods and a_t to denote age. For birth cohort b, calendar year is $b+a_t$. A common calendar event occurs at a different age for each cohort. Households share a calendar-time series of subjective payment-shortfall hazards, but the corresponding risk enters their life-cycle problems at different ages.

The household chooses consumption and saving, whether to work full time, how much working time to devote to human-capital investment, and when to claim Social Security benefits. An interesting mechanism arises when we have endogenous wages in life cycle model. On the one hand, employment and human-capital investment affect earnings. On the other hand, covered earnings and claiming affect scheduled benefits. A payment shortfall generates a standard wealth effect and substitution effect before Social Security insolvency arrives as well as after payments fall. This is the key to understand the distributional incidence of Social Security insolvency.

Each cohort enters at age 18 knowing its full age-specific onset-hazard profile. These profiles map the HRS-disciplined financial-risk index into calendar years and then ages. Hazards are set to zero before the series begins in 1995 and held at their partial-2026 level after its last observation. Knowledge of the entire profile from model entry is a maintained information assumption, specified in Section 4.

Within each year, individuals make their decisions given their individual state vector, the aggregate payment regime of Social Security, and idiosyncratic leisure shock. The amount each individual receive from scheduled benefits is determined by the aggregate payment regime. The aggregate state transition happens at the end of the year and changes the benefits for the following years. As a result, households have precautionary motives not only for saving but also labor supply and human-capital investment.

The model is partial equilibrium. The rental price of human capital, interest rate, tax schedules, and claiming rules are fixed across experiments. Wages remain endogenous through human-capital investment. Household parameters and the inherited program rules are common across the simulated cohorts, so birth year changes the age at which payment risk is encountered without introducing a separate set of cohort-specific preferences or technologies.

2.2. Social Security payment risk

I first explain about aggregate payment regime because this is the novel part of this paper. The aggregate payment regime is $z_t \in \{F_0, L, F_1\}$. The economy begins in the pre-depletion full payment state F_0 where households receive their OASI schedueld payments. Once the Social Security trust-fund runs out, the economy enter the reduced-payment state L in which households receive only a fraction of scehdueld payments. After then, the Congress may enact a policy reform so as to restore full scheduled payments. This state is denoted as F_1. Once the reform is implemented, restoration is permanent so F_1 is an absorbing state.

Let ϕ be the fraction of scheduled payments paid in state L. The payment factor is

$$\chi(z_t) = \begin{cases} 1, & z_t \in \{F_0, F_1\}, \\ \phi, & z_t = L, \end{cases} \qquad 0 < \phi < 1. \tag{1}$$

Let b_t^{pre} denote the household's scheduled OASI payment after the Retirement Earnings Test. Actual benefits are

$$ssb_t = \chi(z_t)\, b_t^{pre}. \tag{2}$$

The earnings record and claiming rules determine b_t^{pre}, which will be discussed in detail below. The factor χ reduces actual payments while leaving scheduled entitlements and their statutory adjustments unchanged. It applies equally to households that claimed before and after depletion, so early claiming cannot gurantee the full payments. The fraction is applied after the earnings test and its future-entitlement credit, and before benefit taxation. Payroll tax rules remain unchanged.

Cohort b at age a_t considers an end-of-year aggregate payment regime transition from F_0 to L with perceived probability $\lambda_b(a_t)$, conditional on depletion not having occurred. Its subjective transition law satisfies

$$\mathcal{Q}_b(L \mid F_0, a_t) = \lambda_b(a_t), \qquad \mathcal{Q}_b(F_0 \mid F_0, a_t) = 1 - \lambda_b(a_t) \tag{3}$$

which differs by cohort b. The hazard vector is obtained by mapping the annual calendar hazards from Section 3.2 into the ages of cohort b. It enters the household's optimization problem at every age. Future onset risk can change current decisions even at ages when the current-year hazard is zero.

Once payments are reduced, the annual probability of restoration of payment shortfall is q_d:

$$\mathcal{Q}(F_1 \mid L) = q_d, \qquad \mathcal{Q}(L \mid L) = 1 - q_d. \tag{4}$$

A constant $q_d > 0$ generates a geometric reduced-payment duration with expected duration $d = 1/q_d$. For a limiting case, $q_d = 0$ gives a permanent shortfall. Restoration returns the paid-benefit fraction to one without repaying missed benefits.

The subjective transition law $\mathcal{Q}_b$ governs household choices. A separate objective law $\mathcal{Q}^*$ governs the aggregate payment histories used in simulation and welfare evaluation. Subjective onset beliefs can differ from the realized onset process. The experiments hold the restoration hazard common to beliefs and the objective process. Section 4 specifies the index-mapped calendar hazards, realized depletion date, paid-benefit fraction, and matched environments. Section 5.1 specifies the duration scenarios. The model takes these payment processes as given and does not model the financing of full-payment restoration which is out of scope of this paper.

2.3. Households and preferences

Households differ in their resources and in the utility cost of adjusting work and consumption. For the baseline model, I retain the demographic structure, permanent types, and preferences of Fan, Seshadri, and Taber (2024).

2.3.1. Demographics and permanent types

Family status $M_t \in \{0, 1, 2\}$ distinguishes single or divorced households, married households with a nonworking spouse, and married households with a working spouse. It evolves exogenously according to

$$\Pr(M_{t+1} = m' \mid M_t = m) = P_t^M(m, m'), \tag{5}$$

estimated from SIPP transition frequencies by Fan, Seshadri, and Taber (2024). Family status scales consumption utility (Gourinchas and Parker 2002; De Nardi et al. 2025), shifts the taste for leisure, and determines whether spousal income is received. Conditional on $M_t = 2$, spousal earnings are log-normal with age-specific mean and variance; otherwise they are zero. Spousal labor supply is stochastic, but exogenous.

A permanent household type is

$$\vartheta_i = (a_{0i}, \pi_i, H_{i0}),$$

where a_{0i} shifts the taste for leisure, π_i is the efficiency of human-capital production, and H_{i0} is initial human capital. The joint distribution of (a_{0i}, π_i) has nine mass points, and initial human capital depends on both components and a residual initial-condition shock. These types generate differences in earnings, accumulated wealth, and work attachment, allowing different responses to the same perceived onset risk. Health remains fixed, and SSDI and SSI receipt are inactive. The horizon ends at age 80 without an additional survival process.

2.3.2. Consumption, leisure, and claiming

Let $p_t \in \{0, 1\}$ indicate full-time work and $\ell_t = 1 - p_t$ leisure. Employment is an extensive-margin choice: households choose a fixed full-time schedule or nonwork, with no part-time or total-hours choice. A worker can allocate part of that schedule to training. Flow utility is

$$u_t(c_t, p_t, d_t; M_t, a_{0i}, \varepsilon_t) = \psi_t(M_t)\frac{c_t^{1-\eta}}{1-\eta} + \exp\{a_{0i} + \alpha^M(M_t) + \varepsilon_t\}\,\ell_t + \nu_t d_t, \tag{6}$$

The consumption term is CRRA with curvature η and a family-size shifter $\psi_t(M_t)$. The leisure term depends on permanent type, the family shifter $\alpha^M(M_t)$, and an independent and identically distributed shock $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ observed before the work decision. More saving can require lower current consumption, while additional employment entails a leisure cost. These margins determine the utility cost of preparing for reduced benefits.

The indicator d_t equals one when the household initiates benefits. Its age-dependent utility payoff is

$$\nu_t = \nu_{62}\,\mathbf{1}\{a_t = 62\} + [\nu_{a,65} + \nu_{b,65}(a_t - 65)]\,\mathbf{1}\{65 \le a_t \le 70\}, \tag{7}$$

This term captures determinants of claiming outside the financial incentives in the model and helps reproduce claiming peaks at ages 62 and 65 (Rust and Phelan 1997). Benefit initiation is irreversible, but nonwork is not: a household that stops working can return in a later year. Retirement therefore denotes nonwork late in life rather than an absorbing state.

2.4. Employment, human capital, and earnings

Human-capital investment links adjustment today to earning capacity later in life. Productive human capital is $H_t > 0$. Conditional on employment, the household chooses the share I_t of working time devoted to training:

$$\mathcal{I}(p_t) = \begin{cases} [0,1], & p_t = 1, \\ \{0\}, & p_t = 0. \end{cases} \tag{8}$$

With $\overline{N} = 2{,}000$ annual hours, training uses $N_t^{\text{training}} = \overline{N} I_t$ and production uses $N_t^{\text{paid}} = \overline{N}(1 - I_t)$ for a worker. The rental price of human capital is normalized to one, giving the hourly wage

$$w_t = H_t, \tag{9}$$

and annual earnings

$$y_t = \overline{N} H_t p_t (1 - I_t). \tag{10}$$

Only production time is compensated. The wage target corresponds to the rental value H_t, whereas annual earnings also reflect employment and the share of time spent training. Human capital evolves through the Ben–Porath technology

$$H_{t+1} = (1-\delta)H_t + p_t\,\xi_t\pi_i I_t^{\alpha_I} H_t^{\alpha_H}, \qquad \mathbb{E}[\xi_t] = 1, \tag{11}$$

Here δ is depreciation, π_i is type-specific investment efficiency, and ξ_t is a mean-one log-normal innovation. New human capital is produced only while working; nonwork implies deterministic depreciation at the same rate applied on the job. There is no additional transitory or persistent wage shock.

Investment sacrifices current earnings in exchange for future productivity. Its return depends on the remaining working horizon. Anticipated reduced benefits can change that horizon and therefore the incentive to invest, subsequent wages, and covered earnings. This channel allows households to adjust future earning capacity as well as the amount they save from current income.

2.5. Social Security entitlements and claiming

Scheduled benefits depend on covered earnings and claiming age. Additional employment can therefore increase current resources and future retirement income. I retain the historical FST rules across all cohorts, including the normal retirement age of 65. Birth-cohort labels determine exposure to payment risk rather than different statutory retirement ages. Appendix C.2 gives the earnings-record approximation, benefit formula, and numerical claiming adjustments.

Claiming is a one-time, irreversible decision available from age 62 and required by age 70. Let s_t indicate whether benefits were initiated before year t and d_t indicate initiation in the current year. Claiming and employment are separate choices, so a household can receive benefits while working. The record state R_t stores Average Indexed Monthly Earnings (AIME) before claiming and the annual scheduled entitlement $\bar{b}_t$ afterward. Covered earnings update AIME only before claiming. Earlier claiming provides income sooner but lowers the annual entitlement, while delayed claiming raises it under the inherited adjustment schedule.

A working claimant can have current benefits withheld under the Retirement Earnings Test. Withholding earns a credit that increases future scheduled entitlements, exchanging current income for later payments. Its effect on work incentives depends on how the household values that exchange (Friedberg 2000; Gruber and Orszag 2003). Perceived payment-shortfall risk changes the value of the future credit without changing its statutory formula. This mechanism links onset beliefs to employment and claiming decisions before depletion.

Let RET_t denote benefits withheld under the test. The scheduled current payment is $b_t^{pre} = \max\{0, \bar{b}_t - \mathrm{RET}_t\}$; actual benefits then equal $\chi(z_t) b_t^{pre}$. The payment fraction applies to benefits already claimed as well as new claims, so claiming early cannot protect a household from the shortfall. Claiming responses reflect liquidity, consumption smoothing, and the expected return to delaying receipt. The experiment covers OASI retired-worker payments; disability benefits remain inactive.

2.6. Budget constraint, taxes, and bequests

Saving transfers resources across ages, while taxes and means-tested transfers affect how much of a gross benefit shortfall reaches consumption. I retain the FST federal income and payroll tax rules, reported in Appendix C.1. Let y_t^s denote spousal earnings and define other gross income as $Y_t^o = \max\{rA_t, 0\} + y_t + y_t^s$.

Social Security taxation depends on the benefit actually received, ssb_t. Lower payments can reduce the taxable portion ssb_t^{tax}, partially offsetting the gross payment loss. Appendix C.3 gives the thresholds and taxable-benefit formulas. Writing $\mathcal{T}$ for the inherited after-tax-income function, disposable income is

$$\Upsilon_t = \mathcal{T}(Y_t^o + ssb_t^{tax}) + ssb_t - ssb_t^{tax}. \tag{12}$$

The household budget constraint is

$$A_{t+1} = A_t + \Upsilon_t - c_t + tr_t, \qquad A_{t+1} \geq \underline{A}_{t+1}, \tag{13}$$

The transfer tr_t provides the consumption floor $\underline{c}$ following Hubbard, Skinner, and Zeldes (1995):

$$tr_t = \max\left\{0, \underline{c} - \left[A_t + \Upsilon_t - \underline{A}_{t+1}\right]\right\}. \tag{14}$$

The age-dependent borrowing bound $\underline{A}_{t+1}$ is the natural debt limit. The floor represents means-tested insurance outside Social Security and limits the consumption consequences of shortfalls for households with few resources. At the terminal age, remaining assets yield warm-glow bequest utility following De Nardi (2004):

$$B(A_{T+1}) = \frac{b_1}{1-\eta}(b_2 + A_{T+1})^{1-\eta}, \tag{15}$$

The parameter b_1 governs the strength of the motive and b_2 makes bequests a luxury. Together, the borrowing limit, transfer floor, and bequest motive shape households' incentives and capacity to self-insure through assets.

2.7. Recursive problem and solution

The payment regime changes current resources and the value of future choices. The recursive formulation brings these two channels together. Permanent type indexes separate household problems and is suppressed below. Before the leisure shock is realized, the individual state is

$$\boldsymbol{X}_t = (A_t, H_t, R_t, s_t, M_t, a_t), \tag{16}$$

The state collects assets, human capital, the Social Security record, claiming status, family status, and age. The aggregate regime indexes separate value functions V^{F_0}, V^L, and V^{F_1}. Cohort-specific onset beliefs index separate pre-depletion solutions, so I suppress the cohort index and write $\lambda(a_t)$ below. The individual state vector does not need an additional coordinate for birth cohort or the common payment regime.

2.7.1. Decisions within a period

For $p \in \{0, 1\}$, let $W_t^p(\boldsymbol{X}_t)$ be the optimized value of work branch p, excluding the current leisure payoff. Write $W_t^1 = W_t^W$ and $W_t^0 = W_t^N$. Conditional on the current regime, the branch problem is

$$W_t^p(\boldsymbol{X}_t) = \max_{c_t, I_t, d_t} \left\{ \psi_t(M_t)\frac{c_t^{1-\eta}}{1-\eta} + \nu_t d_t + \beta\,\mathbb{E}\big[\overline{V}_{t+1}(\boldsymbol{X}_{t+1}) \,\big|\, \boldsymbol{X}_t,\, p_t = p\big] \right\}, \tag{17}$$

subject to the training set (8), human-capital law (11), claiming and benefit rules, and budget constraint (13). Consumption and investment are continuous choices; claiming is discrete, and $I_t = 0$ when $p = 0$. Current benefits and disposable income are evaluated under the current payment regime. The expectation integrates the family, spousal-earnings, and human-capital shocks as detailed in Appendix D.

The leisure shock determines which optimized branch is chosen. When $W_t^W > W_t^N$, the household works if its shock is below

$$\mu_t = a_{0i} + \alpha^M(M_t), \qquad \varepsilon_t^* = \log(W_t^W - W_t^N) - \mu_t. \tag{18}$$

The employment probability conditional on the remaining states $\widetilde{X}_t$ is

$$\Pr(p_t = 1 \mid \widetilde{X}_t) = \Phi\left(\frac{\varepsilon_t^*}{\sigma_\varepsilon}\right). \tag{19}$$

Integrating the maximum of the two branches gives

$$\begin{aligned}\mathcal{E}_t(W_t^W, W_t^N) = \Phi(q_t)W_t^W + [1 - \Phi(q_t)]W_t^N \\ + \exp\left(\mu_t + \frac{\sigma_\varepsilon^2}{2}\right)\Phi(\sigma_\varepsilon - q_t), \qquad q_t = \frac{\varepsilon_t^*}{\sigma_\varepsilon},\end{aligned} \tag{20}$$

where Φ is the standard normal cumulative distribution function. If $W_t^W \leq W_t^N$, nonwork is optimal for every shock realization, and expected utility equals W_t^N plus the unconditional mean leisure payoff. Appendix D.1 derives the closed form and its limiting case. The start-of-period value is

$$V_t(\boldsymbol{X}_t) = \mathcal{E}_t\big(W_t^W(\boldsymbol{X}_t),\, W_t^N(\boldsymbol{X}_t)\big)\,. \tag{21}$$

2.7.2. Payment risk in continuation values

While the current regime determines current benefits, transition probabilities weight next period's regime-specific values. From F_1, full payment continues with certainty. From L, restoration occurs with probability q_d. From F_0, depletion occurs at the end of the current year with subjective probability $\lambda(a_t)$, reducing payments in the following year. Hence,

$$\overline{V}_{t+1}(\boldsymbol{X}_{t+1}) = \begin{cases} V_{t+1}^{F_1}(\boldsymbol{X}_{t+1}), & z_t = F_1, \\ [1 - q_d]V_{t+1}^{L}(\boldsymbol{X}_{t+1}) + q_d V_{t+1}^{F_1}(\boldsymbol{X}_{t+1}), & z_t = L, \\ [1 - \lambda(a_t)]V_{t+1}^{F_0}(\boldsymbol{X}_{t+1}) + \lambda(a_t)V_{t+1}^{L}(\boldsymbol{X}_{t+1}), & z_t = F_0. \end{cases} \tag{22}$$

This expression is where the cohort-specific hazard enters the dynamic programming problem. A higher weight on the reduced-payment continuation value can change the household's current consumption, work, investment, and claiming choices even while current payments remain full. Backward induction also carries future onset risk into decisions at earlier ages when the current onset hazard is zero.

I first solve the restored full-payment problem V^{F_1}, whose payment regime is absorbing. I then solve V^L using V^{F_1} and q_d. Finally, I solve V^{F_0} separately for each cohort's hazard vector using the reduced-payment continuation values. These steps are performed for each permanent type and duration specification. At each age and state, the solver optimizes the feasible work and claiming branches and integrates the shocks. The numerical grids and optimization procedures are reported in Appendix D.

The resulting decision rules are then simulated under the objective payment process $\mathcal{Q}^*$. This separates the beliefs governing preparation from the payment histories on which outcomes and welfare are evaluated.

2.8. Parameterization and benchmark fit

I retain the parameter estimates of Fan, Seshadri, and Taber (2024). Household preferences, human-capital technology, heterogeneity, and inherited tax and benefit rules are held fixed when introducing payment risk. The additional primitives are the subjective onset-hazard profiles, the objective onset process, the paid-benefit fraction, and the restoration hazards. Their quantitative specification is given in Section 4 and Section 5.1. The empirical hazards in Section 3.2 are the subjective onset beliefs used throughout the structural counterfactuals.

Setting perceived onset risk to zero and maintaining $\chi = 1$ returns the full-payment benchmark. Figure 1 compares its employment, wage, consumption, and claiming profiles with the targets used by Fan, Seshadri, and Taber (2024). The fit varies across moments, providing a benchmark check for the inherited model. Household parameters are not re-estimated using the payment-shortfall counterfactuals.

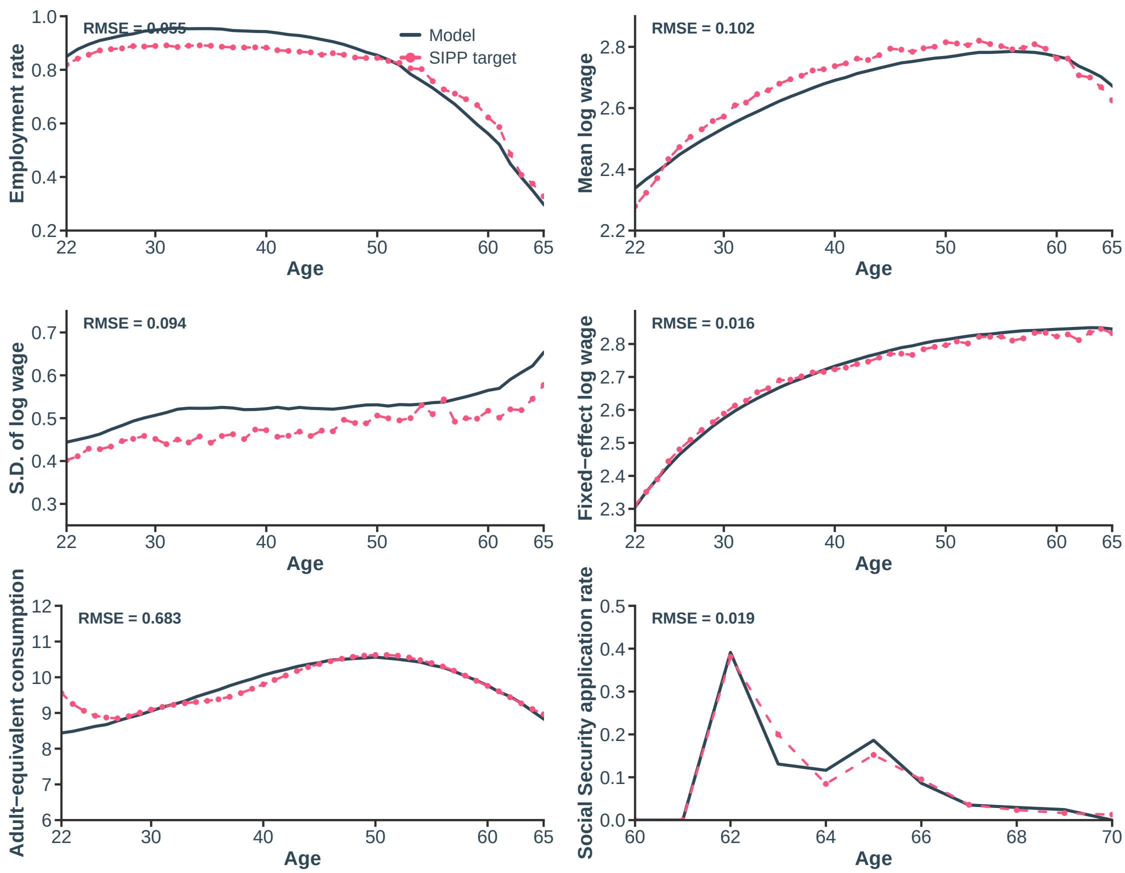

FIGURE 1. Replication of Fan, Seshadri, and Taber (2024) benchmark

Note. The figure compares the Fan, Seshadri, and Taber (2024) benchmark with the corresponding empirical target profiles. Household parameters are inherited from their work and are not re-estimated using the payment-shortfall experiments.

3. Perceived Social Security Benefit Risk

This section presents two pieces of motivating evidence linking perceived Social Security benefit risk to stated economic choices. Direct evidence on OASI reserve depletion is unavailable because depletion has not occurred, and no market insurance protects households against a policy-induced decline in future payments.[1] I therefore use survey questions about the possibility that Social Security will become less generous. These questions measure broad policy risk rather than the structural model's specific reserve-depletion event (Diamond 1997; Shoven and Slavov 2006; Luttmer and Samwick 2018; Caliendo, Gorry, and Slavov 2019).

[1]A life annuity insures idiosyncratic longevity risk; it does not insure the level of Social Security payments.

3.1. Perceived Benefit Risk and Intended Work

Do people approaching retirement report a greater intention to work when they perceive more risk to their own future Social Security payments? This relationship provides descriptive evidence for the model's premise that benefit beliefs are associated with choices made before payments change.

I use the 2006 Health and Retirement Study (HRS) Internet Survey. I restrict the sample to respondents ages 50–61 who answered the relevant questions, are not receiving Social Security, and expect to receive benefits in the future. These respondents have not yet reached the earliest retirement-benefit claiming age of 62. The age-62 regression sample contains 3,359 individuals. Appendix A reports summary statistics. The reported sample means and regressions are unweighted and therefore describe this Internet Survey sample rather than a nationally representative population.

The main explanatory variable is *perceived personal benefit risk*, denoted by R_i^P.[2] It is the respondent's reported probability that Social Security will change during the next ten years in a way that lowers the respondent's own payments relative to current law. Its sample mean is 60.8 percent, with a standard deviation of 28.2 percentage points. I do not call this measure insolvency risk or a subjective depletion hazard because the survey permits many policy mechanisms.

The main outcome is the individual's reported probability of working full time after age 62, denoted by $W_{i,62}$. It measures intended work at an older age, not realized employment. Its sample mean is 47.2 percent and its standard deviation is 37.8 percentage points. The dispersion partly reflects round-number heaping at 0, 50, and 100, a common feature of subjective-probability questions in household surveys (Manski 2004; Hurd 2009).

I also construct *perceived general program risk*, R_i^G, from the reported probability that Social Security will become less generous in general. This separates the risk a person assigns to personal payments from a belief about program-wide generosity. Some respondents may expect a future reform but also expect Congress to protect people close to retirement. Holding R_i^G

[2]The general-risk question (HRS 2006, question P110) reads: "Thinking of the Social Security program in general and not just your own Social Security benefits: On a scale from 0 to 100, what is the percent chance that Congress will change Social Security sometime in the next 10 years, so that it becomes less generous than now?" The personal-risk question follows it: "We just asked you about changes to Social Security in general. Now we would like to know whether you think these Social Security changes might affect your own benefits." The question does not identify a specific mechanism, such as a statutory entitlement change, a higher retirement age, benefit taxation, or reserve depletion.

fixed, the coefficient below compares people with similar views about the program but different perceived risks to their own payments.

I estimate the regression equation

$$W_{i,62} = \alpha + \beta R_i^P + \theta R_i^G + X_i'\gamma + \varepsilon_i. \tag{23}$$

The controls X_i include age fixed effects, gender, education, race and ethnicity, marital status, self-reported health, earnings, and nonhousing wealth. Standard errors are clustered by household, allowing responses within a household to be correlated. Both risk measures are expressed in units of 10 percentage points. The coefficient β therefore measures the conditional difference in the reported work probability associated with a 10-percentage-point difference in perceived personal benefit risk.

Figure 2 shows the result. A 10 percentage-point difference in perceived personal benefit risk is associated with a 0.83 percentage-point difference in the reported probability of working after age 62. A difference of one standard deviation in perceived risk—28.2 percentage points—therefore corresponds to about a 2.3 percentage-point difference in the work probability. This difference is about 5 percent of the sample's average work probability. By comparison, the coefficient on general program risk is -0.32 (standard error 0.32), and its confidence interval includes zero.

The unadjusted survey means point in the same direction. Among people who report a 0–24 percent chance that policy changes will lower their personal payments, the average probability of working after age 62 is 44.9 percent. Among those who report a 75–100 percent chance, the average is 48.9 percent. The average rises across all four risk groups. Because many respondents give round-number answers such as 0, 50, or 100, I treat this panel only as a descriptive group comparison. The regression panels report the conditional relationship after including the controls.

The estimated relationship remains similar in the alternative specifications reported in Table A.2. It changes little when I add the respondent's expected chance of survival and perceived chance of a major economic depression, use controls measured before the belief question, or include pension coverage on the current job. Pension coverage is undefined for respondents who are not working, so I code it as no pension, has a pension, or not currently working. With this control, the coefficient on perceived personal benefit risk is 0.84 (standard error 0.27), close to the baseline 0.83. When the outcome is the reported probability of working

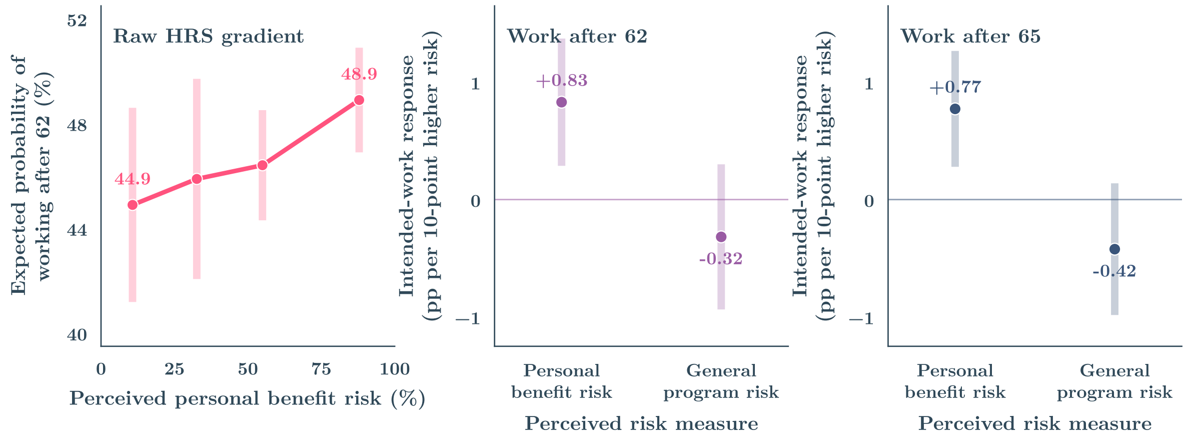

FIGURE 2. Perceived Social Security risk and intended work at older ages

Note. The sample consists of people ages 50–61 who answered the 2006 HRS survey for themselves, are not receiving Social Security, expect to receive benefits in the future, and answered the questions about personal and general risk. The left panel shows the unadjusted average for each perceived-personal-benefit-risk group. The middle and right panels show coefficients from equation (23), estimated separately for the reported probability of working after ages 62 and 65. The bars show 95 percent confidence intervals. The age-62 regression includes 3,359 people, and the age-65 regression includes 3,362. Standard errors are clustered by household.

after age 65, the coefficient is 0.77 (standard error 0.25), or 0.74 (standard error 0.25) with the pension control.

I next examine whether stated work plans contain information about subsequent employment. I follow 1,254 respondents who were ages 58–61 in 2006 to the 2010 HRS, by which time all had reached age 62. The final linear-probability-model sample contains 1,119 respondents with observed full-time employment and nonmissing controls.

A 10-percentage-point higher stated probability of working after 62 in 2006 is associated with a 3.7-percentage-point higher probability of full-time work in 2010 (standard error 0.4), conditional on 2006 full-time status and the same household characteristics. Stated plans therefore have predictive content for later employment. By contrast, the coefficients on perceived personal benefit risk in the 2008 and 2010 full-time employment regressions have confidence intervals that include zero. Table A.3 reports these regressions and the analogous 2008 exercise.

The evidence supports a limited conclusion. Respondents who perceive their own future benefits as less secure report a greater probability of working at older ages, and stated work plans predict later full-time employment. The regressions do not establish that benefit beliefs cause either planned or realized work. Stable employment preferences, financial knowledge, political beliefs, and private information about health or wealth may affect both beliefs and work plans. The survey question also combines OASI reserve-depletion risk with other policy changes that could reduce benefits. I therefore treat β as a descriptive association. It is not a model calibration target and does not enter the welfare calculations.

3.2. Mapping Public Information into Subjective Onset Hazards

The cross-sectional evidence in Section 3.1 does not show how public information about Social Security finances has changed over time. I therefore construct an *OASI financial-risk index*. The index is a summary of public information, not a probability. I then use HRS survey answers to map the index into an empirical series of subjective onset hazards. The model requires an annual risk profile for every cohort, while survey expectations are observed at only a few dates. The index supplies variation between those observations, and the survey mapping disciplines its probability scale. The resulting annual hazards supply the subjective onset beliefs in the structural counterfactuals in Section 4.

Households observe public information about Social Security's finances and the economy that supports those finances. I combine that information into four blocks. The OASI-finance block, weighted 0.30, measures the size and proximity of the financing problem. The macro-stress block, weighted 0.20, measures labor-market conditions because covered earnings generate payroll-tax revenue. The official-uncertainty block, weighted 0.10, measures revisions between successive Trustees Reports. The news-based Economic Policy Uncertainty index of Baker, Bloom, and Davis (2016), weighted 0.40, measures broader uncertainty about government economic policy. The weights are fixed and sum to one.

I build the index in three steps. First, I orient every input so that a larger number indicates greater OASI financial risk. Second, within each block I use a one-sided Kalman filter to summarize the information released by each month. "One-sided" means that a historical value uses no later information. Third, I combine the four filtered blocks using the fixed weights above. The index tracks OASI rather than combined OASDI because the structural experiment covers OASI retired-worker payments, not Disability Insurance. Appendix B lists every variable, transformation, release date, weight, and filtering rule.

Figure 3 shows the sources of movement in the headline index. The fiscal component combines OASI finances and labor-market conditions; the uncertainty component combines revisions in official forecasts with broader economic-policy uncertainty. Each component is centered at 100 and rescaled so that their weighted sum equals the headline at every date.

An index score is not a probability: a score of 120, for example, does not mean that a

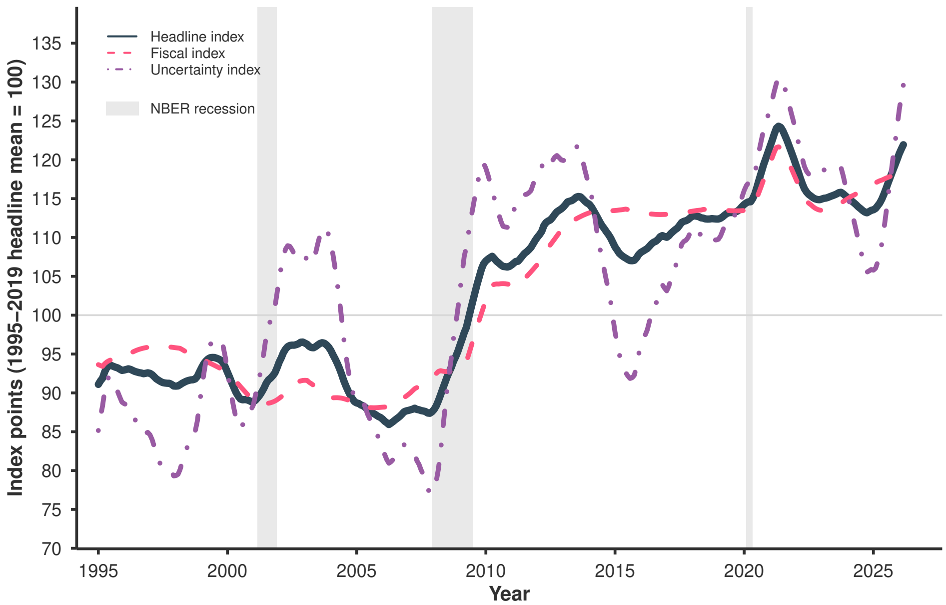

FIGURE 3. OASI financial-risk index and component indexes

Note. The fiscal index rescales the OASI-finance and macro-stress contributions, whose headline weights are 0.30 and 0.20. The uncertainty index rescales the official-uncertainty and economic-policy-uncertainty contributions, whose headline weights are 0.10 and 0.40. The displayed component indexes are normalized so that $I_t^H = 0.70 I_t^F + 0.30 I_t^U$; the 0.70 and 0.30 coefficients are display normalizations, not replacements for the four block weights. Before smoothing, all three indexes average 100 from January 1995 through December 2019, and a higher value indicates a more adverse OASI financial environment. The 2026 endpoint contains data from January through March. Shaded areas mark NBER recessions.

payment shortfall has a 120 percent chance of occurring. To obtain a probability scale, I use the 2006, 2007, and 2011 HRS Internet Surveys, which ask for the chance that Congress will make Social Security less generous within ten years. Let G_y be the calendar-year average of the unsmoothed OASI state. I estimate the subjective onset hazard

$$h_y = \Lambda(\alpha_0 + \alpha_1 G_y),$$

where Λ is the logistic function. The maintained mapping interprets h_y as the annual probability of entering a reduced-payment state in year y, conditional on not entering it earlier.

The survey asks about a ten-year horizon, whereas the model uses an annual hazard.[3] I assume the annual hazard is constant within the survey horizon, so its ten-year counterpart is $1 - (1 - h_y)^{10}$. I fit α_0 and α_1 to the three survey-wave means by weighted nonlinear least

[3]The survey asks about broad legislated changes in generosity; the model's reduced-payment state follows OASI reserve depletion. The mapping treats the survey answers as informative about the level of perceived benefit risk without equating the two mechanisms.

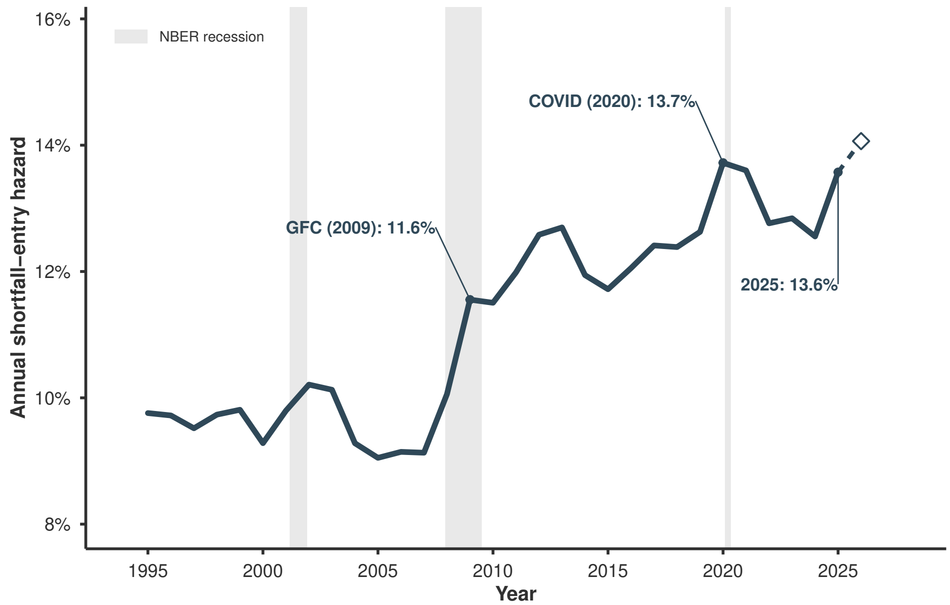

FIGURE 4. Empirical subjective onset hazard

Note. The figure maps the annual OASI financial-risk state into $h_y = \Lambda(\alpha_0 + \alpha_1 G_y)$. The parameters use HRS Internet Survey averages for the 10-year probability that Social Security will become less generous. The expression $1 - (1 - h_y)^{10}$ puts the annual hazard and survey answers on a common horizon. The 2026 endpoint uses data from January through March and is marked as a partial year. Shaded areas mark NBER recessions.

squares, using each wave's sample share as its weight. With three moments and two parameters, the fitted values approximate rather than exactly match all three means. Appendix B reports the objective, fit, and sensitivity of this mapping.

Figure 4 shows the fitted series. It is low and stable from 1995 through 2007, between 9.1 and 10.2 percent, and then rises during the financial crisis. It averages 12.2 percent from 2010 through 2019, reaches 13.7 percent in 2020, and equals 14.1 percent in the partial 2026 data. On the ten-year survey scale, the same path runs from about 64 percent in the late 1990s to 77 percent in 2020.

The two recessions move the index for different reasons. In 2009, labor-market stress and economic-policy uncertainty account for the sharp increase. Those components later mean-revert, but the level remains elevated because the OASI-finance component rises as successive Trustees Reports move projected reserve depletion closer and widen the actuarial deficit. COVID-19 produces a second uncertainty spike in 2020, while the underlying upward drift remains fiscal.

Two cautions govern interpretation. First, h_y is an annual onset hazard, not a cumulative belief. Second, the horizontal axis is calendar time, whereas households make decisions in age time. A household born in 1957 was 38 when the series begins and encountered its highest values only after age 60. A household born in 2005 turned 18 in 2023 and has encountered the elevated values throughout its working life. The same calendar-time path therefore maps into a different sequence of age-specific hazards for every birth cohort.

For model years outside the observed 1995–2026 series, I set the hazard to zero before 1995 and hold it at the partial-2026 estimate thereafter. Cohort b uses $\lambda_b(a) = h_{b+a}$. These extrapolation rules and knowledge of the full hazard profile from age 18 are maintained assumptions; they do not imply that households actually observed future index realizations.

4. Counterfactuals and Welfare Analysis

4.1. Environments

This section specifies counterfactual policy-reform scenarios designed to decompose the total incidence of a reform into an unanticipated benefit-cut effect and an anticipation effect. The first is the surprise-shortfall effect: a household that unexpectedly receives 78 percent of its scheduled OASI payment re-optimizes consumption, saving, work, human-capital investment, and claiming, if eligible, after reduced payments begin. The second is the anticipation effect: a household that assigns positive probability to payment-shortfall onset can adjust those choices before reserve depletion. Table 1 summarizes the environments.

TABLE 1. Counterfactual information and payment histories

Environment	Information used for choices	Objective payment history
Full-payment benchmark (O)	No perceived onset risk	Full scheduled payments
Unexpected shortfall (U)	Full-payment policy through the 2033 event	Reduced payments begin in 2034; restoration hazard q_d
Anticipated shortfall (A)	Cohort-specific onset beliefs and restoration hazard q_d	The same distribution as the surprise environment

The decomposition matters because anticipatory adjustment, or self-insurance, is not free. It creates a tension over the life cycle. On the one hand, it can reduce the marginal utility cost of reduced retirement benefits through consumption-smoothing. On the other hand, it is costly because households reallocate consumption and leisure across ages. Which force dominates is ambiguous ex ante, and a simple comparison across generations conflates these two mechanisms. I therefore simulate the same households with identical parameters, initial conditions, and payment histories. Within each cohort, the anticipated and surprise environments share the same onset date and restoration histories; they differ in the information used before onset. Across cohorts, the common calendar event occurs at different ages.

The decomposition then answers two separate questions. The surprise-shortfall effect shows how much a household loses when the cut arrives without warning. The anticipation effect shows how much of that loss a household can avoid by preparing in advance, after netting out the cost of preparing. Without separating the two, a small total loss could mean either that the cut was small or that households prepared well, and these have different implications for individuals.

4.2. Beliefs

Households' subjective onset beliefs come from the annual financial-risk index mapped to HRS expectations in Section 3.2. Let h_y denote the perceived probability of entering the reduced-payment state at the end of calendar year y, conditional on remaining in the pre-depletion state at the start of that year. The empirical input is

$$h_y = \Lambda(\widehat{\alpha}_0 + \widehat{\alpha}_1 G_y),$$

where G_y is the annual index state. The HRS questions concern broad reductions in Social Security generosity. Interpreting their mapped probabilities as onset hazards for the model's reduced-payment state is a maintained measurement assumption, not a claim that respondents directly forecast reserve depletion.

Calendar-year distribution. The annual hazards imply a distribution over possible onset dates. Let τ denote the end-of-year onset date. Conditional on no onset before year y_0, the probability assigned to year $y \geq y_0$ is

$$\pi_{y_0}(y) \equiv \Pr(\tau = y \mid \tau \geq y_0) = h_y \prod_{v=y_0}^{y-1} (1 - h_v).$$

The product is the probability of remaining in the pre-depletion state until year y. Thus the calendar distribution is derived from the index-mapped hazards; no separate Trustees-informed depletion-year distribution is used. Households reaching the same calendar year share the same conditional onset risk, even though that year occurs at different ages in their lives.

Cohort-specific hazard. A household born in year b reaches age a in calendar year $b + a$. The probability entering its dynamic program is therefore

$$\lambda_b(a) = h_{b+a}. \tag{24}$$

The input h_y is already conditional on no earlier onset, so no further survival adjustment is required. For example, the end-of-2033 hazard enters the 1957 cohort's problem at age 76 and the 2005 cohort's problem at age 28. Because onset occurs at year-end, reduced payments first enter the next period's budget.

The observed hazard series spans 1995–2026. I set hazards to zero before 1995 and hold them at the partial-2026 estimate, 0.140643, afterward. The anticipated environment assumes that each cohort knows its full hazard profile from age 18. The 1957 cohort begins making decisions in 1975 and the 2005 cohort in 2023. Future hazards can influence choices even in years when the current hazard is zero. This use of the full profile is a maintained information assumption, not a reconstruction of information households actually possessed at entry.

Subjective hazards govern choices before onset. The objective simulations instead impose depletion at the end of 2033 and reduced payments beginning in 2034. Households in the anticipated environment are not told that the realized onset date will be 2033.

Onset uncertainty is distinct from uncertainty about restoration. After each reduced-payment year, full scheduled payments resume with probability $q_d = 1/d$ for finite expected duration d, and $q_d = 0$ under permanence. The restoration hazard is common to household beliefs and the objective payment process. Households in the anticipated environment incorporate this duration risk from model entry; those in the surprise environment learn it when payments fall. Both environments face the same realized onset date and distribution of restoration histories.

4.3. Counterfactual scenarios

For each expected duration d, including permanence, I compare the three environments described above. Household parameters, initial condition at age 18, and idiosyncratic shock histories are held fixed across environments.

The full-payment benchmark (O) is the reference environment. Households assign zero probability to payment-shortfall onset and receive full scheduled OASI payments at every age. This payment path is exogenous. The model does not specify the financing adjustments needed to sustain full payments or include their costs in household welfare. The comparisons measure household incidence conditional on the specified payment regimes.

The unexpected-shortfall environment (U) introduces an unexpected reduction in the retirement benefit. Through the end of 2033, households follow the full-payment regime's decision rules. At the start of 2034, they learn that the scheduled benefit is reduced to $\phi = 0.78$. Their full scheduled payments will resume after each reduced-payment year with probability q_d. From that point onward, they re-optimize and follow new optimal decision rules. This environment allows behavioral adjustment after the shortfall begins but contains no anticipation or preparation for it.

In the anticipated-shortfall environment (A), households incorporate the subjective onset hazard $\lambda_b(a)$ and restoration hazard q_d into their decisions from age 18 onward. Each household faces the same realized depletion date and restoration history as its counterpart in U. The two environments share the same path of paid-benefit fractions. Actual benefits can nevertheless differ because anticipation changes labor supply, covered earnings, and benefit entitlements as a result. Differences between environment A and U capture the consequences of anticipation or self-insurance, including how choices made before reserve depletion affect subsequent adjustment and welfare.

4.3.1. Decomposition

Let Y_{bd}^{j} denote an outcome for birth cohort b under expected shortfall duration d in environment $j \in \{O, U, A\}$. The full-payment benchmark is identical across duration scenarios. I retain its d subscript to make the comparisons explicit. Define

$$\Delta_{bd}^{U} = Y_{bd}^{U} - Y_{bd}^{O}, \qquad \text{unexpected-shortfall effect,} \tag{25}$$

$$\Delta_{bd}^{A} = Y_{bd}^{A} - Y_{bd}^{U}, \qquad \text{anticipation effect,} \tag{26}$$

$$\Delta_{bd}^{T} = Y_{bd}^{A} - Y_{bd}^{O}, \qquad \text{total anticipated-shortfall effect.} \tag{27}$$

The unexpected-shortfall effect Δ_{bd}^{U} measures the consequences of an unexpected payment reduction relative to continued full payment. It includes the loss of payment income and the subsequent adjustment of consumption, saving, employment, human-capital investment, and claiming.

The anticipation effect, Δ^A_{bd}, measures how outcomes change when households prepare for the shortfall. The comparison holds the realized depletion date and restoration histories fixed. Differences arise from anticipatory or self-insurance choices and their consequences for subsequent household states and behavior. Preparation can support consumption during the shortfall, but may require reallocation of earlier consumption or leisure. This is the key insight that self-insurance comes with cost, and it varies by the position in life cycle. Hence, its welfare effect is ambiguous ex ante. Moreover, because subjective onset beliefs need not coincide with the objective event process, this comparison does not measure the value of receiving accurate information.

The total anticipated-shortfall effect, Δ^T_{bd}, compares the anticipated environment directly with the full-payment benchmark. It is the primary comparison for assessing the welfare incidence across generations. For outcomes measured in levels, the decomposition is additive,

$$\Delta^T_{bd} = \Delta^A_{bd} + \Delta^U_{bd}.$$

The same identity holds for expected lifetime utility differences. Consumption-equivalent welfare measures, however, are calculated separately for each pair of environments. Since these measures apply a nonlinear transformation to utility, their component effects generally do not sum to the total consumption-equivalent effect. Section 5 reports all three comparisons for each generation and expected shortfall duration.

4.4. Measuring welfare cost

The counterfactuals change several aspects of household life-cycle: consumption, labor supply, claiming Social Security benefits, and the savings. Consumption alone does not capture their welfare consequences. I evaluate these changes using expected lifetime utility and express the resulting utility differences in a percentage of lifetime consumption.

The measure asks by what percentage would consumption in the reference environment have to change at every age to make the cohort as well off as it is in the target environment. For example, a welfare cost of 1 percent means that households would be equally well off, on average, under the target environment or under the reference allocation with consumption reduced by 1 percent through lifetime. Labor supply, claiming age, and savings remain at their reference values in this calculation. Their differences across environments still contribute to the utility difference being expressed in consumption units.

The reference environment depends on the comparing environment. For the unexpected-shortfall and total anticipated-shortfall effects, the reference is the full-payment benchmark O. For the anticipation effect, the reference is the unexpected-shortfall environment U, and the target is the anticipated environment A. The latter comparison asks whether self-insuring for the shortfall raises or lowers welfare relative to encountering the same payment regime without preparation.

In every comparison, a positive reported value denotes a welfare cost. The target delivers lower expected lifetime utility than the reference. A negative value means a welfare gain. Thus, a negative anticipation effect means that preparation improves welfare relative to the surprise environment, even if the anticipated shortfall remains costly relative to full payment.

Let $\mathcal{U}_{bd}^{j}$ denote expected lifetime utility for birth cohort b under expected shortfall duration d in environment j. For stochastic restoration, I first average each person's lifetime utility over restoration histories and then average across people within the cohort. For a target environment j and reference environment k, let $\mathcal{U}_{bd}^{k}(s)$ denote reference utility when consumption at every age from 18 through 80 is multiplied by $s > 0$, with all other reference choices and terminal assets held fixed. The consumption-equivalent welfare cost is defined by

$$\mathcal{U}_{bd}^{k}(s_{bd}^{j,k}) = \mathcal{U}_{bd}^{j}(1), \quad \mathrm{CEV}_{bd}^{j,k} = 100\left(1 - s_{bd}^{j,k}\right). \tag{28}$$

This adjustment translates the utility difference into consumption units. It does not require households to re-optimize.

The model's separable CRRA consumption utility makes the adjustment straightforward to compute. Define the reference consumption-utility component as

$$\mathcal{C}_{bd}^{k} = \mathbb{E}\left[\sum_{a=18}^{80} \beta^{a-18} \psi_a(M_a) \frac{(c_{ba}^{k})^{1-\eta}}{1-\eta}\right],$$

and collect the remaining utility components in $\mathcal{N}_{bd}^{k} = \mathcal{U}_{bd}^{k}(1) - \mathcal{C}_{bd}^{k}$. Scaling consumption then gives

$$\mathcal{U}_{bd}^{k}(s) = s^{1-\eta}\, \mathcal{C}_{bd}^{k} + \mathcal{N}_{bd}^{k}.$$

Consequently, the required multiplier is

$$s_{bd}^{j,k} = \left(\frac{\mathcal{U}_{bd}^{j}(1) - \mathcal{N}_{bd}^{k}}{\mathcal{C}_{bd}^{k}} \right)^{1/(1-\eta)} . \tag{29}$$

Discounting and family-size adjustments remain embedded in the consumption component.

The multiplier equates cohort-average utility. It is a cohort-level welfare measure, rather than an average of separately calculated household consumption equivalents. Each target–reference comparison has its own multiplier, so the reported consumption-equivalent effects generally do not add across the decomposition.

5. Intergenerational Welfare Incidence

Who bears the largest welfare loss from a Social Security payment shortfall, and how does the answer depend on how long reduced payments are expected to last? I compare five birth cohorts—1957, 1965, 1975, 1982, and 2005—that are ages 76, 68, 58, 51, and 28 at the common 2033 Social Security trust-fund depletion date. These cohorts encounter the same payment reduction at different stages of life, with different exposure to reduced retirement benefits and different opportunities to adjust saving, work, human-capital investment, and claiming.

Holding the depletion date and payment reduction fixed, I vary the expected time until full scheduled payments resume, including permanent reduced payments as the limiting case. A short duration primarily exposes cohorts already receiving benefits or approaching retirement horizon. As expected duration increases, the shortfall reaches more of younger cohorts' retirement years and shifts the welfare burden toward them.

To explain this shift, I compare welfare across the full-payment, surprise-shortfall, and anticipated-shortfall environments. The surprise-shortfall comparison measures the welfare loss when households can adjust and re-optimize only after reduced payments begin. The anticipated-shortfall comparison allows them to adjust beforehand as well. Holding the realized paid-benefit fractions fixed across these two environments allows me to understand the effect of self-insurance on welfare, and how its consequences differ with the expected duration of reduced payments.

5.1. Varying expected reduced-payment duration

The experiment varies how long reduced payments are expected to last, holding all else fixed. For each scenario, OASI reserves are depleted at the end of 2033 and payments fall to $\phi = 0.78$ of scheduled benefits in 2034. Restoration resumes payments to their full scheduled benefits but does not compensate the lost amount of earlier benefits. The welfare comparisons measure the incidence of these payment histories without including any associated tax increases or spending reductions.

I consider expected durations of $d \in \{1, 2, 3, 5, 10, 20, 40\}$ years together with permanent reduced payments. For each finite d, full payments resume with probability $q_d = 1/d$ after each year in the reduced-payment regime:

$$\Pr(F_{1,t+1} \mid L_t) = q_d, \qquad \Pr(L_{t+1} \mid L_t) = 1 - q_d, \qquad q_d = \frac{1}{d},$$

where L denotes reduced payments and F_1 denotes restored full payments. The realized duration D follows a geometric distribution with mean d. For example, $d = 5$ implies a 20 percent chance of restoration after each reduced-payment year, rather than a known restoration date. At $d = 1$, restoration occurs with certainty after one year; under permanence, $q_d = 0$ and full payments never resume. The restoration hazard governs both household expectations and the simulated payment histories.

Welfare effects are expressed as percentages of lifetime consumption using the CEV defined in equation (29). I report three comparisons: the cost of an unexpected shortfall relative to full payment ($\mathcal{C}^U_{cd}$), the additional cost or benefit of anticipation or self-insurance relative to surprise ($\mathcal{C}^A_{cd}$), and the total cost of an anticipated shortfall relative to full payment ($\mathcal{C}^T_{cd}$). Positive values indicate welfare losses, while negative values indicate welfare gains. Each CEV is calculated using its own reference environment as discussed above, so the surprise and anticipation CEV need not sum to the total CEV.

For each cohort and expected duration, I simulate 20,000 individuals. The same individuals face the same underlying idiosyncratic shocks across the three environments. However, their endogenous decisions can differ. For each individual, I first average lifetime utility over possible restoration histories. Then I average across individuals and calculate the CEV from these cohort-average utilities. Thus, the reported CEV measures a cohort-level welfare effect rather than an average of individual CEV.

5.2. The duration frontier

Figure 5 shows how the expected duration of reduced payments changes which cohorts bear the largest welfare costs. The estimates use HRS-disciplined, index-mapped subjective hazards. Filled circles report the total cost of an anticipated shortfall relative to full payment, while open squares report the cost of an unexpected shortfall relative to the same benchmark. The largest model-implied welfare cost shifts from the 1965 cohort at five years to the 1975 cohort at twenty years. Under permanence, the three youngest cohorts bear nearly equal costs. Figure E.6 reports all eight duration scenarios.

At an expected duration of five years, the 1965 cohort has the largest total cost, at 0.34 percent of lifetime consumption. This cohort is age 68 at depletion and receives reduced benefits from the start of the shortfall. The 1957 cohort is also immediately affected, but has fewer remaining benefit years. Because the remaining years that are exposed to insolvency is smaller, the 1957 cohort's total cost is 0.204 percent, lower than that of the 1965. Younger cohorts are less exposed because the Congress can make a restoration reform before they reach retirement age. At twenty years, the largest estimate is 0.65 percent for the 1975 cohort, compared with 0.61 percent for 1965. The 2005 cohort remains less affected, at 0.14 percent.

Under permanent reduced payments, total welfare costs rise from 0.36 percent for the 1957 cohort to 0.89 percent for 1965, then level off approximately 1.26 percent for the three younger cohorts. All retirement benefits of these younger cohorts are subject to the permanent reduction, whereas older cohorts received part of their full scheduled benefits before depletion. Their similar total costs nevertheless conceal different welfare consequences of self-insurance. Across the three younger cohorts, the surprise-shortfall cost increases toward the youngest cohort, while the anticipation comparison becomes more favorable. These opposing effects leave their total welfare costs close together.

The open squares report welfare costs when households do not anticipate the shortfall but re-optimize once payments fall. Longer shortfalls shift the burden toward younger cohorts even in this environment. As expected duration increases from five years to permanence, the surprise-shortfall cost rises modestly for the 1957 cohort, from 0.17 to 0.23 percent of lifetime consumption. But this increases sharply for the 2005 cohort, from approximately zero to 1.33 percent. Anticipation changes this pattern differently across cohorts. Under permanence, estimated welfare costs are higher with anticipation for the four older cohorts and lower for 2005. Section 5.4 measures these anticipation effects directly, using the surprise environment

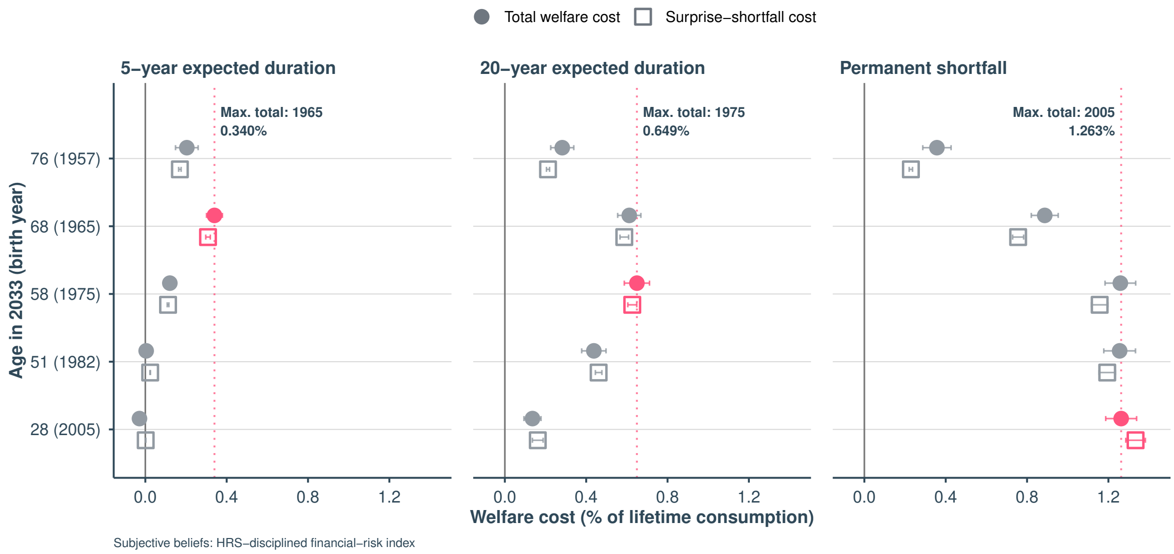

FIGURE 5. Expected reduced-payment duration and intergenerational welfare costs

Note. Payments fall to 78 percent of scheduled benefits in 2034. Finite spells end each year with probability $1/d$, where d is expected duration; permanence means no restoration. Circles and squares report anticipated- and surprise-shortfall CEV costs relative to full payment. Positive values denote losses. Pink identifies the largest total-cost point estimates. Bars show paired 95-percent conditional Monte Carlo intervals with parameters and empirical inputs held fixed.

as the reference environment.

Model-implied total welfare costs do not increase at every step as expected duration rises. For the 1957 cohort, the estimate falls slightly from 0.282 percent at twenty years to 0.275 at forty years, even though its surprise-shortfall welfare cost increases. Some short-duration welfare costs are also slightly negative. For example, the estimated total welfare cost for the 2005 cohort at five years is -0.029 percent.

5.3. Benefit Exposure

Expected duration determines how much of a cohort's retirement overlaps with reduced payments. A brief shortfall reaches households already receiving benefits, while a persistent shortfall also reaches cohorts whose retirement lies farther in the future. This timing difference helps explain why the welfare burden shifts toward younger cohorts as duration increases. The welfare estimates alone do not show its magnitude because they also incorporate changes in saving, work, human-capital investment, and claiming. Hence, I measure benefit exposure using the scheduled payments households would receive in the full payment benchmark.

TABLE 2. Mechanical benefit exposure across cohorts and durations

	Exposure share E_{cd} (percent)			Pr(ever reduced)	Expected reduced years
Cohort (age in 2033)	$d = 1$	$d = 20$	Permanent	$d = 20$	Permanent
1957 (76)	5.1	18.3	19.7	1.000	4
1965 (68)	6.2	51.9	66.7	1.000	12
1975 (58)	0.0	55.1	100.0	0.785	17.17
1982 (51)	0.0	38.5	100.0	0.548	17.17
2005 (28)	0.0	11.8	100.0	0.168	17.17

Note. Exposure is the discounted share of full-payment benchmark OASI payments falling in reduced-payment years (equation (30)). Permanent denotes zero restoration hazard. Pr(ever reduced) reports the probability of receiving at least one reduced payment when expected duration is twenty years. Expected reduced years counts years with positive reduced OASI payments through the model's terminal age of 80, averaged across individuals under permanence. All measures use the full-payment benchmark's simulated earnings and claiming paths.

Let s_{ca} denote the share of cohort c discounted lifetime scheduled OASI payments received at age a along the benchmark. These shares sum to one. Let p_{cda} denote the probability that payments are reduced when the cohort reaches that age, given expected duration d. Expected benefit exposure is

$$E_{cd} = \sum_a s_{ca} p_{cda}. \tag{30}$$

This measure is the expected share of scheduled benefit wealth exposed to the shortfall or insolvency. An exposure of zero means that no scheduled payments fall within the reduced-payment regime. An exposure of one means that the entire benefit stream is exposed to insolvency. WIth payments reduced by 22 percent, the corresponding expected accounting loss is $0.22E_{cd}$ of scheduled benefit wealth. For each cohort, I hold its benchmark payment shares fixed as duration varies. Changes in exposure therefore arise from the probability that reduced payments persist into its benefit-receipt years. This calculation uses the benchmark allocation to measure benefit exposure.

Table 2 reports benefit exposure at three durations, together with the probability of receiving any reduced payment and the expected number of reduced-payment years. Under a one-year duration, only the two oldest cohorts receive reduced payments. The exposed shares of lifetime benefit wealth are 5.1 percent for 1957 and 6.2 percent for 1965. The three younger cohorts have zero exposure because full payments resume before they begin receiving retirement benefits.

At an expected shortfall duration of twenty years, exposure reaches the highest at 55.1 percent of discounted lifetime benefits for the 1975 cohort and falls to 38.5 percent for the 1982 cohort. Their probabilities of receiving at least one reduced payments are 78.5 and 54.8 percent, respectively. The 2005 cohort is less exposed: it has a 16.8 percent change of receiving a reduced payment, and 11.8 percent of its lifetime benefit wealth is exposed. Its later retirement leaves more time for full payments to resume before benefit receipt begins.

These differences reflect both the timing of benefit receipt and the share of lifetime benefits remaining when the shortfall begins. Older cohorts are immediately affected, but have already received part of their benefit stream. Younger cohorts begin receiving benefits later, so their exposure depends more strongly on whether reduced payments persist. Under permanence, 19.7 percent of the 1957 cohort's discounted lifetime benefits and 66.7 percent of the 1965 cohort's benefits are exposed. Exposure reaches 100 percent for the 1975, 1982, and 2005 cohorts because all their benefits arrive after depletion. These three cohorts receive reduced benefits for an average of 17.17 years, compared with 12 years for 1965 and 4 years for 1957. All benefit receipt years are counted through the model's terminal age of 80.

Figure 6 compares benefit exposure under five-year, twenty-year, and permanent shortfalls. Figure E.7 all eight duration scenarios. When the expected shortfall is brief, exposure is concentrated among cohorts already receiving benefits. As expected duration increases, a larger share of younger cohorts' retirement benefits becomes exposed. This timing pattern helps explain why longer shortfalls shift the welfare burden toward younger cohorts.

5.3.1. Why duration matters beyond exposed benefit loss

The benefit-exposure measure explains how the timing of reduced payments distributes expected losses across birth cohorts. It does not capture uncertainty about how much each cohort will ultimately lose. A shortfall that certainly lasts one year, a shortfall with an uncertain restoration date, and a permanent reduction differ in both respects. The following example separates expected benefit loss from duration uncertainty, clarifying why accounting exposure alone cannot summarize the consequences of the reduced payment.

For illustration, each cohort receives the same scheduled annuity by age. One unit annually from age 67 through age 110, discounted to 2026 at a factor of 0.97. Social Security trust-fund is depleted at the end of 2033, and payments fall to 78 percent of scheduled amounts in 2034. The annuity schedule is imposed, so there are no saving, work, or claiming decisions in this calculation. Differences across cohorts arise from when their scheduled payments overlap with the shortfall.

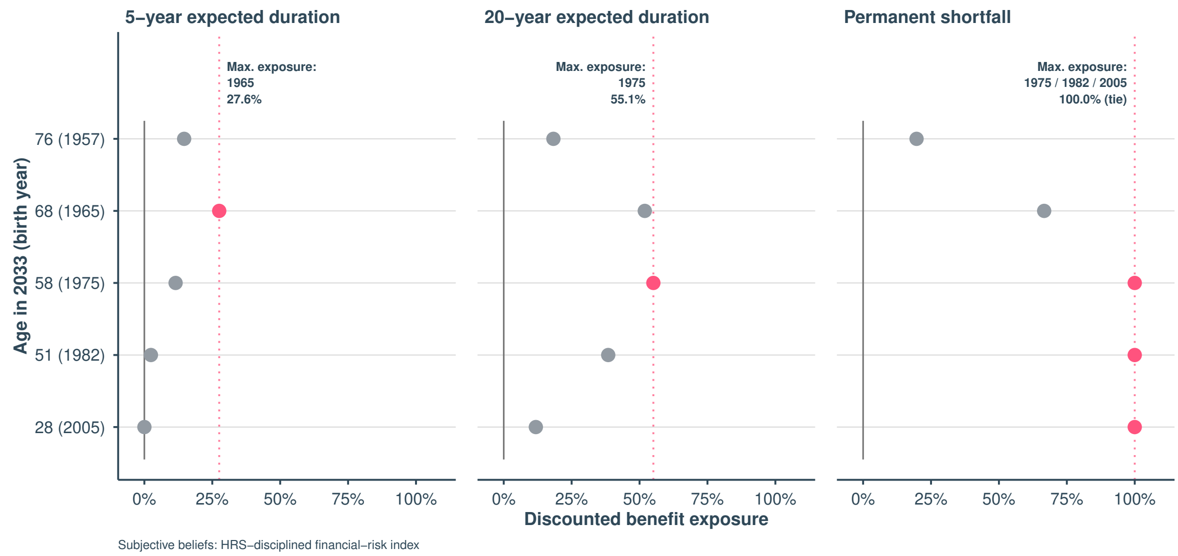

FIGURE 6. Mechanical benefit exposure across cohorts and durations

Note. The figure reports the discounted share of each cohort's scheduled OASI benefits expected to be received while payments are reduced. Benefits are evaluated along the full-payment benchmark path, so differences across cohorts and durations reflect life-cycle timing and the persistence of the reduced-payment state rather than behavioral responses. Panels select five-year, twenty-year, and permanent shortfalls; permanence means zero restoration hazard. Rows give age in 2033, with birth year in parentheses. Pink markers and dotted vertical lines identify the largest exposure in each panel, including tied cohorts. These highlights rank benefit exposure, rather than welfare costs.

Let m denote the ratio of paid to scheduled discounted annuity wealth under a realized payment history. Its exposure measures the average fraction of scheduled wealth received. Its certainty equivalent, $CE(m)$, also accounts for variation across payment histories, using the benchmark CRRA coefficient.[4] The total discount relative to scheduled annuity wealth is

$$1 - CE(m) = \underbrace{1 - \mathbb{E}[m]}_{\text{expected accounting loss}} + \underbrace{\mathbb{E}[m] - CE(m)}_{\text{uncertainty discount}}, \tag{31}$$

The first term records missing benefits on average. The second term captures the additional discount associated with uncertainty about the fraction received. Both are measured relative to scheduled OASI wealth. They are distinct from the structural model's estimated welfare costs.

Figure 7 shows why the two terms must be distinguished. Exposed losses increase with

[4]For clarity, the certainty equivalent is computed as

$$CE(m) = u^{-1}(\mathbb{E}[u(m)]) = \begin{cases} \left(\mathbb{E}[m^{1-\gamma}]\right)^{1/(1-\gamma)}, & \gamma \neq 1, \\ \exp\{\mathbb{E}[\log m]\}, & \gamma = 1. \end{cases}$$

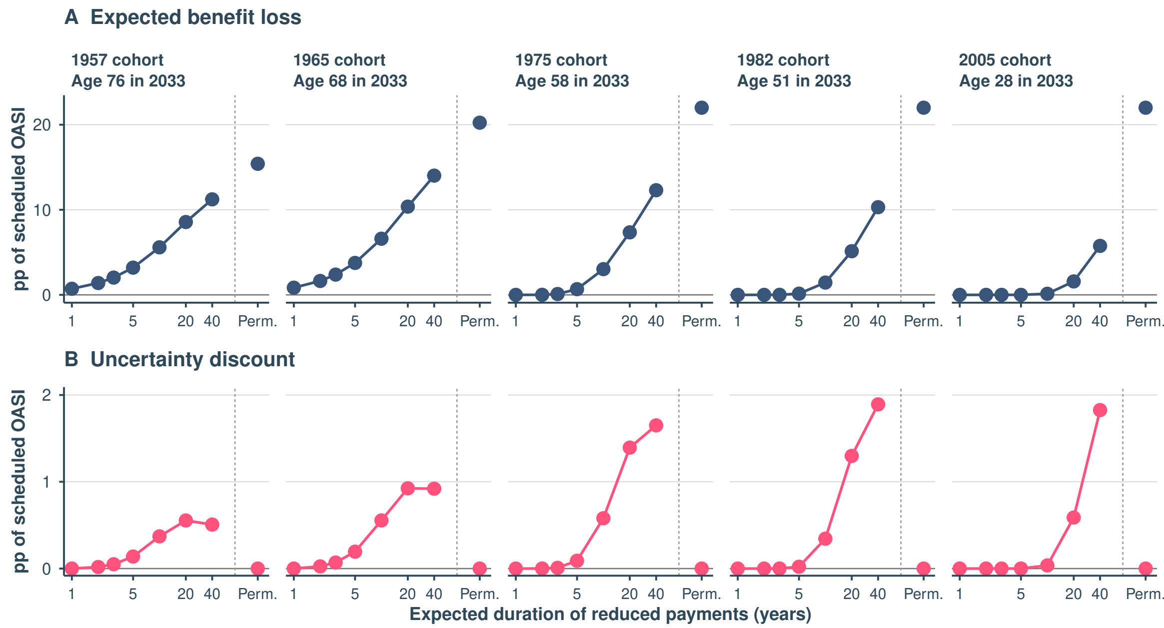

FIGURE 7. A fixed-behavior benchmark for benefit loss and uncertainty

Note. Columns identify birth cohorts and their ages in 2033. Row A reports $100(1 - \mathbb{E}[m])$; row B reports $100(\mathbb{E}[m] - CE(m))$, both in percentage points of scheduled OASI wealth. The scheduled annuity pays one unit annually at ages 67–110, discounted to 2026 at 0.97. Payments fall to 78 percent in 2034. Restoration has annual probability (1/d) for finite expected duration (d); Perm. denotes no restoration. Finite durations use a logarithmic horizontal scale. The CRRA coefficient is 4.0363.

duration as more scheduled payments fall within the shortfall. For younger cohorts, those losses become substantial when reduced payments persist into retirement. The uncertainty discount follows a different pattern. It is zero for both a one-year duration and permanence because the payment history is known in either case. For other expected durations, restoration is uncertain, and different realized durations produce different benefit losses. A permanent shortfall can generate a large expected loss even though its duration carries no uncertainty.

The distinction also matters when expected losses are identical. For the 1982 cohort, consider three restoration processes that each produce an expected accounting loss of 4.2 percentage points of scheduled OASI wealth. A known restoration date produces no uncertainty discount. A constant annual restoration hazard produces a discount of 1.1 percentage points, while a mixture of quick restoration and permanent reduced payments produces a discount of 2 percentage points. These processes deliver the same benefits on average, but differ in the possibility of a persistent shortfall. Appendix C.4 provides the construction and further comparisons.

The example clarifies what the duration frontier in Section 5.2 captures. Changing restoration process alters both expected benefit losses and uncertainty about those losses, with different consequences on cohorts. Translating these payment-risks into lifetime welfare requires a structural model, which accounts for other household resources and the costs of adjusting saving, work, investment, and claiming.

5.4. The Welfare Effect of Anticipation

The preceding exercises show how payment duration changes benefit exposure and uncertainty about future payments. I now ask whether preparing for those losses improves lifetime welfare. This comparison is important because resources available when payments fall reveal only one side of self-insurance. Households can accumulate more assets or adjust work or change the timing of claiming retirement benefits. But obtaining those resources may require earlier sacrifices of consumption and leisure. The anticipation effect evaluates this reallocation.

Figure 8 reports the anticipation CEV, $\mathcal{C}^{A}_{cd}$, for five, twenty-year, and permanent shortfalls. A positive value means that the anticipated allocation delivers lower lifetime welfare than the surprise environment. A negative value means that it delivers higher welfare or welfare gain. Each comparison is expressed as a percentage of consumption in the surprise environment and is computed directly from the two allocations.

Under permanent reduced payments, anticipation raises the estimated welfare cost by 0.129 percent of reference lifetime consumption for the 1957 cohort, 0.134 percent for 1965, and 0.106 percent for 1975. For the 2005 cohort, anticipation instead produces a welfare gain of 0.073 percent. The 1982 estimate is a cost of 0.063 percent, but it is not statistically different from zero from Monte Carlo simulations. Thus, under the same permanent payment reduction, anticipation improves the youngest cohort's lifetime allocation while worsening it for several older cohorts.

These welfare differences cannot be inferred from the resources households hold when the shortfall begins. Appendix E.2 shows that the 1957, 1965, and 1975 cohorts enter the first reduced-payment year with more assets and higher consumption under anticipation than under surprise. Nevertheless, their lifetime anticipation effects are costs. The additional resources help them absorb reduced payments, but the lifetime comparison also values the earlier choices required to obtain those resources. Because subjective hazards enter decisions from age 18, these effects include adjustments made long before the realized depletion date, including for cohorts already retired when payments fall. The life cycle profiles document changes in savings, consumption, labor supply, human-capital investment and claiming.

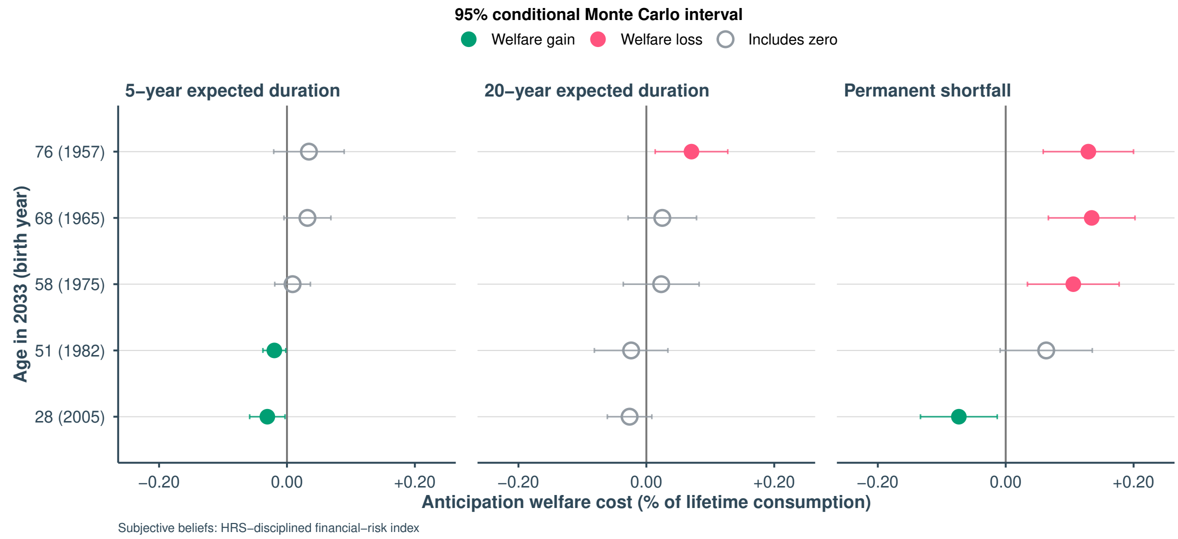

FIGURE 8. Welfare effect of anticipation by cohort and duration

Note. Points compare anticipated and surprise shortfalls under the same objective payment process: payments fall to 78 percent in 2034, with annual restoration probability $(1/d)$ for finite expected duration (d) and no restoration under permanence. Values are percentages of lifetime consumption in the surprise environment; positive values denote welfare costs. Bars are paired 95-percent conditional Monte Carlo intervals from 20,000 simulated individuals per cohort–duration cell. Green and pink filled markers identify intervals below and above zero, respectively; gray open markers identify intervals including zero. Rows show age in 2033 and birth year. Intervals cover finite-simulation variation only.

The interpretation depends on the distinction between subjective beliefs and the objective payment process. Households choose under the index-mapped distribution of perceived insolvency risk, while welfare is evaluated under the specified restoration process. Positive welfare costs describe the consequences of preparation or self-insurance under these belief inputs.

Expected duration also changes the comparison. At five years, anticipation produces estimated welfare gains of 0.02 and 0.031 percent for the 1982 and 2005 cohorts, respectively. At twenty years, only the 1957 cohort's interval excludes zero, indicating a welfare cost of 0.07 percent. Figure E.8 reports the complete welfare analysis of anticipation.

The objective of this exercise is to show how anticipation changes the burden implied by exposure alone. Self-insurance can support consumption after payment fall without improving the lifetime allocation, and its welfare consequences vary across cohorts and expected durations. Intergenerational incidence therefore depends both on which benefits are exposed and on the choices households make in response to perceived risk.

5.5. Accounting Incidence versus Welfare Incidence

An accounting calculation asks a simpler question: how much scheduled benefit income goes unpaid? Comparing the accounting and structural model establishes whether the cohorts losing the most benefits are also those bearing the largest welfare cost. Table 3 makes this comparison under permanent reduced payments, with benefits falling to 78 percent of scheduled amounts in 2034.

Panel A of Table 3 calculates unpaid benefits using earnings and claiming decisions from the full-payment benchmark. It reports the loss as a share of discounted scheduled OASI benefits, as a share of discounted benchmark lifetime consumption, and as a present-value index measured in 2033. Panel B of Table 3 reports the surprise- and anticipated-shortfall CEV relative to full payment. These welfare measures value the resulting consumption, leisure, and bequests after households adjust their choices.

TABLE 3. Accounting incidence versus welfare incidence

	1957	1965	1975	1982	2005
Panel A. Accounting loss: unpaid scheduled OASI payments, behavior fixed					
Share of scheduled OASI wealth lost (%)	4.3	14.7	22.0	22.0	22.0
Loss as % of lifetime consumption	0.29	0.99	1.49	1.49	1.49
Loss in present value at 2033 (peak $= 100$)	34.1	90.4	100.0	80.8	40.1
Panel B. Welfare: consumption-equivalent cost (% of lifetime consumption)					
Surprise shortfall	0.23	0.76	1.16	1.19	1.33
Anticipated shortfall	0.36	0.89	1.26	1.26	1.26

Note. Panel A values unpaid scheduled OASI payments along the full-payment benchmark path, with payments reduced by 22 percent from 2034 onward and discount factor $\beta = 0.97$. The first two rows use cohort-specific age-18 valuation; the third values losses in 2033 and normalizes the largest to 100. Panel B reports permanent-shortfall CEV costs relative to full payment, using HRS-disciplined, index-mapped subjective hazards. Positive values denote welfare losses. The accounting and welfare measures value different objects; their difference is not a measure of income recovered through adjustment.

The common-date accounting measure places substantially less weight on younger cohorts' losses because their unpaid benefits arrive farther in the future. With the 1975 cohort's present-value loss normalized to 100, the 2005 cohort's index is only 40.1. Their anticipated-shortfall welfare costs, however, are both approximately 1.26 percent of lifetime consumption. The ranking of loss also reverses between 1965 and 2005 cohort: the younger cohort has a smaller present-value accounting loss but a larger welfare loss.

This difference partly reflects what calendar-time discounting measures. In the full-payment benchmark, the 1975 and 2005 cohorts have identical scheduled benefit profiles by age, but the younger cohort receives those benefits thirty calendar years later. At the assumed discount factor, that delay reduces the present value to 40 percent of the older cohort's value. It does not imply that the younger cohort loses a smaller share of its own retirement income.

The accounting loss and the model-implied CEV answer different questions even though both are expressed relative to benchmark consumption. For the 1975, 1982, and 2005 cohorts, unpaid benefits equal 1.49 percent of discounted lifetime consumption. Their welfare cost of approximately 1.26 percent means something different: reducing benchmark consumption by 1.26 percent at every age would produce the same lifetime utility as the anticipated shortfall. This welfare comparison accounts for the timing of consumption, changes in leisure, and bequest under the households' chosen allocation.

The accounting percentage can be either larger or smaller than the model-implied welfare cost. For the three younger cohorts, it is larger; for 1957, the accounting loss is 0.29 percent while the welfare cost is 0.36 percent. Missing benefit income does not translate into a percentage loss of welfare. Nor does the gap measure how much income households recover through adjustment, because the two measures value different objects. Table 3 establishes why accounting unpaid dollars alone are insufficient to assess intergenerational welfare incidence. Appendix E.4 provides the calculation in details.

5.6. Summary of the mechanism

The distributional consequences of OASI reserve depletion depend critically on what happens afterward. The depletion date and the size of the payment reduction alone are insufficient to identify which generations lose the most. Restoration determines how much of each cohort's retirement overlaps with reduced payments. A brief shortfall concentrates exposure among current beneficiaries, while a persistent shortfall reaches benefits that younger workers have yet to receive. Any assessment of intergenerational incidence must therefore specify how long reduced payments may last.

The capacity to prepare must also be evaluated over the household's entire lifetime. More assets or higher consumption when payments fall can indicate successful self-insurance while concealing its earlier costs. The relevant welfare question is whether the resulting allocation, including those earlier sacrifices, leaves the household better off. Under the estimated subjective onset beliefs and the specified objective payment process, anticipation improves this allocation for some cohorts and worsens it for others. Resources accumulated before depletion are an incomplete measure of the benefits of self-insurance.

Restoration expectations matter through both exposure and uncertainty. Even a long expected shortfall leaves some possibility that full payments return before younger cohorts retire. Permanence removes that possibility entirely. It can consequently impose large expected benefit losses while carrying no uncertainty about restoration. This distinction explains why permanent reductions should be treated as a separate endpoint and why expected losses alone cannot summarize the risks households face.

Counting unpaid benefits alone can give a misleading picture of which generations bear the greatest burden. Younger cohorts' losses receive less weight when discounted to a common calendar date, even when those losses represent a large share of their retirement income. Households can also self-insure the shortfall through saving, work, and claiming decisions, but doing so can require sacrifices of consumption or leisure at other ages. The welfare measure accounts for these consequences throughout life. It therefore assesses both how households live with reduced benefits and what it costs them to prepare, which the present value of missing benefits cannot capture.

5.7. Discussion

Luttmer and Samwick (2018) establish that perceived uncertainty about Social Security benefits carries a substantial welfare cost. They elicit subjective benefit distributions and certainty equivalents, finding higher risk premia among older respondents. They interpret this age pattern as consistent with fewer opportunities to adjust future saving and labor supply. Their survey valuation can incorporate respondents' anticipated adjustments without specifying a behavioral model. This paper addresses a complementary question: how a common payment shortfall distributes lifetime welfare losses across cohorts when its duration and households' responses are specified explicitly.

The age gradient in Luttmer and Samwick (2018) describes how households value protection against perceived benefit uncertainty. However, it does not establish which generation bears the greatest lifetime loss from a common payment shortfall. Older households may value protection highly because they have fewer opportunities to adjust, yet have relatively few benefit years remaining. Younger households may have greater adjustment capacity while facing a reduction in their entire retirement benefit stream. Expected shortfall duration determines which of these exposures matters, so greater vulnerability to uncertainty need not imply greater welfare loss from reduced payments. A second distinction concerns the interpretation of self-insurance. Their survey premium can incorporate whatever adjustments respondents expect to make, but does not separately identify whether anticipating a particular shortfall improves the lifetime allocation relative to being surprised. Preparation can produce more assets and higher consumption when payments fall while still leaving households worse off once earlier sacrifices are counted. This finding concerns choices made under subjective onset beliefs and evaluated under a specified payment process. It complements their valuation of certainty by showing why both the cohort ranking and the welfare consequences of preparation depend on the payment history against which those choices are evaluated.

The contribution is consequently an explanation of intergenerational incidence under alternative payment durations, together with a model-based assessment of anticipatory adjustment. The welfare objects differ from those in Luttmer and Samwick (2018): their risk premium measures the discount associated with uncertainty around expected benefits, whereas the total CEV here includes the consequences of receiving lower benefits and changing household choices, expressed relative to lifetime consumption. The magnitudes and cohort rankings therefore answer different questions. Together, the approaches show why understanding Social Security risk requires both evidence on how households value uncertainty and an account of when payment losses occur and how households respond.

6. Conclusion

How long reduced OASI payments last determines which generations bear the largest welfare cost. In the model, a 22 percent payment reduction beginning in 2034 falls most heavily on the 1965 cohort when its expected duration is five years. Under permanent reductions, the largest burden fall on the 1975, 1982, and 2005 cohorts. The reason is the timing of benefit receipt: a brief shortfall mainly affects current retirement beneficiaries, while a persistent shortfall reaches retirement income that younger cohorts have yet to receive.

Preparation or self-insurance changes these costs, but its success must be assessed over the life cycle. Households can enter the shortfall with more assets and higher consumption while having sacrificed enough earlier consumption or leisure to leave lifetime welfare lower. Under the HRS-disciplined subjective hazards, anticipation reduces the permanent-shortfall cost for the youngest cohort and raises it for several older cohorts. Accounting losses also give an incomplete picture. The 2005 cohort's unpaid benefits are worth about 40 percent of the 1975 cohort's loss when valued in 2033, yet their welfare costs are nearly identical. Discounted benefit income and lifetime welfare measure different consequences of the same shortfall.

These estimates describe intergenerational incidence conditional on the assumed payment process, calibrated parameters, and belief inputs. Within this model results, the implication is clear: a comparison of generations must specify both their ages when Social Security becomes insolvent and how long the payment reduction may last. The trust-fund depletion date alone does not identify who bears the largest welfare costs.

References

Baker, Scott R., Nicholas Bloom, and Steven J. Davis. 2016. "Measuring Economic Policy Uncertainty." *Quarterly Journal of Economics* 131 (4): 1593–1636. https://doi.org/10.1093/qje/qjw024.

Caliendo, Frank N., Aspen Gorry, and Sita Slavov. 2019. "The Cost of Uncertainty about the Timing of Social Security Reform." *European Economic Review* 118: 101–125. https://doi.org/10.1016/j.euroecorev.2019.05.008.

De Nardi, Mariacristina. 2004. "Wealth Inequality and Intergenerational Links." *Review of Economic Studies* 71 (3): 743–768. https://doi.org/10.1111/j.1467-937X.2004.00302.x.

De Nardi, Mariacristina, Eric French, John Bailey Jones, and Rory McGee. 2025. "Why Do Couples and Singles Save during Retirement? Household Heterogeneity and Its Aggregate Implications." *Journal of Political Economy* 133 (3): 750–792. https://doi.org/10.1086/733421.

Diamond, Peter A. 1997. "Insulation of Pensions from Political Risk." In *The Economics of Pensions: Principles, Policies, and International Experience*, edited by Salvador Valdés-Prieto, pp. 33–57. Cambridge University Press. https://www.nber.org/papers/w4895.

Durbin, James and Siem Jan Koopman. 2012. *Time Series Analysis by State Space Methods.* No. 38 in Oxford Statistical Science Series. 2 ed. Oxford: Oxford University Press.

Fan, Xiaodong, Ananth Seshadri, and Christopher R. Taber. 2024. "Estimation of a Life-Cycle Model with Human Capital, Labor Supply, and Retirement." *Journal of Political Economy* 132 (1): 48–95. 10.1086/726232.

Friedberg, Leora. 2000. "The Labor Supply Effects of the Social Security Earnings Test." *Review of Economics and Statistics* 82 (1): 48–63. https://doi.org/10.1162/003465300558623.

Gourinchas, Pierre-Olivier and Jonathan A. Parker. 2002. "Consumption over the Life Cycle." *Econometrica* 70 (1): 47–89. https://doi.org/10.1111/1468-0262.00269.

Gruber, Jonathan and Peter Orszag. 2003. "Does the Social Security Earnings Test Affect Labor Supply and Benefits Receipt?" *National Tax Journal* 56 (4): 755–773. https://doi.org/10.17310/ntj.2003.4.03.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hubbard, R. Glenn, Jonathan Skinner, and Stephen P. Zeldes. 1995. "Precautionary Saving and Social Insurance." *Journal of Political Economy* 103 (2): 360–399. https://doi.org/10.1086/261987.

Hurd, Michael D. 2009. "Subjective Probabilities in Household Surveys." *Annual Review of Economics* 1: 543–562. https://doi.org/10.1146/annurev.economics.050708.142955.

Kitao, Sagiri. 2014. "Sustainable Social Security: Four Options." *Review of Economic Dynamics* 17 (4): 756–779.

Kitao, Sagiri. 2018. "Policy Uncertainty and Cost of Delaying Reform: The Case of Aging Japan." *Review of Economic Dynamics* 27: 81–100.

Luttmer, Erzo F. P. and Andrew A. Samwick. 2018. "The Welfare Cost of Perceived Policy Uncertainty:

Evidence from Social Security." *American Economic Review* 108 (2): 275–307. https://doi.org/10.1257/aer.20151703.

Manski, Charles F. 2004. "Measuring Expectations." *Econometrica* 72 (5): 1329–1376. https://doi.org/10.1111/j.1468-0262.2004.00537.x.

Rust, John and Christopher Phelan. 1997. "How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets." *Econometrica* 65 (4): 781–831. https://doi.org/10.2307/2171940.

Shoven, John B. and Sita N. Slavov. 2006. "Political Risk versus Market Risk in Social Security." Working Paper 12135, National Bureau of Economic Research. https://www.nber.org/papers/w12135.

U.S. Social Security Administration. 2025. "The 2025 Annual Report of the Board of Trustees of the Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds." , Board of Trustees, Federal Old-Age and Survivors Insurance and Federal Disability Insurance Trust Funds. https://www.ssa.gov/oact/tr/2025/.

Appendix A. Empirical Motivation: HRS Sample

TABLE A.1. Summary statistics for the HRS estimation sample

	Mean	SD	Median
A. Demographics			
Age	56.3	2.9	56
Female	59.3%		
Years of education	13.7	2.5	14
Black	12.9%		
Hispanic	8.0%		
Married or partnered	77.7%		
Fair or poor health	16.0%		
B. Economic variables			
Currently employed	79.8%		
Has pension on current job (among employed)^d	58.8%		
Annual earnings ($000, 2004 dollars)	35.3	113.7	24.4
Nonhousing wealth ($000, 2004 dollars)	94.1	323.0	9.6
Liquid wealth ($000, 2004 dollars)	24.2	90.2	4.5
IRA balance ($000, 2004 dollars)	72.2	342.9	0.0
C. Beliefs and expectations			
Perceived personal benefit risk (%)	60.8	28.2	60
Perceived general program risk (%)	68.1	24.1	75
Subjective probability of surviving to 75 (%)^a	65.7	27.1	75
Perceived chance of a major economic depression (%)^b	47.2	27.7	50
Expected probability of working after 62 (%)	47.2	37.8	50
Expected probability of working after 65 (%)^c	30.2	33.7	20

Note. $N = 3{,}359$ except where noted: $^aN = 3{,}289$, $^bN = 3{,}307$, $^cN = 3{,}354$; these three items have limited additional survey nonresponse within the estimation sample. $^dN = 2{,}665$, the employed subsample; pension coverage on the current job is undefined for respondents who are not currently working, so the share is computed only among the employed rather than treating non-workers as lacking a pension. Earnings, nonhousing wealth, liquid wealth, and IRA balance are right-skewed; the median is a more representative summary than the mean for those rows. A median IRA balance of 0.0 reflects that most respondents report no IRA holdings. Liquid wealth and IRA balance are components of nonhousing wealth and do not sum to it exactly because nonhousing wealth includes other financial assets. Binary variables report only the sample share.

Source. Author's calculations from the 2006 HRS Internet Survey.

TABLE A.2. Belief-to-intended-work regressions across specifications

	(1)	(2)	(3)	(4)
Panel A. Work after 62 ($W_{i,62}$)				
Perceived personal benefit risk (R_i^P)	0.830***	0.858***	0.915***	0.837***
	(0.277)	(0.281)	(0.279)	(0.267)
General program risk (R_i^G)	−0.318	−0.534*	−0.501	−0.278
	(0.315)	(0.320)	(0.316)	(0.305)
Observations	3,359	3,243	3,222	3,342
R^2	0.110	0.132	0.152	0.174
Panel B. Work after 65 ($W_{i,65}$)				
Perceived personal benefit risk (R_i^P)	0.773***	0.783***	0.838***	0.737***
	(0.252)	(0.255)	(0.254)	(0.245)
General program risk (R_i^G)	−0.423	−0.602**	−0.551*	−0.347
	(0.286)	(0.289)	(0.289)	(0.280)
Observations	3,362	3,249	3,225	3,345
R^2	0.077	0.104	0.108	0.123
Age fixed effects	Yes	Yes	Yes	Yes
Demographic and 2006 wealth/earnings controls	Yes	Yes	—	Yes
Subjective survival and depression expectations		Yes		
2004 (pre-belief) controls, incl. 2004 work status			Yes	
Pension coverage on current job (3-category)				Yes

Note. Each column reports coefficients from equation (23) estimated jointly on perceived personal benefit risk (R_i^P) and general program risk (R_i^G), both in units of 10 percentage points. Standard errors, clustered by household, are in parentheses. Column (1) is the specification used in Figure 2 and reported in the main text: age fixed effects, gender, education, race and ethnicity, marital status, self-reported health, and inverse-hyperbolic-sine-transformed nonhousing wealth and earnings, all measured in 2006. Column (2) adds the person's subjective probability of surviving to 75 and perceived chance of a major depression. Column (3) replaces the 2006 demographic and financial controls with their 2004 (pre-belief) counterparts, plus 2004 employment status and hours, to check whether the relationship reflects conditions already in place before the belief was elicited. Column (4) adds pension coverage on the current job as a three-category control (no pension, has a pension, or not currently working), since pension coverage is undefined for non-workers. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Source. Author's calculations from the 2006 (and, for column (3), 2004) HRS Internet Survey.

TABLE A.3. Predictive content of stated work plans for later full-time employment

	2008 (ages 60–61 in 2006)			2010 (ages 58–61 in 2006)		
	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable	$W_{i,62}$	Full-time work	Full-time work	$W_{i,62}$	Full-time work	Full-time work
Perceived personal benefit risk (R_i^P, per 10pp)	1.110*	0.240	0.020	1.604***	0.291	-0.189
	(0.585)	(0.661)	(0.669)	(0.432)	(0.503)	(0.481)
General program risk (R_i^G, per 10pp)	0.773	0.446	0.124	0.002	0.948	0.762
	(0.697)	(0.795)	(0.795)	(0.507)	(0.589)	(0.558)
Stated plan: work after 62, $W_{i,62}$ (per 1pp)			0.298***			0.369***
			(0.055)			(0.038)
2006 full-time work status		Yes	Yes		Yes	Yes
Household controls[a]	Yes	Yes	Yes	Yes	Yes	Yes
Observations	601	567	559	1,254	1,129	1,119
R^2	0.175	0.506	0.549	0.139	0.259	0.329

Note. Sample restricted to respondents who were ages 60–61 (columns 1–3) or 58–61 (columns 4–6) in 2006, so that all respondents have reached age 62 by the outcome year (2008 or 2010, respectively). Columns (1) and (4) regress the stated probability of working full time after 62, $W_{i,62}$ (0–100 scale), on the two risk measures and household controls; coefficients are already in percentage points of the stated probability. Columns (2)/(5) and (3)/(6) are linear probability models for full-time work in 2008/2010. They include the same regressors and 2006 full-time work status; columns (3) and (6) additionally include the 2006 stated plan $W_{i,62}$. Coefficients on the risk measures and $W_{i,62}$ are multiplied by 100 and expressed in percentage points. Standard errors, clustered by household, are in parentheses. [a] Same controls as column (1) of Table A.2: age fixed effects, gender, education, race and ethnicity, marital status, self-reported health, and inverse-hyperbolic-sine-transformed nonhousing wealth and earnings, all measured in 2006. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Source. Author's calculations from the 2006, 2008, and 2010 HRS.

Appendix B. Measuring the OASI Financial-Risk Index

This appendix describes the construction of the empirical subjective-onset-hazard series. The input history is monthly from January 1995 through March 2026. I construct parallel OASI and OASDI series, but the paper reports OASI as its headline measure.

B.1. Information set and timing

Table B.1 lists the four blocks. All inputs are oriented before filtering so that a larger value means more adverse OASI financial conditions. The three trust-fund variables come from the intermediate-assumption OASI projections in successive SSA Trustees Reports. A report enters the filter only in its release month. In other months, the trust-fund variables are missing in the measurement equation, although tables and figures carry forward the most recently published values.

TABLE B.1. Variables used in the OASI financial-risk index

Block	Public series	Oriented measurement
Trust-fund level	75-year actuarial balance, open-group unfunded obligation, and projected intermediate-assumption reserve-depletion year	Actuarial deficit, unfunded obligation, and inverse years to depletion
Macro stress	BLS total nonfarm payroll employment, average hourly earnings, and unemployment, obtained from FRED	Unemployment and negative year-over-year growth in the product of employment and average hourly earnings
Official uncertainty	Successive SSA Trustees Reports	Absolute changes in the three trust-fund measures between consecutive Trustees Reports
Economic-policy uncertainty	Baker, Bloom, and Davis (2016) monthly US news-based Economic Policy Uncertainty index	Log EPU, released to the index construction after the source's revision window closes

Note. The three trust-fund components receive equal within-block weights, as do the two macro-stress components. The headline weights are 0.30, 0.20, 0.10, and 0.40 in the order shown. Official text is excluded. Source: Author's calculations from SSA Trustees Reports, BLS, FRED, and the Baker, Bloom, and Davis (2016) US Economic Policy Uncertainty index.

Each input enters the construction only after it would have become public. Employment, earnings, and unemployment for month t enter in month $t + 1$, matching the BLS release lag. The paper uses latest-revised FRED histories shifted by one month rather than ALFRED first-release vintages. It therefore reproduces publication timing but not the data vintage available in real time; the one-sided history is an approximation to a real-time information set. The EPU index for month t can be revised in months $t+1$ and $t+2$ as newspaper archives fill in, so it enters in month $t+3$.

B.2. Transformations and measurements

Constructing the index takes two steps. This subsection turns each raw input into a standardized block measurement, one number per block per month. The next subsection filters those measurements over time and combines the four blocks into a single index.

Let m_{jt} denote component j in the month t it becomes available to the index construction, signed so that a larger value means worse OASI finances. The actuarial deficit is minus the reported actuarial balance. Depletion proximity is one over the number of years from the report year to the projected depletion year. Wage-bill stress is minus the year-over-year growth in the product of payroll employment and average hourly earnings.

Each component is standardized in two steps. I take logs of the three positive trust-fund components and the EPU index, leaving the others in levels, so that $m^*_{jt} = \log m_{jt}$ for those four series and $m^*_{jt} = m_{jt}$ otherwise. I then compute

$$z_{jt} = \frac{m^*_{jt} - \mu_j}{\sigma_j}, \qquad x_{jt} = \frac{\operatorname{asinh}(z_{jt}) - \mu^a_j}{\sigma^a_j},$$

where μ_j and σ_j are the mean and standard deviation of m^*_{jt}, and μ^a_j and σ^a_j are the mean and standard deviation of $\operatorname{asinh}(z_{jt})$.

The first step puts each component in standard-deviation units. The inverse hyperbolic sine then compresses extreme values without changing their order, so a ten-standard-deviation spike counts about three times, not ten times, a one-standard-deviation move. Because this compression changes the scale, the second step re-standardizes so that every component again has mean zero and unit variance, and equal weights within a block correspond to equal contributions. All means and standard deviations are computed once from observations until December 2011, the last HRS survey year used for calibration, and are held fixed.

The four block measurements are then simple averages of transformed components. Trust-fund level is the mean of transformed actuarial deficit, unfunded obligation, and depletion proximity. Macro stress is the mean of transformed unemployment and wage-bill stress. Official uncertainty is the mean absolute change in the three transformed trust-fund components between consecutive Trustees Reports, passed through the same fixed transformation. The EPU block is the transformed, lagged EPU index. If a component is missing, it enters as zero, its calibration mean, and the weights are not rescaled. A block is missing only when all of its components are missing.

B.3. Kalman-filtering and aggregation

The four blocks of an index are not observed at the same frequency. Trust-fund and official-uncertainty measurements change only when SSA releases a new Trustees Report, roughly once a year. Macro stress and EPU are observed every month. Filtering addresses two problems created by this timing issue. It carries a block forward between releases and, when a release arrives, updates the block gradually rather than all at once.

I filter each block measurement M_t^k separately, where k indexes the trust-fund, macro-stress, official-uncertainty, and economic-policy-uncertainty blocks. The construction therefore uses four univariate filters rather than one multivariate factor model.[5] Each block has a latent state F_t^k that evolves as

$$F_t^k = \phi_k F_{t-1}^k + \eta_t^k, \qquad \mathrm{Var}(\eta_t^k) = q_k,$$

and that is observed with noise,

$$M_t^k = H_k F_t^k + \varepsilon_t^k, \qquad \mathrm{Var}(\varepsilon_t^k) = R_k.$$

The persistence ϕ_k governs how much of the state carries from one month to the next, and q_k is the variance of the state innovation. The loading H_k converts state units into measurement units, and R_k is the variance of measurement error. I fix ϕ_k, q_k, H_k, and R_k as calibration choices rather than estimate them. The recursions are the standard scalar linear state-space updates (Hamilton 1994; Durbin and Koopman 2012).

The preceding subsection already reduces each block to a single standardized number per month. The measurement and the state are therefore in the same units, and $H_k = 1$ for each k. Setting $R_k = 1$ is a normalization rather than a claim about how noisy the data are. With a scalar state and measurement, the filter depends on the prior variance, state innovation variance

[5]This is not merely a computational simplification. In a single multivariate filter, the common state would be updated only by the series observed in a given month, so in the eleven months between Trustees Reports the macro and EPU series would receive all effective weight. Filtering the blocks separately and aggregating with fixed weights keeps the stated weights operative in every month, and it lets each block carry its own persistence.

q_k, and measurement variance only through the ratios $P^k_{1|0}/R_k$ and q_k/R_k. Scaling all three by a common constant leaves every filtered state unchanged, so fixing $R_k = 1$ fixes units only. Conditional on the fixed persistence and initialization, q_k governs how much weight a new observation receives.

The filter is a recursion of a point estimate of the block state and the variance of that estimate. I write $F^k_{t|s}$ for the estimate of F^k_t formed from measurements dated no later than month s, and $P^k_{t|s}$ for the variance of that estimate. The pair $t \mid t-1$ therefore describes beliefs held before month t's measurement is seen, and $t \mid t$ describes beliefs held after. Only $s \leq t$ appears anywhere in the construction, which is the sense in which the filter is one-sided. The recursion begins in January 1995 at $(F^k_{1|0}, P^k_{1|0}) = (0, 1)$. The state starts at zero because every block measurement is standardized. The prior variance equals the measurement variance, so the initial belief carries the same weight as a single observation would.

In every month, I form a prediction. Taking the conditional mean and variance of the state equation gives

$$F^k_{t|t-1} = \phi_k F^k_{t-1|t-1}, \qquad P^k_{t|t-1} = \phi_k^2 P^k_{t-1|t-1} + q_k.$$

The state decays toward zero at rate ϕ_k. The variance evolves in two offsetting ways: the factor ϕ_k^2 shrinks it, while the state innovation variance $q_k = 0.10$ adds to it. The second effect dominates at the values used here, so an additional month without a release always leaves the block state less precise than the month before.

When M^k_t is observed, I update:

$$K^k_t = \frac{P^k_{t|t-1}}{P^k_{t|t-1} + 1}, \qquad F^k_{t|t} = F^k_{t|t-1} + K^k_t \big(M^k_t - F^k_{t|t-1}\big), \qquad P^k_{t|t} = (1 - K^k_t) P^k_{t|t-1}.$$

The Kalman gain K^k_t is the weight placed on the measurement innovation, $M^k_t - F^k_{t|t-1}$. When the prior variance $P^k_{t|t-1}$ is large relative to measurement variance $R_k = 1$, the update moves farther toward the new measurement. When the prior variance is small, the update moves less. Incorporating the observation also reduces the filtered variance.

When no measurement arrives, the update is skipped and $(F^k_{t|t}, P^k_{t|t}) = (F^k_{t|t-1}, P^k_{t|t-1})$. This step is what reconciles the annual and monthly release calendars. Between two Trustees Reports the trust-fund state is carried forward unchanged while its variance accumulates at q_k per month, so by the time the next report appears the filter has become substantially more willing to move. At the parameters used here, an annually released block moves about 0.65 of the way toward a new reading, while a monthly block moves about 0.26 of the way. A once-a-year Trustees Report therefore has a large effect at release, and a single monthly print has a small one, without either frequency being rescaled by hand. This is what makes the annual and monthly blocks comparable within one index.

The persistence parameters carry the substantive assumption. I set $\phi_k = 1$ for trust-fund level and official uncertainty, so the construction holds those states fixed between Trustees Report releases. I set $\phi_k = 0.98$ for macro stress and EPU, which implies a half-life of about 34 months. A labor-market or policy-uncertainty spike therefore decays in the index, while a published revision to program finances remains until the next report.

Let w_k denote the headline weight of block k,

$$w_{TF} = 0.30, \qquad w_{MS} = 0.20, \qquad w_{OU} = 0.10, \qquad w_{EPU} = 0.40,$$

which are fixed, nonnegative, and sum to one. I standardize each filtered block by its January 1995–December 2011 mean and population standard deviation, yielding Z_t^k. The monthly state used for the HRS mapping is the weighted sum

$$G_t = \sum_k w_k Z_t^k.$$

Conditional on the staged input history, later observations do not revise an earlier G_t.

The figures rescale G_t to a readable level; the HRS mapping uses the unrescaled state. All display moments are taken over January 1995–December 2019, a benchmark window that excludes COVID-19. Write μ_G and σ_G for the mean and population standard deviation of G_t over that window, and μ_{Z^k} for the mean of Z_t^k over the same window. The headline index and the contribution of block k are

$$I_t^H = 100 + 10\,\frac{G_t - \mu_G}{\sigma_G}, \qquad C_t^k = 10\,w_k\,\frac{Z_t^k - \mu_{Z^k}}{\sigma_G},$$

so that the contributions add up to the index exactly, $I_t^H - 100 = \sum_k C_t^k$.

The decomposition figure groups the four contributions into a fiscal and an uncertainty index,

$$I_t^F = 100 + \frac{C_t^{TF} + C_t^{MS}}{0.70}, \qquad I_t^U = 100 + \frac{C_t^{OU} + C_t^{EPU}}{0.30},$$

which gives $I_t^H = 0.70 I_t^F + 0.30 I_t^U$ exactly. The constants 0.70 and 0.30 scale the two displayed indexes and do not replace the block weights w_k; any pair summing to one reproduces I_t^H. Before smoothing, all three indexes have a January 1995–December 2019 mean of 100. Both index figures display a one-sided 12-month trailing mean, which neither the HRS mapping nor the structural model uses. The annual state G_y is the calendar-year mean of the unsmoothed monthly OASI values of G_t.

B.4. Mapping the index to HRS beliefs

The HRS Internet Surveys provide the probability scale. I use the general Social Security expectations item in 2006, 2007, and 2011. It asks respondents, thinking about Social Security in general rather than only their own benefits, for the percent chance that Congress will change the program within the next 10 years so that it becomes less generous. The wording is essentially constant across waves; in 2011 the item appears in Collection A of a randomized wording experiment. Valid answers lie between zero and 100 and are divided by 100.

The response is a proxy for belief about a future policy-induced reduction in generosity, not an objective forecast of the statutory depletion date. The Internet Surveys also contain personalized questions. In 2006 and 2007, current recipients report the chance that current benefits will be cut, while nonrecipients report the chance that changes will reduce future benefits. The 2011 personalized follow-up is conditional on a positive answer to the general question and, for nonrecipients, a positive chance of ever receiving Social Security; it is restricted to respondents younger than 70. The 2011 survey also fields an alternative general "benefits will be cut" wording. These personalized and alternative-wording items are diagnostic alternatives and do not enter the reported calibration.

Let G_y be the annual state and define

$$h_y = \Lambda(\alpha_0 + \alpha_1 G_y), \qquad \Lambda(z) = \frac{\exp(z)}{1 + \exp(z)}.$$

Because the survey asks about a 10-year horizon, the fitted survey counterpart is

$$P_y^{10}(\alpha_0, \alpha_1) = 1 - \left[1 - \Lambda(\alpha_0 + \alpha_1 G_y)\right]^{10}.$$

The public replication archive contains wave-level weighted means and observation counts, but not respondent microdata. I fit the two mapping parameters by weighted nonlinear least squares, using survey-wave observation shares as weights:

$$(\widehat{\alpha}_0, \widehat{\alpha}_1) = \underset{\substack{-10 \le \alpha_0 \le 2, \\ 0 \le \alpha_1 \le 10}}{\arg\min} \sum_y \frac{N_y}{\sum_s N_s} \left[\bar{p}_y^{HRS} - P_y^{10}(\alpha_0, \alpha_1) \right]^2 .$$

The objective uses only annual index states and weighted survey means for 2006, 2007, and 2011; no post-2011 index row enters it.

TABLE B.2. HRS moments used to estimate the subjective-probability mapping

Survey year	G_y	Observations	HRS 10-year risk	Model 10-year risk	Model annual h_y
2006	-0.692	1,208	63.3%	61.7%	9.1%
2007	-0.704	1,487	60.3%	61.6%	9.1%
2011	1.552	2,180	72.1%	72.1%	12.0%

Note. The HRS entries are archived weighted means of the standard 10-year "less generous" question. The archive does not contain the respondent records needed to reconstruct the weights independently. Model 10-year risk is $1 - (1 - h_y)^{10}$.

Source. Author's calculations from the HRS Internet Survey and the annual OASI financial-risk state.

The estimates are

$$\widehat{\alpha}_0 = -2.203, \qquad \widehat{\alpha}_1 = 0.135.$$

At $G_y = 0$, they imply an annual hazard of 9.9 percent and a 10-year risk of 64.9 percent. Identification is limited: two parameters are fit to three aggregate wave means. Leaving out 2006 or 2007 yields slopes of 0.151 and 0.114, respectively. Leaving out 2011 leaves two almost identical G_y values and pushes the numerical solution toward the parameter bounds ($\widehat{\alpha}_0 = 2.000$ and $\widehat{\alpha}_1 = 6.148$). I therefore use leave-one-wave-out results as a sensitivity diagnostic and do not report respondent-bootstrap standard errors from data that are not in the archive. The partial 2026 state is $G_y = 2.917$ and implies an annual hazard of 14.1 percent.

B.5. Mapping calendar hazards into the model

The estimated annual hazard h_y supplies the subjective onset beliefs in the main structural counterfactuals. The observed series covers 1995–2026. Calendar years before 1995 receive zero hazard, and years after 2026 use the last available annual hazard, 0.140643. There are no missing interior years. For cohort b, the annual pre-depletion transition at age a is $\lambda_b(a) = h_{b+a}$. The input is already a conditional annual probability; it is not divided by a survival probability again.

Conditional on no onset before calendar year y_0, the implied probability of onset by year y is

$$1 - \prod_{v=y_0}^{y} (1 - h_v).$$

The dynamic program uses the annual hazards rather than this cumulative probability. Each cohort knows its entire profile from model entry, including the specified extrapolation. This is an information assumption for the counterfactual, not evidence of historical household foresight.

Full-payment restoration, when included, is governed by the distinct restoration hazard q_d.

Appendix C. Social Security and Tax Rules

This appendix records the fixed Social Security and tax rules retained from Online Appendix C of Fan, Seshadri, and Taber (2024). Dollar amounts are in 2004 dollars. The benchmark combines the 2004 federal income and payroll tax schedules with the 1999 Retirement Earnings Test (RET). These historical rules are model primitives: the payment-shortfall experiment changes only the fraction of each scheduled OASI retirement payment that is actually paid.

C.1. Taxes and after-tax resources

The payroll tax consists of a 6.2-percent Social Security tax on earnings up to \$87,900 and an uncapped 1.45-percent Medicare tax. State income taxes are omitted. Federal taxes use head-of-household filing status, a \$3,100 personal exemption, and a \$7,150 standard deduction. Table C.1 gives the piecewise-linear map retained from FST.

TABLE C.1. After-tax income schedule in 2004 dollars

Marginal tax rate	Pre-tax income Y	After-tax income $\mathcal{T}(Y)$
0.0765	$Y \leq 10{,}250$	$0.9235Y$
0.1765	$10{,}250 < Y \leq 20{,}450$	$9{,}465.88 + 0.8235(Y - 10{,}250)$
0.2265	$20{,}450 < Y \leq 49{,}150$	$17{,}865.58 + 0.7735(Y - 20{,}450)$
0.3265	$49{,}150 < Y \leq 87{,}900$	$40{,}065.03 + 0.6735(Y - 49{,}150)$
0.2645	$87{,}900 < Y \leq 110{,}750$	$66{,}163.15 + 0.7355(Y - 87{,}900)$
0.2945	$110{,}750 < Y \leq 172{,}950$	$82{,}969.33 + 0.7055(Y - 110{,}750)$
0.3445	$172{,}950 < Y \leq 329{,}350$	$126{,}851.43 + 0.6555(Y - 172{,}950)$
0.3645	$Y > 329{,}350$	$229{,}371.63 + 0.6355(Y - 329{,}350)$

Note. The displayed marginal rate combines federal income, Social Security payroll, and Medicare taxes. Above the Social Security taxable maximum, the 6.2-percent component drops out, producing the lower fifth-row rate.

C.2. Earnings records, claiming, and scheduled benefits

Scheduled benefits depend on the earnings record and the claiming decision. These rules create a second connection between current employment and future retirement resources. I retain the historical FST rules across all cohorts; the birth-cohort labels determine exposure to payment risk rather than cohort-specific statutory retirement ages.

C.2.1. Claiming and the earnings record

Let s_t indicate whether benefits were initiated before year t, and d_t indicate initiation in the current year. Claiming is available between ages 62 and 70 and follows

$$s_{t+1} = \max\{s_t, d_t\}, \qquad d_t \in \begin{cases} \{0\}, & s_t = 1 \text{ or } a_t < 62, \\ \{0,1\}, & s_t = 0 \text{ and } 62 \le a_t < 70, \\ \{1\}, & s_t = 0 \text{ and } a_t = 70. \end{cases} \tag{C.1}$$

Claiming and employment are separate decisions, so a household may receive benefits while working. A single record state R_t stores the earnings average before claiming and the annual entitlement afterward:

$$R_t = \begin{cases} \text{AIME}_t, & s_t = 0, \\ \bar{b}_t, & s_t = 1. \end{cases} \tag{C.2}$$

Both quantities use a common grid. Before claiming, the record approximates Average Indexed Monthly Earnings (AIME), the monthly average of the 35 highest years of covered earnings. Following Fan, Seshadri, and Taber (2024), it updates according to

$$\text{AIME}_{t+1} = \text{AIME}_t + \max\left\{0, \frac{\min\{y_t, 87{,}900\}}{35 \times 12} - m(a_t)\,\text{AIME}_t\right\}, \tag{C.3}$$

where $m(a) = \mathbf{1}\{a - \underline{a} \ge 35\}\,\text{share}_{\min}(a)$ and $\text{share}_{\min}(a)$ is the SIPP-based estimate of the share attributable to the lowest earnings year. Covered earnings are capped at \$87,900 in 2004 dollars. The approximation allows sufficiently high earnings to replace an earlier low year and prevents a low-earnings year from reducing the record. Real earnings enter directly, without a separate wage-index adjustment.

C.2.2. Scheduled benefits and the earnings test

The monthly Primary Insurance Amount applies the progressive formula

$$\begin{aligned} \text{PIA}(R) = {} & 0.90\min\{R, 612\} \\ & + 0.32\min\{\max(R - 612, 0), 3{,}689 - 612\} \\ & + 0.15\max\{R - 3{,}689, 0\}, \end{aligned} \tag{C.4}$$

with bend points in 2004 dollars. At claiming, AIME is converted into an annual entitlement using

$$\bar{b}_t = 12 \cdot g(a_t) \cdot \mathrm{PIA}(R_t), \quad g(a) = \begin{cases} 1 - 0.0667(65-a), & 62 \le a < 65, \\ 1, & a = 65, \\ 1 + 0.06 \min\{a - 65, 4\}, & 65 < a < 70, \\ 1.24, & a \ge 70. \end{cases} \tag{C.5}$$

The inherited normal retirement age is 65 and the earliest claiming age is 62. Early claiming reduces benefits by approximately 6.67 percent per year; delayed claiming earns a 6-percent annual credit through age 69. These rules are held fixed when comparing birth cohorts.

A claimant working before age 70 faces the Retirement Earnings Test under the 1999 rules expressed in 2004 dollars. Benefits withheld are

$$\mathrm{RET}_t = \begin{cases} \min\{\bar{b}_t,\ \max\{0, (y_t - 10{,}885)/2\}\}, & 62 \le a_t < 65, \\ \min\{\bar{b}_t,\ \max\{0, (y_t - 17{,}575)/3\}\}, & 65 \le a_t < 70, \\ 0, & \text{otherwise}, \end{cases} \tag{C.6}$$

Withholding produces a future-entitlement credit:

$$\bar{b}_{t+1} = \bar{b}_t \left[1 + \frac{\mathrm{RET}_t}{\bar{b}_t} \Big(0.067 \cdot \mathbf{1}\{62 \le a_t < 65\} + 0.06 \cdot \mathbf{1}\{65 \le a_t < 70\} \Big) \right], \tag{C.7}$$

The ratio is set to zero when $\bar{b}_t = 0$. Covered earnings no longer update AIME after claiming. The scheduled current payment is $b_t^{pre} = \max\{0, \bar{b}_t - \mathrm{RET}_t\}$, and the regime factor in equation (2) determines the payment actually received.

The earnings test exchanges current payments for future entitlements. Its incentive effect therefore depends on how households value those future payments (Friedberg 2000; Gruber and Orszag 2003). Perceived shortfall risk changes that valuation without changing the statutory credit. Similarly, claiming earlier cannot protect benefits from the common payment factor. Claiming responses reflect liquidity, consumption smoothing, and the expected return to delayed receipt. The experiment covers OASI retired-worker benefits; disability benefits remain inactive.

C.3. Taxable Social Security benefits

The earnings test and its entitlement credit are applied before the payment fraction, and benefit taxation is computed from the payment actually received. Let $Y_t^o = \max\{rA_t, 0\} + y_t + y_t^s$ denote other gross income.

Benefit taxation depends on actual OASI payments. Provisional income is

$$\widetilde{Y}_t = Y_t^o + \tfrac{1}{2} ssb_t. \tag{C.8}$$

The tentative taxable amount and the statutory cap give

$$T_t^{ss} = \tfrac{1}{2}\min\{ssb_t,\ \widetilde{Y}_t - 25{,}000,\ 9{,}000\} + 0.85\max\{0,\ \widetilde{Y}_t - 34{,}000\}. \tag{C.9}$$

$$ssb_t^{tax} = \begin{cases} 0, & \widetilde{Y}_t \leq 25{,}000, \\ \min\{0.85\, ssb_t,\ T_t^{ss}\}, & \widetilde{Y}_t > 25{,}000. \end{cases} \tag{C.10}$$

Benefits are untaxed below the lower threshold, and at most 85 percent of the payment is taxable. The after-tax-income function in Appendix C.1 is applied to other gross income plus taxable Social Security benefits. The untaxed portion of the actual payment is then added to obtain disposable income in equation (12). Thus, the payment factor affects both gross benefits and their tax treatment.

C.4. Understanding the fixed-behavior examples

These examples unpack the fixed-behavior benchmark in Figure 7. They hold behavior fixed and ask how different actual-payment paths change the value of scheduled OASI payments. The goal is not to measure life-cycle welfare. Instead, the examples show why the expected accounting loss and the uncertainty discount in equation (31) measure different objects.

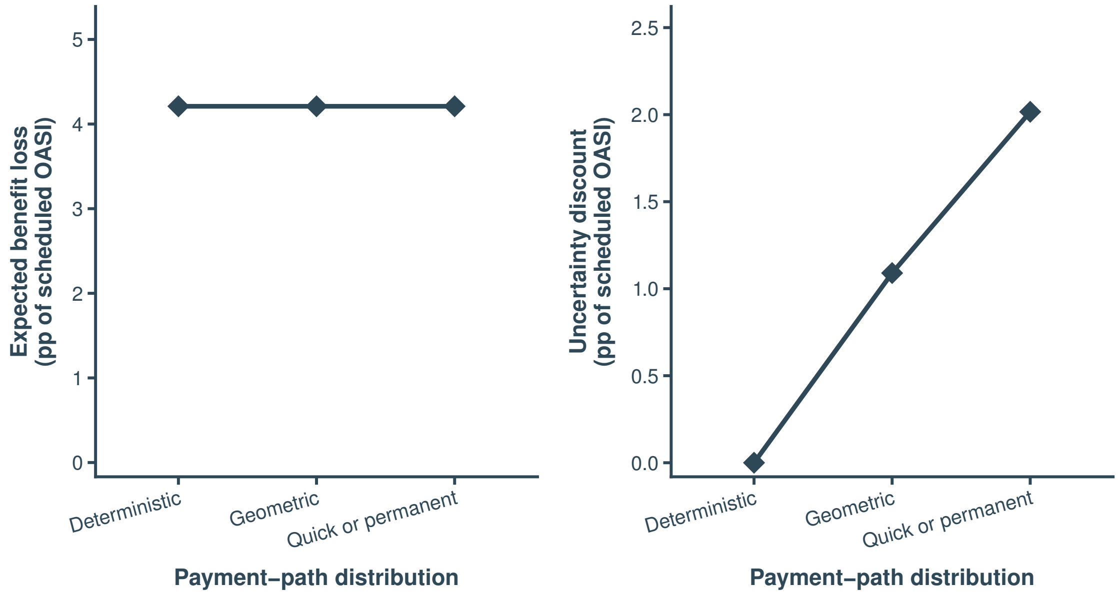

FIGURE C.1. Equal expected accounting loss does not imply equal duration risk

Note. All three policies are mean-matched for the 1982 birth cohort claiming at age 67. The calibrated geometric hazard and the quick-restoration-or-permanent mixing probability are then held fixed. Panel A verifies the equal expected loss at the reference household. Panel B reports the uncertainty discount. Values are percentage points of normalized scheduled OASI wealth.

The same average loss can imply different uncertainty discounts. Figure C.1 considers the 1982 cohort and three full-payment-restoration processes. Each process has the same expected accounting loss: 4.2 percentage points of scheduled OASI wealth. In the first, the restoration date is known, so there is no uncertainty discount. In the second, restoration can occur in any year, so the discount is 1.1 percentage points. In the third, full payments either return quickly or never return, so the discount is 2.0 percentage points. The average loss is the same in all three cases. What changes is uncertainty about full-payment restoration.

Separating onset and duration uncertainty under index-mapped beliefs. Figure C.2 uses the same index-mapped annual onset hazards as the structural model, conditional here on no depletion through 2026. The annual hazard is held at 0.140643 after 2026. Full-payment restoration has a 5-percent annual hazard, giving an expected reduced-payment duration of twenty years. The onset distribution is enumerated through the last possible annuity-payment year, 2115. All later-onset probability is assigned the no-loss payoff, so no probability is discarded or renormalized.

The joint uncertainty discount ranges from 0.625 to 1.331 percentage points of scheduled

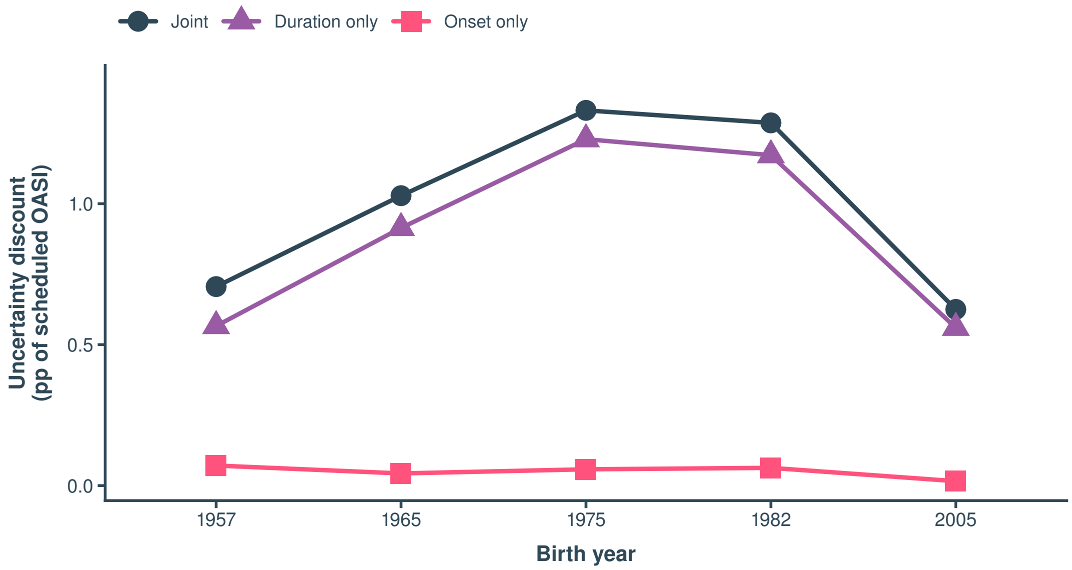

FIGURE C.2. Onset and duration uncertainty under index-mapped subjective hazards

Note. The onset distribution follows HRS-disciplined, index-mapped hazards conditional on solvency through 2026; the last observed annual hazard is held fixed afterward. Restoration has annual probability 0.05. Lines report joint, duration-only, and onset-only uncertainty discounts in percentage points of scheduled OASI wealth. Removing a source of risk replaces its payoff variation by the corresponding conditional expectation. Behavior is fixed and all payoff-relevant histories are enumerated.

OASI wealth across cohorts. The figure also shows onset-only and duration-only uncertainty. Removing one source means replacing payoffs by their conditional expectation over that source, preserving the expected benefit ratio. A symmetric two-factor allocation assigns duration uncertainty 85.1, 92.3, 94.0, 93.1, and 93.5 percent of the joint discount for the 1957, 1965, 1975, 1982, and 2005 cohorts, respectively. Duration is the larger component in each case, but onset uncertainty is no longer negligible.

Payment depth and duration determine which cohorts are exposed. Figure C.3 compares a deep but temporary payment shortfall with a smaller permanent payment shortfall. The two policies are designed to give the 1982 cohort the same expected loss, but they do not affect other cohorts in the same way. The deeper temporary shortfall is concentrated on people who receive OASI payments during 2034–2053, while the permanent shortfall reaches younger cohorts as well. Thus, matching the average loss for one cohort does not equalize losses across generations. This is the life-cycle reason that duration matters.

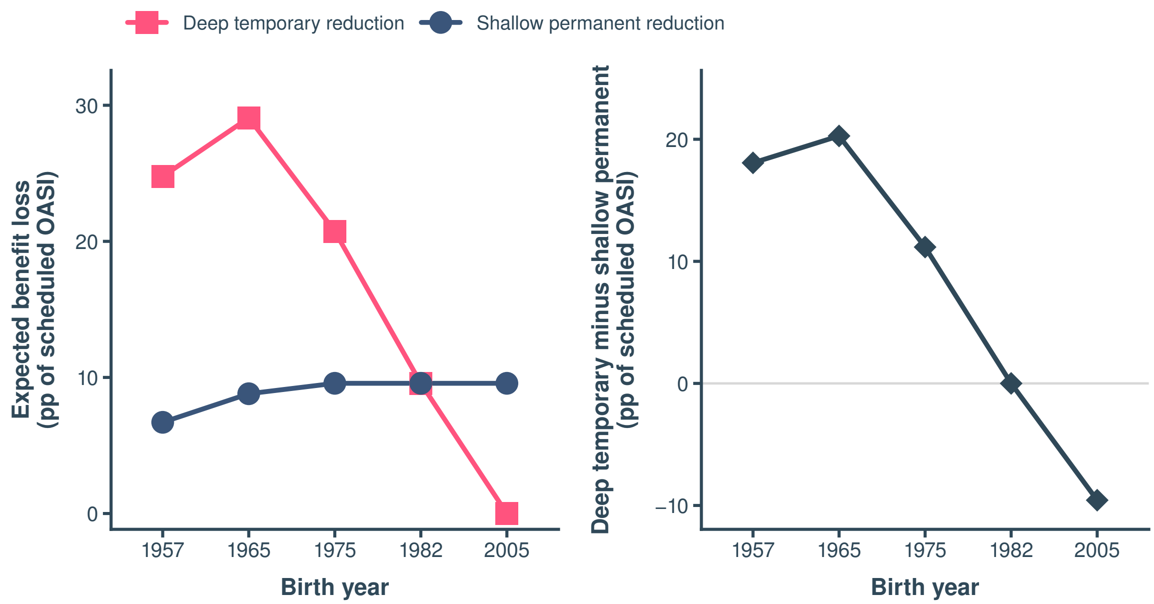

FIGURE C.3. Severity and duration redistribute the same reference loss

Note. The deep-temporary policy pays 50 percent of the scheduled amount from 2034 through 2053. The shallow-permanent policy pays 90.432 percent from 2034 onward. The policies are mean-matched only for the 1982 cohort claiming at age 67. Both policies are deterministic and therefore have zero uncertainty discount conditional on onset.

Appendix D. Model Details and Computation

This appendix completes the description of the life-cycle model in Section 2. It contains two parts. First, I derive the closed-form integration of the leisure-taste shock stated in Equation (20) and its limiting case. Second, I write out the expectation operator over next-period shocks in the Bellman equation (17) and identify the cases in which its integrals degenerate. The economic environment is that of Fan, Seshadri, and Taber (2024); labor supply is binary throughout, so part-time work is outside the model.

D.1. Analytic integration of the leisure-taste shock

Section 2 defines the conditional branch values $W_t^W(\widetilde{X}_t)$ and $W_t^N(\widetilde{X}_t)$, each optimized over consumption, claiming, and (for the working branch) investment, and each excluding the current leisure payoff $\exp(\mu_{it} + \varepsilon_{it})$. When $W_t^W > W_t^N$, the household works if and only if $\varepsilon_{it} < \varepsilon_t^*$, with the cutoff given by Equation (18), and the work probability is $\Phi(q_t)$ with $q_t = \varepsilon_t^*/\sigma_\varepsilon$ as in Equation (19).

The value of the state $\widetilde{X}_t$ before the shock is realized is the expectation of the maximum over the two branches. Splitting the expectation at the cutoff,

$$E[V_t \mid \widetilde{X}_t] = \Pr(\varepsilon < \varepsilon_t^*)\, W_t^W + \Pr(\varepsilon \geq \varepsilon_t^*)\, W_t^N + E\big[e^{\mu_{it}+\varepsilon_{it}} \mathbf{1}\{\varepsilon_{it} \geq \varepsilon_t^*\}\big]\,.$$

The last term is the mean of a log-normal random variable truncated from below. Completing the square in the normal density gives the standard result

$$E\big[e^{\mu_{it}+\varepsilon_{it}} \mathbf{1}\{\varepsilon_{it} \geq \varepsilon_t^*\}\big] = \exp\left(\mu_{it} + \frac{\sigma_\varepsilon^2}{2}\right) \Phi(\sigma_\varepsilon - q_t)\,,$$

so that the pre-shock expected value has the closed form

$$E[V_t \mid \widetilde{X}_t] = \Phi(q_t)\, W_t^W + [1 - \Phi(q_t)]\, W_t^N + \exp\left(\mu_{it} + \frac{\sigma_\varepsilon^2}{2}\right) \Phi(\sigma_\varepsilon - q_t)\,. \tag{D.1}$$

This is the operator $\mathcal{E}_t$ in Equation (20). When $W_t^W \leq W_t^N$, the limiting expression as $q_t \to -\infty$ applies: the household never works and

$$E[V_t \mid \widetilde{X}_t] = W_t^N + \exp\left(\mu_{it} + \frac{\sigma_\varepsilon^2}{2}\right), \tag{D.2}$$

which is the nonwork value plus the unconditional mean of the log-normal taste shifter.

Two computational implications follow. First, because the shock is integrated exactly, no quadrature over ε_{it} is required and ε_{it} never appears in the state vector; the two conditional values W_t^W and W_t^N are sufficient statistics for the work margin. Second, Equation (D.1) combines the two conditional branch values analytically. The payment-regime extension changes the continuation values inside W_t^W and W_t^N but leaves this integration step unchanged.

D.2. Expectation operator in the Bellman equation

Conditional on a work branch $p \in \{0, 1\}$ and the associated choices of consumption, investment, and claiming, the continuation value in Equation (17) is the expectation over three sources of next-period uncertainty: the family transition, next-period spousal earnings, and the human-capital innovation. Writing $F_{\zeta,t+1}$ for the spousal earnings distribution and F_ξ for the distribution of the Ben–Porath innovation ξ^H in Equation (11), the expectation operator applied to the state $\boldsymbol{X}_t$ of Equation (16) is

$$\mathbb{E}_t\left[\overline{V}_{t+1}\right] = \sum_{M'} P_t^M(M_t, M') \iint \overline{V}_{t+1}(A', H', R', s', M')\, dF_{\zeta,t+1}\, dF_{\xi}, \tag{D.3}$$

where $\overline{V}_{t+1}$ is the regime-mixed continuation value of Equation (22). Two of these integrals collapse in identifiable cases. The spousal-income integral is degenerate at zero income unless the next-period family state is married-working ($M' = 2$); when $M' = 2$, spousal log earnings are log-normal with age-specific mean and variance and the integral is evaluated with two Gauss–Hermite nodes. The human-capital integral is degenerate when the household does not work: with $p_t = 0$, Equation (11) reduces to deterministic depreciation $H_{t+1} = (1-\delta)H_t$, and the innovation is irrelevant. The idiosyncratic expectation in equation (D.3) contains no survival, medical-expenditure, job-offer, previous-employment, or separate wage shock. The current payment regime also determines current OASI payments and resources. The regime transition and belief group enter through $\overline{V}_{t+1}$ and add no dimension to the shocks integrated in (D.3).

D.3. Numerical implementation

I solve the model by backward induction from age 80, when the continuation value is the bequest function in equation (15). Assets, human capital, and the Social Security record are discretized on age-varying grids. Parameter search uses $(N_A, N_H, N_R) = (18, 16, 11)$; the reported counterfactuals use $(31, 16, 21)$. The record grid represents AIME before claiming and the claiming-age-adjusted entitlement after claiming, consistent with equation (C.2).

At each state and age, I solve the work and nonwork branches separately and compare the admissible claiming choices. Conditional on the discrete choices, I optimize consumption and, in the work branch, human-capital investment with a Nelder–Mead simplex routine. I evaluate the consumption-floor boundary separately to retain corner solutions.

I integrate the human-capital innovation with five-node Gauss–Hermite quadrature, spousal log earnings with two-node Gauss–Hermite quadrature, and family status with a finite Markov sum. I use tetrahedral interpolation for continuation states between grid points. The analytically integrated leisure shock combines the work and nonwork branch values using equation (D.1).

Within every branch, I update AIME before claiming, convert the record to the claiming-adjusted entitlement at initiation, apply the earnings test and its future-entitlement credit, and then apply the payment factor $\chi(z_t)$. I calculate taxable benefits from the payment actually received and apply the after-tax-income map last. This order ensures that the reduced-payment state changes actual OASI payments but not AIME, the claiming adjustment, or the earnings-test credit. The adjusted entitlement is interpolated on the common record grid after claiming. All environments use the same grid and numerical settings.

Appendix E. Additional life-cycle results

This appendix provides complete life-cycle profiles for the permanent reduced-payment endpoint. The main text connects expected reduced-payment duration to mechanical benefit exposure, welfare incidence, and anticipatory adjustment. The material below reports the underlying cohort profiles and the complete eight-scenario duration comparisons accompanying the main-text figures.

E.1. Surprise-shortfall adjustment is age dependent

The surprise-shortfall effect in equation (25) measures the model-implied consequence of an unexpected decline in actual OASI payments. In Figure E.1, each cohort follows the full-payment benchmark through 2033; all later differences are ex-post behavioral responses. Reduced payments raise the marginal value of resources, but life-cycle position determines which adjustment margins remain available.

For the 1957 cohort, reduced payments begin at age 77, after claiming is complete. Actual OASI payments fall by 1.41 model units at that age, consumption falls by 0.087 units, and assets end 3.67 units below the benchmark at age 80. Claiming cannot move, human-capital adjustment is negligible, and employment rises by at most 0.49 percentage points. The old-age response is therefore almost pure dissaving: households absorb part of the lost transfer by consuming less and running down accumulated assets. They have little ability to self-insure after the surprise, but they also lose payments for relatively few remaining years; the surprise-shortfall welfare cost is 0.23 percent of lifetime consumption.

The 1965 cohort encounters the shortfall near the retirement transition. Reduced payments begin at age 69, when claiming is effectively complete but a substantial retirement horizon remains. Employment rises by 3.89 percentage points at onset. Consumption reaches 0.35 model units below the benchmark at age 71, and assets are 9.65 units lower by age 80. Human capital eventually rises by 0.21 units, but the short horizon limits the return to new investment. This combination—many remaining payment years but few remaining adjustment years—produces the largest point-in-time consumption and asset losses among the three cohorts discussed, even though its lifetime welfare loss, 0.76 percent, is not the largest.

The 1975 cohort enters the reduced-payment state at age 59, before becoming eligible to claim, and therefore retains a broader set of adjustment margins. The share that has claimed by

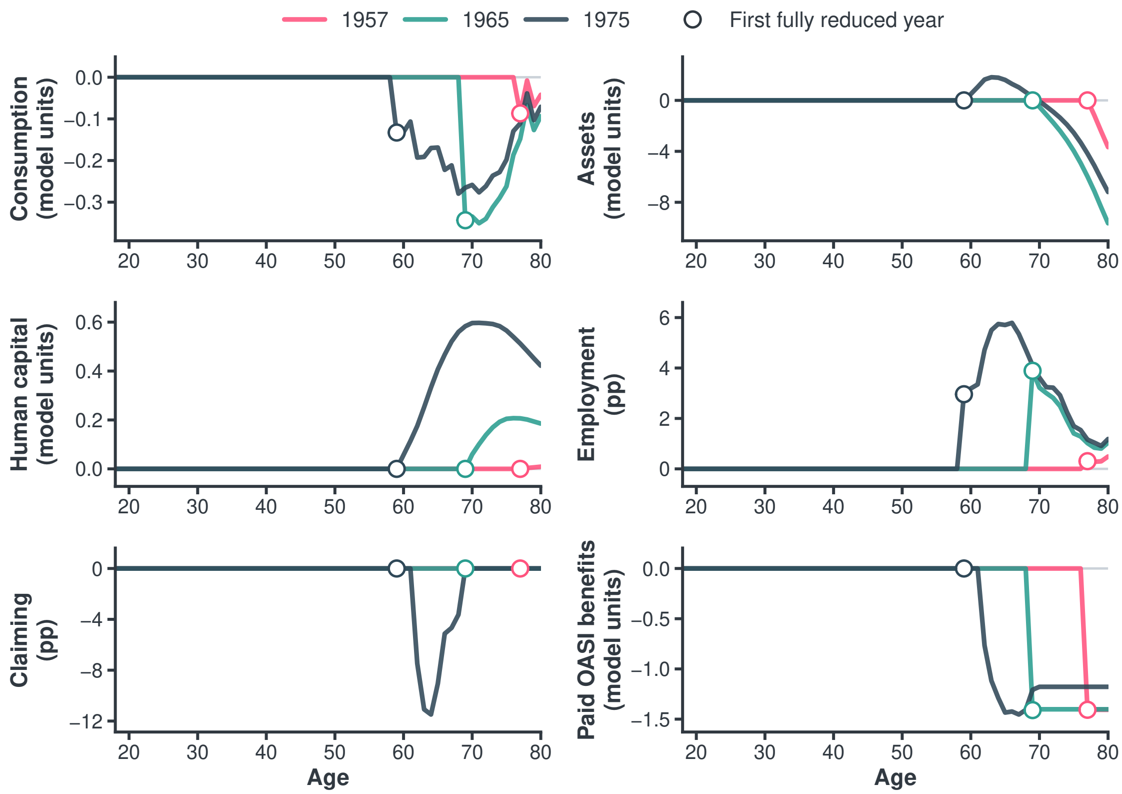

FIGURE E.1. Life-cycle responses to a surprise OASI payment shortfall

Note. Lines show the 1957, 1965, and 1975 birth cohorts, which are ages 76, 68, and 58 in the 2033 depletion year and ages 77, 69, and 59 in the first reduced-payment year. These cohorts encounter reduced payments after, during, and before the claiming window, respectively. The figure plots the surprise-shortfall environment minus the full-payment FST-based benchmark at each age. Employment and claiming differences are percentage points; all other differences are in model units. Open circles mark the first reduced-payment age. Structural parameters, initial conditions, and simulation draws are common across scenarios.

age 64 is 11.49 percentage points lower than in the benchmark. Employment rises by as much as 5.79 percentage points at age 66, and the longer expected work horizon raises the return to training: human capital peaks 0.60 model units above the benchmark at age 71. Assets initially rise by 1.82 units at age 63 and then fall to 7.19 units below the benchmark by age 80 as the accumulated buffer replaces part of the missing payment income. Because the cohort can adjust work, productivity, saving, and claiming jointly, its maximum consumption decline, 0.28 model units, is smaller than that of the 1965 cohort.

Two objects organize these responses. Mechanical benefit exposure records how much of a cohort's scheduled payment stream falls in reduced-payment years. Adjustment capacity records which self-insurance margins remain available when the surprise occurs.

Panel A of Table E.1 reports the surprise-shortfall welfare cost. It rises from 0.23 percent for the 1957 cohort to 0.76 percent for 1965 and 1.16 percent for 1975 as younger cohorts face more remaining years of reduced payments. The largest short-run consumption decline is nevertheless non-monotone in age because households gain more adjustment margins when the event arrives before retirement. The common payment loss therefore propagates differently through retirement, asset accumulation, and human-capital investment at different ages.

E.2. Anticipation changes saving and claiming

Equation (26) measures the anticipation effect by comparing the anticipated-shortfall environment Y^A with the surprise-shortfall environment Y^U under identical objective payment histories. The asset profiles reveal precautionary saving as one adjustment margin.

At the first reduced-payment age, assets in the anticipated environment are 9.14, 9.42, and 6.40 model units higher for the 1957, 1965, and 1975 cohorts. Consumption is higher than under surprise by 0.085, 0.107, and 0.082 units, while employment is lower by 0.45, 1.75, and 1.59 percentage points. These are matched profile differences; isolating the causal contribution of the asset buffer would require constraining other choices. Anticipation reallocates adjustment across ages.

Building these buffers requires earlier adjustment. The largest reductions in consumption relative to surprise are 0.074, 0.072, and 0.071 model units for the 1957, 1965, and 1975 cohorts, respectively, at ages 23, 22, and 23. These early responses are possible because households know their full hazard profiles from model entry, even before the observed hazard series begins. At the first reduced-payment age, the 1975 cohort has 0.095 model units more human capital than under surprise. The corresponding differences are -0.013 for 1957 and 0.044 for 1965. Human-capital responses therefore also vary with the cohort's risk history and remaining working life.

Claiming also shifts resources across time. For the 1965 cohort, the share that has claimed by age 63 is 8.92 percentage points higher under anticipation, although cumulative claiming is equal across environments by age 69. Early claiming cannot preserve the full-payment rate because ϕ applies to old and new claimants alike. It instead moves liquidity into pre-depletion years and reduces exposure to the lower expected return on payments delayed farther into the future.

Figure E.3 shows this reallocation across claiming ages. Positive values indicate that antici-

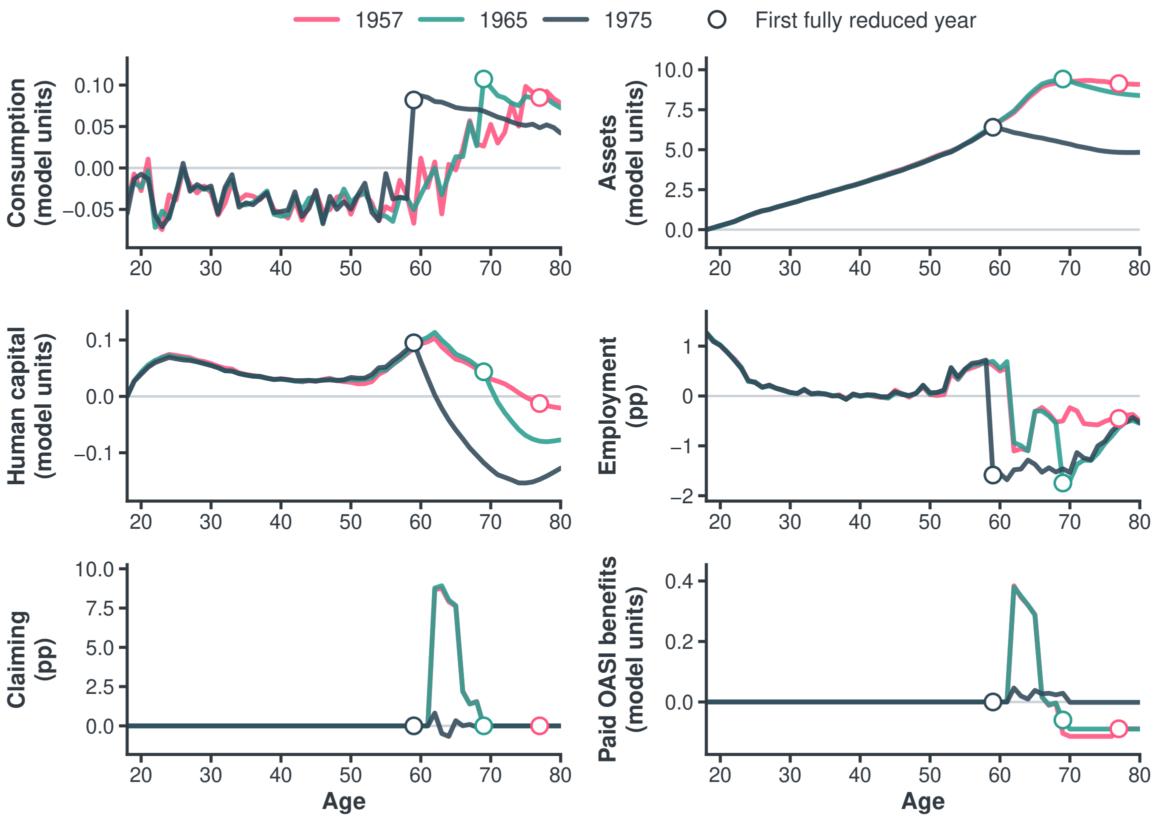

FIGURE E.2. Life-cycle adjustment under anticipation

Note. Lines show the 1957, 1965, and 1975 birth cohorts, which are ages 76, 68, and 58 in the 2033 depletion year and ages 77, 69, and 59 in the first reduced-payment year. These cohorts encounter reduced payments after, during, and before the claiming window. The figure plots the anticipated-shortfall environment minus the otherwise identical surprise-shortfall environment at each age, thereby isolating the anticipation effect. Employment and claiming differences are percentage points; all other differences are in model units.

pation moves applications into an age, while negative values indicate that it moves applications away from that age. For the 1965 cohort, the increase at the earliest claiming ages is followed by fewer applications at later ages. The figure therefore shows a change in when households claim, not a lasting increase in the number of claimants.

More broadly, anticipation changes the timing of adjustment rather than eliminating the underlying loss. Households may save more, reallocate consumption across ages, work more before reduced payments begin and less when it occurs, claim benefits earlier, and—when they have a long enough horizon—invest in skills that raise future earnings. These actions can reduce the immediate drop in consumption and employment when payments fall. However, they also require earlier sacrifices in consumption or leisure, so smoother adjustment does not necessarily translate into higher welfare over the full life-cycle.

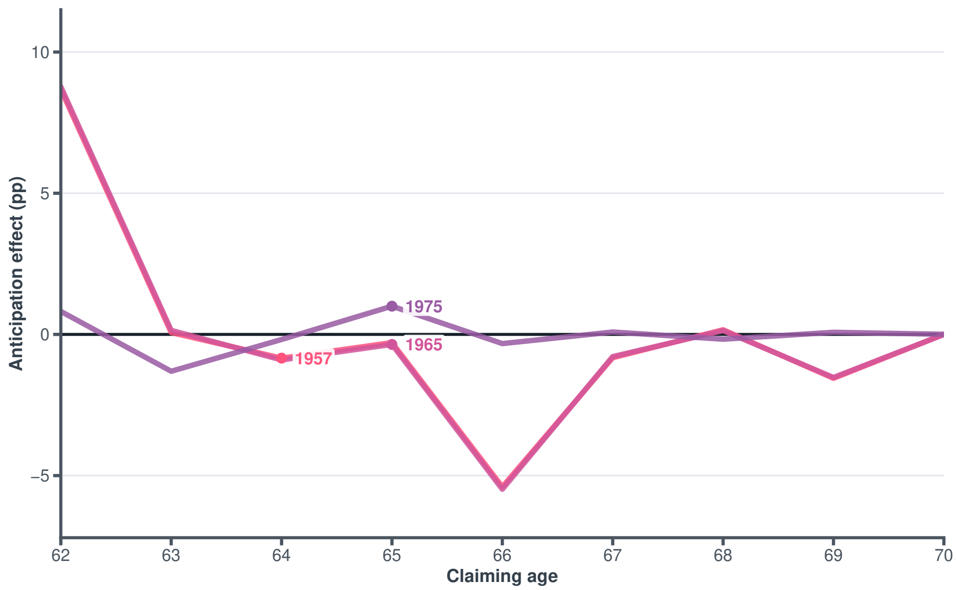

FIGURE E.3. Anticipation changes the timing of Social Security claiming

Note. Lines report the difference in application rates between the anticipated-shortfall and surprise-shortfall environments at each claiming age, in percentage points. Both scenarios face the same realized 2033 reserve-depletion event and permanent reduced payments; they differ in whether households anticipate before the event. The 1957, 1965, and 1975 birth cohorts experience the reduced payments after, during, and before the Social Security claiming window, respectively. Positive values indicate that anticipation moves applications into an age, while negative values indicate that it moves applications away from that age.

E.3. Welfare at the permanent reduced-payment boundary

Sections E.1 and E.2 report the surprise-shortfall and anticipation comparisons. The total CEV is computed directly from the anticipated-shortfall versus full-payment comparison. It is not obtained by adding the two component CEVs, because each CEV applies a nonlinear transformation to a different target–reference utility gap.

Figure E.4 plots the total effect. Total welfare costs in the index-mapped simulation are 0.36, 0.89, 1.26, 1.26, and 1.26 percent for the 1957, 1965, 1975, 1982, and 2005 cohorts. The costs rise across the first three cohorts and then flatten. For the three youngest cohorts, reserve depletion precedes claiming eligibility, so nearly the entire future scheduled OASI payment stream is exposed to permanently reduced payments.

The three matched comparisons clarify why total costs flatten for the three youngest cohorts. The surprise-shortfall cost rises across them, while the anticipation comparison becomes more favorable for the 2005 cohort, which has many remaining working years after the payment reduction.

Figure E.5 summarizes the employment and welfare responses across all five cohorts. The employment response changes sharply with life-cycle position. The two oldest cohorts make relatively small changes in total employment, whereas the 1975, 1982, and 2005 cohorts work 0.590, 0.609, and 0.612 additional years over the life-cycle, respectively. These younger cohorts have more time to respond because reserve depletion occurs well before they become eligible to claim benefits. Their welfare costs are also similar, at about 1.26 percent of lifetime consumption because nearly their entire future scheduled OASI payment stream is exposed to the payment shortfall.

Figure E.4 and Table E.1 show that life-cycle position, not distance in birth years, organizes welfare incidence. The 1965 cohort is 68 at depletion, while the 1975 cohort is 58 and not yet eligible to claim. Their total welfare costs are 0.887 and 1.259 percent. The 1975 and 2005 cohorts are thirty years apart but have similar costs of 1.259 and 1.263 percent. Once depletion precedes claiming eligibility, nearly the entire scheduled payment stream is exposed. A longer adjustment horizon still changes saving, employment, human-capital investment, and claiming, but these changes do not monotonically reduce welfare costs. Differences in birth year need not translate into proportional welfare differences.

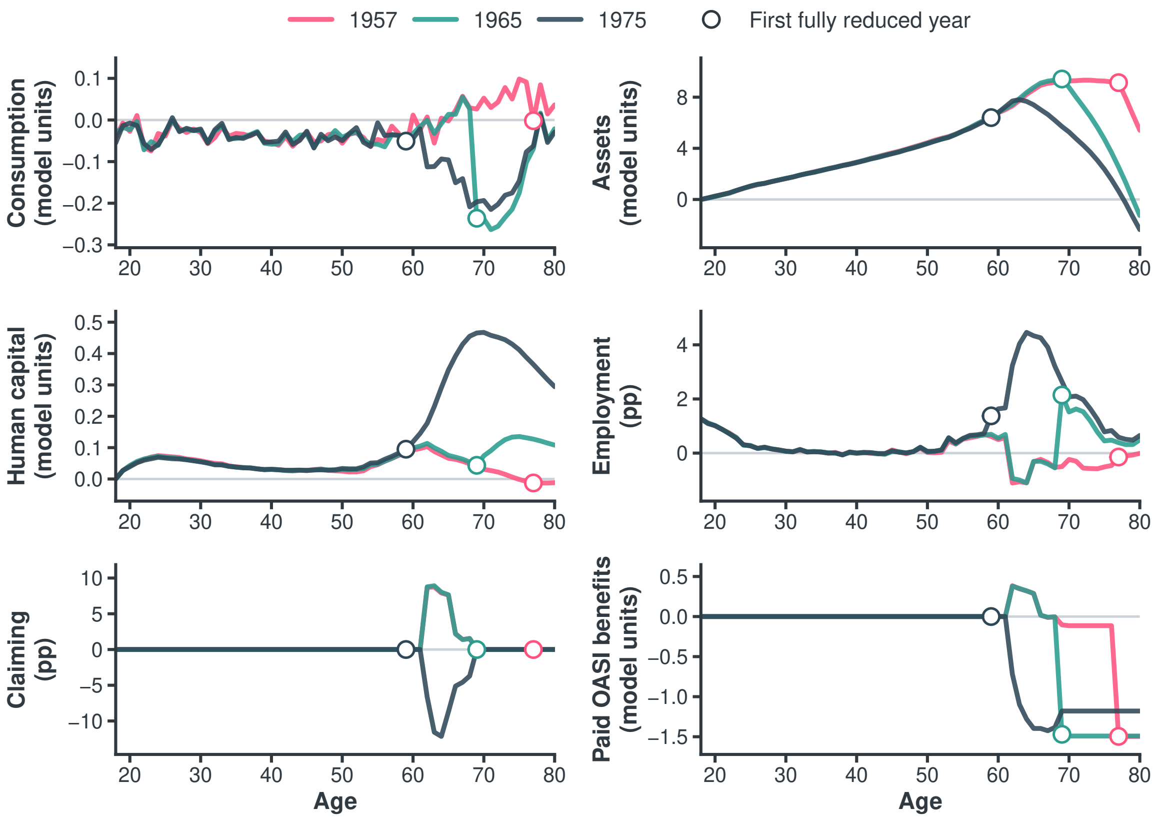

FIGURE E.4. Total life-cycle effect of the anticipated payment shortfall

Note. Lines show the 1957, 1965, and 1975 birth cohorts, which are ages 76, 68, and 58 in the 2033 depletion year and ages 77, 69, and 59 in the first reduced-payment year. These cohorts encounter reduced payments after, during, and before the claiming window, respectively. The figure plots the anticipated-shortfall environment minus the full-payment FST-based benchmark at each age. Employment and claiming differences are percentage points; all other differences are in model units. Structural parameters, initial conditions, and simulation draws are common across scenarios. Flow-outcome differences telescope across the three environments; each welfare CEV is computed separately from its stated target–reference comparison.

E.4. Details of the accounting comparison

Table 3 compares two valuation concepts using the same canonical simulated households. Panel A is a fixed-behavior accounting calculation. It values unpaid scheduled OASI payments along the full-payment benchmark path: payments from 2034 onward are multiplied by $1 - \phi = 0.22$ and discounted at $\beta = 0.97$. Panel B reports CEVs from the matched structural environments. The purpose is to compare what the concepts measure, not to treat their difference as a behavioral decomposition.

Panel A illustrates two accounting conventions. The lifetime convention discounts each

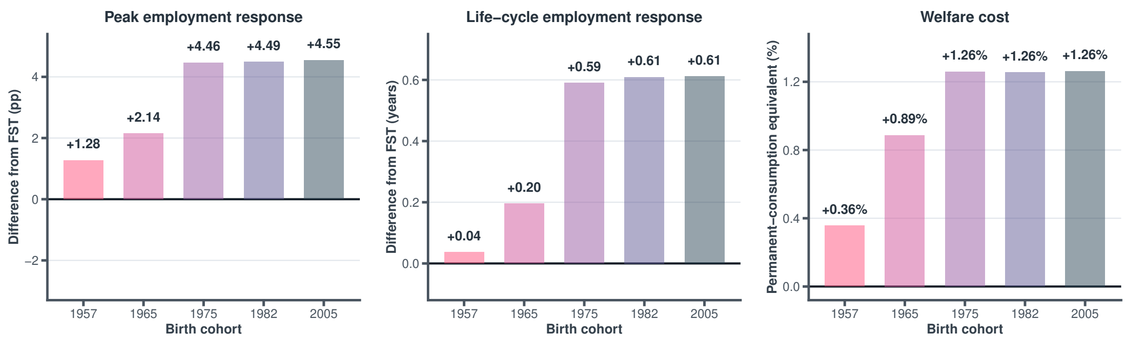

FIGURE E.5. Employment responses and welfare costs by birth cohort

Note. All outcomes compare the anticipated-shortfall environment with the full-payment FST benchmark. The left panel reports the signed age-specific employment-rate difference with the largest absolute magnitude, in percentage points. The middle panel reports the sum of age-specific employment-rate differences over the life-cycle, in years. The right panel reports the consumption-equivalent variation as a percentage of the reference scenario's age-18–80 consumption path, with welfare losses shown as positive values. All three panels report the same five birth cohorts. Reduced payments begin after the claiming window for the 1957 cohort, during it for the 1965 cohort, and before it for the 1975, 1982, and 2005 cohorts.

cohort's loss to age 18 and expresses it as a share of scheduled OASI wealth or lifetime consumption. The common-date convention discounts each cohort's unpaid payments to 2033 and normalizes the largest value to 100. These conventions answer different questions, so they need not rank cohorts in the same order.

Under the lifetime convention, the loss as a share of lifetime consumption rises from 0.29 percent for the 1957 cohort to 1.49 percent for the 1975 cohort and then plateaus. Once reserve depletion precedes the claiming window, the entire scheduled payment stream is exposed in this stationary model. Under the common-date convention, the loss instead peaks for the 1975 cohort and declines for younger cohorts: the index is 40 for the 2005 cohort and 100 for the 1975 cohort because the younger cohort's unpaid payments occur much later. These rankings differ because the lifetime convention uses cohort-specific valuation dates and denominators. Changing one common valuation date to another would not change the ranking of present-value dollar losses.

The CEV supplies a third ranking because it measures a different object. The 2005 cohort has a common-date payment-loss index of 40, below the 1965 cohort's index of 90, but their total CEVs are 1.263 and 0.887 percent, respectively. The CEV values consumption, leisure, and choices in each full model environment rather than pricing only the missing transfers at an external date.

The levels should be read with the same distinction in mind. Accounting losses as shares of lifetime consumption are 0.293, 0.992, and 1.488 percent for 1957, 1965, and the three fully exposed younger cohorts. Total CEV costs are 0.357, 0.887, and approximately 1.26 percent, respectively. The oldest cohort's welfare cost is about 22 percent above its accounting loss, whereas the other cohorts' welfare costs are about 11–16 percent below. These gaps compare different valuation concepts. They cannot be labeled the share recovered by saving, work, human-capital investment, or claiming.

Separately, the three matched model comparisons clarify the welfare plateau among the youngest cohorts. Their surprise-shortfall CEVs rise from 1.157 to 1.194 to 1.334 percent, while the anticipation comparison changes from a cost of 0.106 percent of reference lifetime consumption for the 1975 cohort to a gain of 0.073 percent for the 2005 cohort. The total CEVs, computed directly rather than by adding component CEVs, are approximately 1.26 percent for all three cohorts.

Two qualifications apply. First, the full-payment benchmark assigns the same age profiles of scheduled payments and consumption to the 1975, 1982, and 2005 cohorts because the model contains no calendar-time growth; this stationarity produces the exact accounting plateau. Second, $\beta = 0.97$ is an illustrative discounting convention shared with the annuity calculation, not a claim about the trust fund's actuarial interest rate. Other discount rates would change the common-date levels and can change their relative magnitudes.

TABLE E.1. Life-cycle responses to permanent reduced OASI payments by birth cohort

	1957	1965	1975	1982	2005
Age in event year	76	68	58	51	28
Panel A. Surprise-shortfall effect					
Consumption at first reduced-payment age	-0.087	-0.343	-0.133	-0.082	-0.049
Employment at first reduced-payment age	0.305	3.890	2.960	1.180	0.265
Assets at first reduced-payment age	0.000	0.000	0.000	0.000	0.000
Human capital at first reduced-payment age	0.000	0.000	0.000	0.000	0.000
Paid OASI at first reduced-payment age	-1.407	-1.410	0.000	0.000	0.000
Consumption-equivalent welfare cost	0.229	0.756	1.157	1.194	1.334
	(0.004)	(0.014)	(0.020)	(0.020)	(0.025)
Panel B. Anticipation effect					
Consumption at first reduced-payment age	0.085	0.107	0.082	0.045	0.017
Employment at first reduced-payment age	-0.450	-1.745	-1.585	-0.950	-0.150
Assets at first reduced-payment age	9.135	9.421	6.402	4.711	1.520
Human capital at first reduced-payment age	-0.013	0.044	0.095	0.034	0.059
Paid OASI at first reduced-payment age	-0.089	-0.060	0.000	0.000	0.000
Consumption-equivalent welfare cost	0.129	0.134	0.106	0.063	-0.073
	(0.036)	(0.035)	(0.037)	(0.037)	(0.031)
Panel C. Total anticipated-shortfall effect					
Consumption at first reduced-payment age	-0.002	-0.236	-0.051	-0.037	-0.031
Employment at first reduced-payment age	-0.145	2.145	1.375	0.230	0.115
Assets at first reduced-payment age	9.135	9.421	6.402	4.711	1.520
Human capital at first reduced-payment age	-0.013	0.044	0.095	0.034	0.059
Paid OASI at first reduced-payment age	-1.495	-1.470	0.000	0.000	0.000
Consumption-equivalent welfare cost	0.357	0.887	1.259	1.255	1.263
	(0.035)	(0.034)	(0.038)	(0.040)	(0.039)

Note. Payments fall to 78 percent of scheduled OASI in 2034 and never recover. Panels A, B, and C compare surprise with full payment, anticipation with surprise, and anticipation with full payment, respectively. Anticipation uses HRS-disciplined, index-mapped subjective hazards. Employment differences are percentage points; other age-specific outcomes are in model units. CEV costs are percentages of reference consumption over ages 18–80, with losses positive. Parentheses give paired conditional Monte Carlo standard errors from 20,000 individuals per cohort, holding model parameters and empirical inputs fixed.

E.5. Complete duration comparisons

Figures E.6–E.8 retain all eight expected-duration scenarios underlying the selected comparisons in Figures 5, 6, and 8. They use the same simulations and conditional Monte Carlo intervals as the main text.

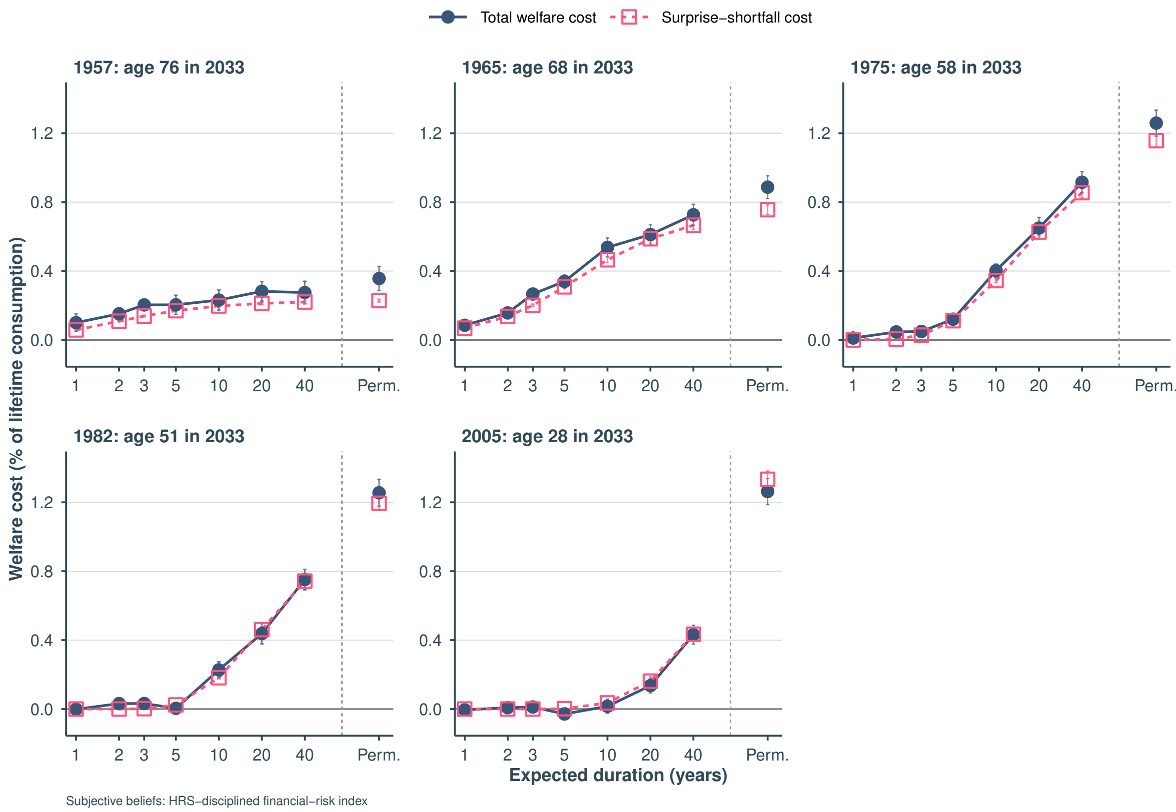

FIGURE E.6. Welfare costs across the complete duration frontier

Note. Each panel identifies a birth cohort and its age in 2033. Navy circles show total welfare cost (anticipated shortfall versus full payment); pink open squares show surprise-shortfall cost (surprise shortfall versus full payment). Losses are positive. Lines connect the seven finite expected-duration scenarios as visual guides on a logarithmic horizontal scale. Permanent denotes zero restoration hazard and is shown separately. Vertical bars are paired 95-percent conditional Monte Carlo intervals from 20,000 simulated people per cohort–duration cell. These intervals cover finite-simulation variation with parameters and empirical inputs fixed, not model, estimation, or multiple-comparison uncertainty.

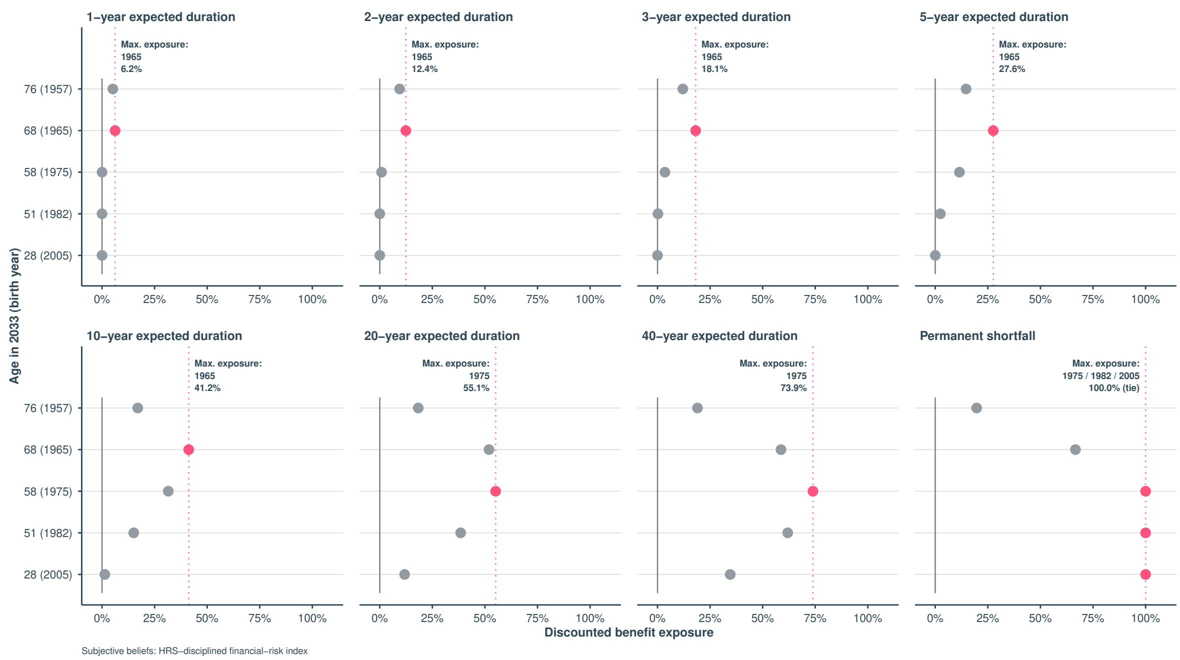

FIGURE E.7. Mechanical benefit exposure across all duration scenarios

Note. Panels report all eight expected-duration scenarios; rows give age in 2033, with birth year in parentheses. Each point is the discounted share of scheduled OASI benefits expected to be received while payments are reduced, evaluated along the full-payment benchmark path. Pink markers and dotted vertical lines identify each panel's largest exposure, including ties. Permanent denotes zero restoration hazard. These exposure rankings do not measure welfare-cost rankings.

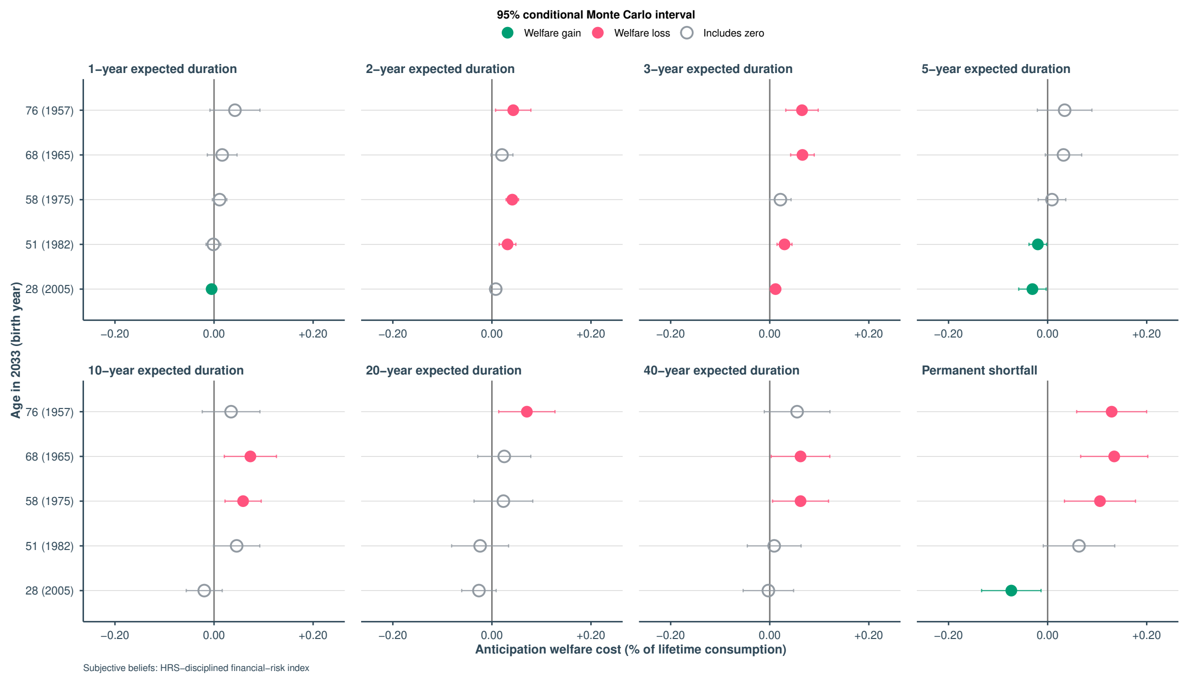

FIGURE E.8. Welfare effect of anticipation across all duration scenarios

Note. Each point reports $\mathcal{C}^{A}_{cd}$, with positive values indicating welfare losses and negative values indicating welfare gains from anticipation relative to surprise. Green filled markers identify paired 95-percent conditional Monte Carlo intervals entirely below zero; pink filled markers identify intervals entirely above zero. Gray open markers identify intervals including zero or absolute contrasts no larger than 10^{-10} percentage points. The one-year 2005 estimate is a small gain; its two-year interval includes zero. Horizontal bars show intervals, and vertical lines mark zero. Rows give age in 2033, with birth year in parentheses. The intervals use 20,000 simulated people per cohort–duration cell and cover finite-simulation variation only.