Taxes and Transfers with Nonlinear Wage Dynamics

Nezih Guner Eugenio Renedo Emre Enes Yavuz

August 20, 2026

Abstract

We study how the specification of wage risk in quantitative life-cycle models changes the welfare evaluation of taxes and transfers. We estimate four wage processes: a canonical linear process; a flexible nonlinear process with age dependence, non-normality, and state-dependent persistence; and two intermediate processes. Comparisons across the four isolate the role of each feature: age dependence, non-normality, and nonlinear persistence. The estimated processes imply different allocations of wage risk across permanent, persistent, and transitory components: the richer processes assign less risk to short-lived transitory shocks and more to long-lived permanent and persistent components. We embed each process in the same incomplete-markets life-cycle economy with endogenous labor supply. Removing federal tax progressivity and means-tested transfers raises mean welfare by 0.72% of lifetime consumption under the nonlinear process but lowers it by 0.47% under the non-normal process. These welfare differences reflect both the total amount of wage risk and its allocation across permanent heterogeneity, persistent shocks, and transitory shocks. Holding this amount and allocation fixed, non-normal shocks create rare but severe low-income states that increase the value of public insurance, while the state-dependent mean reversion that characterizes the nonlinear specification works in the opposite direction: unusually bad persistent states tend to be undone by subsequent favorable shocks, so the wage process partly insures itself, and that lowers the value of public insurance.

Keywords: Wage risk, Nonlinear wage dynamics, Social insurance, Taxes and transfers, Life-cycle models

JEL Codes: D15, D31, E21, H21, I38

Corresponding Author: Eugenio Renedo (eugenio.renedo@cemfi.edu.es). Guner: CEMFI and Banco de España. Renedo: CEMFI. Yavuz: Reddit. Renedo acknowledges financial support from Grant PRE2022-000905 funded by MCIN/AEI/ 10.13039/501100011033 and by "ESF +", the Maria de Maeztu Unit of Excellence CEMFI MDM-2016-0684 and CEMFI CEX2020-001104-M. We thank Manuel Arellano, Stephane Bonhomme, Anmol Bhandari, Mariacristina De Nardi, Matthias Doepke, Serdar Ozkan, and Josep Pijoan Mas for helpful comments. The views expressed herein are those of the authors and not necessarily those of Banco de España or the Eurosystem.

1. Introduction

Progressive taxes and means-tested transfers provide insurance against bad labor-market outcomes and redistribute resources toward households with low earning capacity. Quantitative life-cycle models with incomplete markets and endogenous labor supply are the standard laboratory for evaluating these policies. At the heart of these models is a process for individual wages. This paper asks whether the specification of that process matters for the welfare evaluation of tax-and-transfer policy. The answer is yes: the value of progressive taxes and transfers changes in economically meaningful ways when the canonical wage process is replaced by richer wage dynamics.

The value of social insurance depends on the dispersion and persistence of the wage process that households face. The canonical wage process used in many quantitative models is convenient and parsimonious: log wages consist of a permanent component, a persistent AR(1) component, and a transitory shock, with normally distributed innovations and age-invariant parameters. But a growing empirical literature shows that this process misses important features of the data. Earnings changes are highly leptokurtic and left-skewed (Guvenen et al. 2021); the persistence and the dispersion of shocks vary over the life cycle (Karahan and Ozkan 2013); and the persistence of past shocks is state-dependent: unusual shocks erase the memory of past shocks (Arellano, Blundell, and Bonhomme 2017). These departures from the canonical process, age dependence, non-normality, and nonlinear persistence, are precisely the features that can change private self-insurance and, therefore, the value of public insurance.

We study this question in three steps. First, we estimate four wage processes on the same sample of male hourly wages from the Panel Study of Income Dynamics (PSID). We focus on wages rather than earnings because earnings combine wage risk with endogenous hours choices, while our policy counterfactuals directly affect labor-supply incentives. The first process is the canonical linear process. The second is our benchmark nonlinear process, estimated using the quantile-based method of Arellano, Blundell, and Bonhomme (2017); it allows for age dependence, non-normality, and nonlinear persistence. We also estimate two intermediate processes that isolate particular departures from the canonical model. An age-dependent process, in the spirit of Karahan and Ozkan (2013), retains Gaussian shocks and linear persistence while allowing persistence and shock variances to vary over the life cycle. A non-normal process, following Gálvez and Paz-Pardo (2026), allows flexible non-Gaussian conditional distributions while preserving linear dependence of the persistent component on its lag. The estimated processes differ not only in shock shape but also in the allocation of wage risk across components: the richer processes assign less risk to short-lived transitory shocks and more to long-lived permanent and persistent components.

Second, we embed each estimated wage process in the same life-cycle economy with incomplete markets, endogenous labor supply, progressive federal income taxation, and means-tested transfers. We recalibrate each economy to match the same aggregate targets. This ensures that the welfare comparisons are not driven by different baseline aggregate calibrations. We then evaluate three revenue-neutral counterfactuals: replacing the progressive federal income tax with a proportional tax, removing means-tested transfers, and removing both federal progressivity and means-tested transfers.

Third, we ask how well the different wage processes reproduce a set of untargeted moments. The model is not calibrated to match the cross-sectional distribution of consumption or hours, so we ask which wage process best reproduces these moments in the PSID. The key statistic is the pass-through of persistent wage shocks to consumption, measured using the approach of Blundell, Pistaferri, and Preston (2008). This statistic is informative for welfare because persistent shocks are hard to self-insure and are therefore the shocks for which public insurance is most valuable. We also examine life-cycle profiles of consumption and hours dispersion. This exercise favors the nonlinear process: it comes closest to the observed pass-through of persistent wage shocks to consumption and performs best overall on the untargeted moments. The linear and age-dependent processes pass too much persistent wage risk into consumption, while the non-normal process captures the fat-tailed shape of wage changes but loads more risk onto permanent components and implies too much pass-through relative to the nonlinear benchmark.

We find that the same tax-transfer reform can have different welfare implications under different wage processes. Removing both federal progressivity and means-tested transfers while keeping Social Security in place raises mean welfare by 0.72% of lifetime consumption under the nonlinear process, but lowers mean welfare by 0.47% under the non-normal process. Welfare effects also vary sharply by permanent type: high permanent types gain the most from reducing progressivity and transfers, while low permanent types bear the largest losses when transfers are removed. These distributional effects are especially steep under the nonlinear and non-normal processes, which assign a larger role to permanent heterogeneity. Decomposing the cross-process welfare differences shows that most of the gap reflects both the total amount of wage risk and its allocation across permanent heterogeneity, persistent shocks, and transitory shocks. But the shape of risk also matters. Nonlinear mean reversion lowers the value of social insurance because the wage process partly insures itself: bad persistent realizations tend to move back toward the middle of the distribution. Non-normality works in the opposite direction: fat-tailed and skewed shocks generate rare but severe bad states, precisely the states in which means-tested transfers are valuable.

Our paper contributes to three literatures. First, it contributes to the quantitative public-finance literature that uses heterogeneous-agent life-cycle models to evaluate taxes and transfers, including Conesa and Krueger (2006), Guner, Lopez-Daneri, and Ventura (2016), Heathcote, Storesletten, and Violante (2017), Kindermann and Krueger (2022), Ferriere et al. (2023), and Guner, Kaygusuz, and Ventura (2023). These papers show that tax and transfer policies must be evaluated in models that account for heterogeneity, self-insurance, labor supply, and general equilibrium. We add that the stochastic process for wages is itself a quantitatively important input for such policy evaluations.[1]

Second, the paper relates to the literature on income and wage dynamics in life-cycle models. Canonical processes of the type used by Storesletten, Telmer, and Yaron (2004) and Kaplan (2012) have been central to the study of consumption, savings, and labor supply under incomplete markets. Recent empirical work, however, has documented richer dynamics, including age-varying persistence (Karahan and Ozkan 2013), non-normal earnings changes (Guvenen et al. 2021), and nonlinear persistence (Arellano, Blundell, and Bonhomme 2017). Golosov, Troshkin, and Tsyvinski (2016) show that higher moments of shocks matter for optimal redistribution and insurance. We bring these empirical advances into a tax-and-transfer exercise and quantify which features matter for welfare.

Finally, our paper is closest to work that studies the structural implications of richer earnings or wage dynamics. De Nardi, Fella, and Paz-Pardo (2020) estimate nonlinear household earnings processes and study their implications for consumption, self-insurance, and welfare. De Nardi et al. (2021) document how wage, earnings, family, and government insurance interact in the Netherlands and the United States. De Nardi, Fella, and Paz-Pardo (2025) estimate rich wage processes and show that they lead to different conclusions about the design of benefit systems. Guvenen, Ozkan, and Madera (2024) study consumption dynamics and welfare under non-Gaussian earnings risk, and Gálvez and Paz-Pardo (2026) study how richer earnings dynamics affect life-cycle portfolio choice. We complement this work by focusing on wage risk, modeling labor supply endogenously, and decomposing how age dependence, non-normality, nonlinear persistence, and the allocation of risk across wage components shape the value of taxes and transfers.

The rest of the paper is organized as follows. Section 2 specifies and estimates the four wage processes and reports the estimation results. Section 3 presents the life-cycle model, its calibration, and the untargeted validation exercise. Section 4 evaluates the tax-transfer counterfactuals and the value of social insurance. Section 5 concludes.

[1] More broadly, the paper is related to work on how taxes and transfers affect inequality and labor supply. See, among others, Guner, Kaygusuz, and Ventura (2012) and Borella, De Nardi, and Yang (2023).

2. Wage Dynamics and Estimation

The empirical object in this section is the residual log hourly wage. Let ω_{ij} denote the log wage of individual i at age j, after removing the observable socioeconomic characteristics and time effects described below. We assume that log wages consist of four components:

$$\omega_{ij} = \kappa_j + \zeta_i + \eta_{ij} + \varepsilon_{ij}, \tag{1}$$

where κ_j is a deterministic age profile common to all individuals, ζ_i is a permanent component drawn at the beginning of the working life, η_{ij} is a persistent component that follows a Markov process, and ε_{ij} is a transitory component. After estimating κ_j, we define $x_{ij} \equiv \omega_{ij} - \kappa_j$ and use x_{ij} to identify the stochastic components.

Equation (1) is the standard starting point in quantitative life-cycle models with heterogeneous agents; see, for example, Huggett (1996), Storesletten, Telmer, and Yaron (2004), Kaplan (2012), Guner, Lopez-Daneri, and Ventura (2016), and Heathcote, Storesletten, and Violante (2017). We estimate four versions of this decomposition that differ in three dimensions: age dependence, non-normality, and nonlinear persistence. We first describe the canonical linear process and the benchmark nonlinear process. By canonical, we mean a specification in which the permanent and transitory components are normally distributed, and the persistent component follows a Gaussian AR(1) process with constant persistence and an age-invariant innovation variance. Thus, the canonical process features Gaussian shock distributions, age-invariant parameters, and linear persistence with a constant autoregressive coefficient.

The benchmark nonlinear process relaxes all three restrictions, following Arellano, Blundell, and Bonhomme (2017). Its conditional distributions may vary with age and need not be Gaussian, and local persistence may depend on age, the lagged persistent state, and the current innovation. This flexibility accommodates state-dependent mean reversion: an unusually favorable draw for a worker with a low lagged persistent component, or an unusually adverse draw for a worker with a high lagged persistent component, can partly undo the effects of past shocks and move the worker back toward the middle of the distribution.

We then introduce two intermediate specifications. The age-dependent process retains Gaussian shocks and a linear AR(1) structure, but allows the persistence coefficient, the variance of the innovations to the persistent process, and the variance of transitory shocks to vary over the life cycle, as in Karahan and Ozkan (2013). The non-normal process, following Gálvez and Paz-Pardo (2026), allows flexible, non-Gaussian distributions for all three components. It also allows the distributions of the transitory and persistent components to vary with age. The persistent component nevertheless remains linear in its lag: its slope may vary across innovation quantiles, but not with the level of the lagged persistent state. Comparing the four processes helps us distinguish the roles of age dependence, non-normality, and nonlinear persistence.

2.1. Data and Sample Selection

We use data from the Panel Study of Income Dynamics (PSID), harmonized through the Comparative Panel File (Turek, Kalmijn, and Leopold 2021), for the period 1980–2019. The PSID interviewed households annually up to and including the 1997 wave, and every two years thereafter. Because income and hours refer to the previous calendar year, the annual waves cover income and hours through 1996. The sample is restricted to households headed by men aged 25 to 64. We exclude the Survey of Economic Opportunity sample and observations with missing information on age, education, race, or state of residence.

The analysis uses hourly wages for male household heads aged 25 to 64, calculated as annual real labor income divided by total annual hours worked. Following standard practice (e.g., Heathcote, Perri, and Violante 2010), we drop observations with hourly wages below one-half of the applicable state minimum wage, and with fewer than 520 or more than 5,200 yearly hours worked. We also remove observations with implausibly large wage changes that are immediately reversed, which likely reflect measurement error. To obtain a panel suitable for estimating wage dynamics, we retain wage observations that are linked over five consecutive periods in the annual part of the sample and over three consecutive periods in the biennial part. Appendix A provides additional details on the data construction.

We residualize log wages using a first-stage regression. The controls include state-of-residence and year fixed effects, an indicator for whether the wife earns income, and fixed effects for household size and the number of children in the household. Finally, the education and ethnicity of the household head are included, with education additionally interacted with year fixed effects to capture the changing returns to education over time.[2] From the resulting residuals, we estimate the deterministic age profile κ_j using a fourth-order Hermite polynomial in age. Figure 1 plots the estimated profile. The residual wage net of this profile, $x_{ij} = \omega_{ij} - \kappa_j$, is the object used in the estimation below.

2.2. Four Wage Processes

Linear process. The canonical process assumes that the permanent component, the transitory component, and the innovations to the persistent component are Gaussian and age invariant.

[2] Education is the CPF's four-level, ISCED-based classification of the highest grade completed: grades 1-8 (below secondary), 9-11 (some high school), 12-15 (high-school diploma up to some college), and 16-17 (college degree or more). Ethnicity takes five values: white, Black, Asian, American Indian, and other.

FIGURE 1. Deterministic age profile

NOTES: The figure plots the estimated deterministic age profile, κ_j, for residual log hourly wages. The profile is estimated from PSID data for 1980–2019 after removing observable demographic and household characteristics in a first-stage regression, and fitting a fourth-order Hermite polynomial in age.

In particular,

$$\zeta_i \sim \mathcal{N}(0, \sigma_\zeta^2), \qquad \varepsilon_{ij} \sim \mathcal{N}(0, \sigma_\varepsilon^2), \tag{2}$$

and the persistent component follows

$$\eta_{ij} = \rho\eta_{i,j-1} + \nu_{ij}, \qquad \nu_{ij} \sim \mathcal{N}(0, \sigma_\nu^2), \qquad \eta_{i,j_0-1} = 0, \tag{3}$$

where $j_0 = 25$ is the first working age. The draws are mutually independent across individuals, ages, and components; serial dependence in η_{ij} arises only through the AR(1) recursion. Thus, conditional on $\eta_{i,j-1}$, the persistent component is normally distributed with mean $\rho\eta_{i,j-1}$ and variance σ_ν^2. Persistence is governed by the single parameter ρ, the conditional variance is constant, and all higher conditional moments are pinned down by normality.

Nonlinear process. Following Arellano, Blundell, and Bonhomme (2017), the nonlinear process replaces the Gaussian parametric distributions with flexible conditional quantile functions and allows the transition of the persistent component to depend nonlinearly on its lagged state and on age. Let $h_j \equiv j - j_0$ denote the number of years since age j_0. Let u_i^ζ, $\{u_{ij}^\eta\}_j$, and $\{u_{ij}^\varepsilon\}_j$ be mutually independent uniform random variables on $(0, 1)$, also independent across individuals. We write

$$\zeta_i = Q_\zeta(u_i^\zeta \mid j_{i0}),$$

$$\varepsilon_{ij} = Q_\varepsilon(u^\varepsilon_{ij} \mid h_j),$$
$$\eta_{ij} = Q_\eta(u^\eta_{ij} \mid \eta_{i,j-1}, h_j),$$

where j_{i0} is the age at which individual i first enters the estimation sample. The age j_{i0} should not be confused with the first working age $j_0 = 25$: j_0 is a constant, common to all individuals, whereas j_{i0} is individual-specific because workers enter the panel at different ages, and the estimator cannot condition on histories before a worker is first observed. For any fixed values of the conditioning variables, $Q_x(u \mid \cdot)$ denotes the uth conditional quantile of component x. Drawing u uniformly from $(0, 1)$ and evaluating the quantile function therefore generates a draw from the corresponding conditional distribution. The uniform variable u^η_{ij} can be interpreted as the rank of the current persistent innovation within that distribution.

The conditional quantile function for the persistent component is approximated by a tensor product of Hermite polynomial bases,

$$Q_\eta(u^\eta_{ij} \mid \eta_{i,j-1}, h_j) = \sum_{k_1=0}^{K_1} \sum_{k_2=0}^{K_2} a^\eta_{k_1,k_2}(u^\eta_{ij}) H_{k_1}(\eta_{i,j-1}) H_{k_2}(h_j), \tag{4}$$

where H_k is the probabilists' Hermite polynomial of order k. We set $K_1 = 3$ and $K_2 = 2$, which gives 12 basis functions, including interactions between the lagged persistent state and age.

The persistent component at entry into the estimation sample, denoted by $\eta_{i0} \equiv \eta_{i,j_{i0}}$, is specified as

$$Q_{\eta_0}(u^\eta_{i0} \mid \zeta_i, j_{i0}) = \sum_{k_1=0}^{K_3^\zeta} \sum_{k_2=0}^{K_3} a^{\eta_0}_{k_1,k_2}(u^\eta_{i0}) H_{k_1}(\zeta_i) H_{k_2}(j_{i0}). \tag{5}$$

We set $K_3^\zeta = K_3 = 1$. The four resulting basis terms allow the initial persistent component to depend on the permanent component, age at entry, and their interaction. This initial-conditions specification accounts for the fact that individuals enter the estimation panel at different ages and permits dependence between their permanent component and the persistent state at first observation.

The transitory and permanent quantile functions are given by

$$Q_\varepsilon(u^\varepsilon_{ij} \mid h_j) = \sum_{k=0}^{K_4} a^\varepsilon_k(u^\varepsilon_{ij}) H_k(h_j), \tag{6}$$

and

$$Q_\zeta(u_i^\zeta \mid j_{i0}) = \sum_{k=0}^{K_5} a_k^\zeta(u_i^\zeta) H_k(j_{i0}), \tag{7}$$

with $K_4 = K_5 = 1$. Thus, at each quantile, the transitory component depends linearly on age and the permanent component depends linearly on age at entry. Because the coefficients vary across quantiles, however, age can affect more than the conditional mean: it can also change the dispersion and shape of the distributions. Conditioning Q_ζ on j_{i0} does not make the permanent component vary with age within an individual. Each individual draws ζ_i once; the conditioning allows the distribution of permanent types among individuals first observed at different ages to differ.

This specification relaxes the canonical process along three dimensions. First, the conditional distributions of the persistent and transitory components can vary over the life cycle, while the distributions at entry can depend on the individual's age when first observed. Second, the coefficient functions are estimated flexibly across quantiles rather than being restricted by a Gaussian location-scale specification. The implied distributions may therefore be asymmetric, fat-tailed, and conditionally heteroskedastic. Third, the effect of the lagged persistent state is not summarized by a single AR(1) coefficient. Instead, local persistence is defined as

$$\rho(\eta_{i,j-1}, u_{ij}^\eta, j) \equiv \frac{\partial Q_\eta(u_{ij}^\eta \mid \eta_{i,j-1}, h_j)}{\partial \eta_{i,j-1}}, \tag{8}$$

and can vary with age, the level of the lagged persistent state, and the rank of the current innovation. This flexibility allows the state-dependent mean reversion emphasized above.

For each quantile function, we estimate the coefficient functions on a grid $\tau_1 < \cdots < \tau_L$ in $(0,1)$. We set $L = 11$ and $\tau_l = l/(L+1)$. Between adjacent grid points, we use linear interpolation. Below the lowest grid point and above the highest grid point, we append Laplace tails. For a generic component x, the approximated quantile function is

$$\widehat{Q}_x(u \mid \cdot) = \begin{cases} \sum_k a_k^x(\tau_1) H_k(\cdot) + \frac{1}{\lambda_L^x} \ln\left(\frac{u}{\tau_1}\right), & u \leq \tau_1, \\ \frac{\tau_{l+1} - u}{\tau_{l+1} - \tau_l} \sum_k a_k^x(\tau_l) H_k(\cdot) + \frac{u - \tau_l}{\tau_{l+1} - \tau_l} \sum_k a_k^x(\tau_{l+1}) H_k(\cdot), & \tau_l < u \leq \tau_{l+1}, \\ \sum_k a_k^x(\tau_L) H_k(\cdot) - \frac{1}{\lambda_U^x} \ln\left(\frac{1-u}{1-\tau_L}\right), & u > \tau_L, \end{cases} \tag{9}$$

where $(\cdot)$ denotes the conditioning arguments of the corresponding quantile function, with k understood as a multi-index when there is more than one conditioning variable. The hat distinguishes this interpolated approximation from the abstract quantile functions in equations

(4)–(7). The parameters λ_L^x and λ_U^x govern the lower and upper tails. The middle line flexibly interpolates the central part of the conditional distribution, while the first and third lines extend the quantile function beyond the extreme estimated quantiles with exponentially decaying tails.

Age-dependent process. Following Karahan and Ozkan (2013), the age-dependent process retains the Gaussian, single-lag structure of the linear process but allows persistence and the innovation variances to change over the life cycle. The permanent component is still given by $\zeta_i \sim \mathcal{N}(0, \sigma_\zeta^2)$, but the persistent component follows

$$\eta_{ij} = \rho_j \eta_{i,j-1} + \nu_{ij}, \qquad \nu_{ij} \sim \mathcal{N}(0, \sigma_{\nu,j}^2), \qquad \eta_{i,j_0} \sim \mathcal{N}(0, \sigma_{\eta,j_0}^2). \tag{10}$$

The variance of the initial persistent state, σ_{η,j_0}^2, is estimated separately. The transitory component is also age-dependent,

$$\varepsilon_{ij} \sim \mathcal{N}(0, \sigma_{\varepsilon,j}^2).$$

We parameterize the age profiles parsimoniously as cubic polynomials,

$$\rho_j = \sum_{m=0}^{3} c_m^\rho \left(h_j\right)^m, \qquad \sigma_{\nu,j}^2 = \sum_{m=0}^{3} c_m^\nu \left(h_j\right)^m, \qquad \sigma_{\varepsilon,j}^2 = \sum_{m=0}^{3} c_m^\varepsilon \left(h_j\right)^m. \tag{11}$$

Conditional on age and the lagged persistent state, the transition remains linear and Gaussian: its mean is $\rho_j \eta_{i,j-1}$ and its variance is $\sigma_{\nu,j}^2$. Thus, unlike in the nonlinear process, persistence does not depend on the level of the lagged state or on the realization of the current innovation. Setting the age-slope coefficients to zero yields an age-invariant Gaussian specification, with the initial persistent variance estimated separately. The comparison with the linear process therefore measures the importance of allowing persistence and shock variances to vary systematically over the life cycle.

Non-normal process. Following Gálvez and Paz-Pardo (2026), the non-normal process uses the same quantile-function framework as the nonlinear process but restricts the persistent component to depend linearly on its lag. The permanent component, the initial persistent component, and the transitory component are specified by the quantile functions in equations (5)-(7), using the same quantile grid, interpolation, and Laplace-tail construction in equation (9). Consequently, the distributions of the permanent component, the transitory component, and the initial persistent state may all be non-Gaussian. The one-year-ahead conditional distribution of the persistent component may also be non-Gaussian. The difference from the benchmark nonlinear process lies in the conditional quantile function governing these subsequent persistent transitions. It is restricted to the additive form used by Gálvez and

Paz-Pardo (2026),

$$Q_\eta\left(u^\eta_{ij} \mid \eta_{i,j-1}, h_j\right) = b_0(u^\eta_{ij}) + b_1(u^\eta_{ij})\eta_{i,j-1} + g_1(u^\eta_{ij})h_j + g_2(u^\eta_{ij})h^2_j, \tag{12}$$

where $u^\eta_{ij} \sim U(0,1)$. The functions b_0, b_1, g_1, and g_2 are piecewise-linear splines in the quantile index, with Laplace tails appended at the extremes. Because these coefficient functions vary across quantiles, the conditional distribution may exhibit skewness, fat tails, and quantile-specific dispersion. Moreover, because g_1 and g_2 vary with the quantile index, age can shift different parts of the conditional distribution by different amounts rather than affecting only its mean.

The local slope with respect to the lagged persistent state is

$$\frac{\partial Q_\eta(u^\eta_{ij} \mid \eta_{i,j-1}, h_j)}{\partial \eta_{i,j-1}} = b_1(u^\eta_{ij}).$$

Persistence may therefore differ across innovation quantiles. For any given innovation rank, however, the slope does not vary with $\eta_{i,j-1}$ or age. The transition is thus linear in the lagged persistent state and rules out the state-dependent mean reversion that is central to the nonlinear specification. Age still affects the conditional distribution through $g_1(u^\eta_{ij})h_j + g_2(u^\eta_{ij})h^2_j$, but not through an age-varying coefficient on the lagged persistent state.

2.3. Estimation

Linear and age-dependent processes. We estimate the linear and age-dependent processes by minimum distance, matching age-indexed autocovariances of x_{ij}. Let $N_{(j,j+s)}$ denote the number of individuals observed at both ages j and $j+s$. The empirical covariance is

$$\widehat{\text{cov}}(x_j, x_{j+s}) = \frac{1}{N_{(j,j+s)} - 1} \sum_{i=1}^{N_{(j,j+s)}} x_{ij}x_{i,j+s}, \tag{13}$$

where the sum is over individuals for whom both observations are available.

Under the linear process, the model-implied covariance is

$$\text{cov}(x_j, x_{j+s}) = \rho^s \left(\frac{1-\rho^{2(h_j+1)}}{1-\rho^2}\right)\sigma^2_\nu + \sigma^2_\zeta + \mathbf{1}\{s=0\}\sigma^2_\varepsilon, \tag{14}$$

where $h_j = j - j_0$. The first term is the covariance generated by the persistent component: its variance at age j is propagated forward by ρ^s for s years. The permanent component contributes σ^2_ζ at every lag because it is common to all observations of the same individual. The transitory component contributes only when $s = 0$ because it is independent over time.

For the age-dependent process, we first propagate the variance of the persistent component according to

$$\text{Var}(\eta_j) = \rho_j^2 \text{Var}(\eta_{j-1}) + \sigma^2_{\nu,j}, \qquad \text{Var}(\eta_{j_0}) = \sigma^2_{\eta,j_0}, \tag{15}$$

and then use

$$\text{cov}(x_j, x_{j+s}) = \left(\prod_{r=1}^{s} \rho_{j+r}\right) \text{Var}(\eta_j) + \sigma^2_\zeta + \mathbf{1}\{s = 0\}\sigma^2_{\varepsilon,j}. \tag{16}$$

The product of the age-specific persistence coefficients carries the age-j persistent state forward to age $j + s$. As in the linear process, the permanent component contributes at every lag and the transitory component contributes only to the contemporaneous variance.

Let $\widehat{M}$ and $M(\Theta)$ denote the vectors of empirical and model-implied covariances, stacked across the observed age-lag combinations. For the linear process, $\Theta = (\rho, \sigma^2_\zeta, \sigma^2_\nu, \sigma^2_\varepsilon)$. For the age-dependent process, Θ contains the polynomial coefficients $\{c^\rho_m, c^\nu_m, c^\varepsilon_m\}^3_{m=0}$, together with σ^2_ζ and σ^2_{η,j_0}. We choose the parameters to solve

$$\min_{\Theta} \; \big(\widehat{M} - M(\Theta)\big)' W \big(\widehat{M} - M(\Theta)\big). \tag{17}$$

The diagonal weights are chosen so that the squared residual of each covariance is weighted by $\sqrt{N_{(j,j+s)}}$. Moments based on more observation pairs therefore receive greater weight.

Nonlinear and non-normal processes. The nonlinear and non-normal processes share the same latent-variable structure and are estimated with the quantile-function-based stochastic EM algorithm of Arellano, Blundell, and Bonhomme (2017). The two estimators differ only in the specification of the transition for the persistent component: the nonlinear process allows the transition to depend flexibly on the lagged state and age, equation (4), while the non-normal process restricts this dependence to be linear in the lagged state, equation (12). The permanent component, the initial persistent component, the transitory component, the quantile grid, and the treatment of the tails are otherwise common to both processes.

Let Λ collect all parameters to be estimated: the quantile-function coefficients at the grid points and the lower- and upper-tail parameters, for every component. If the permanent, persistent, and transitory components were observed separately, estimation would be straightforward: we would run the corresponding quantile regressions for each component and estimate the tail parameters by maximum likelihood. The difficulty is that, for each individual, we observe only the residual wage history $x_i = (x_{ij})_j$, where

$$x_{ij} = \zeta_i + \eta_{ij} + \varepsilon_{ij}.$$

Thus, the same wage history can be generated by many different decompositions into permanent, persistent, and transitory components.

The stochastic EM algorithm treats these components as missing data. Starting from an initial parameter vector $\widehat{\Lambda}^{(0)}$, iteration t consists of two steps.

E-step. Holding the current estimate $\widehat{\Lambda}^{(t)}$ fixed, we fill in the missing wage components. For each individual, we draw a permanent component and a path for the persistent component, (ζ_i, η_i). Given these draws, the transitory component is pinned down at every observed age by the adding-up restriction,

$$\varepsilon_{ij} = x_{ij} - \zeta_i - \eta_{ij}.$$

The draw is taken from the posterior distribution of the latent components conditional on the worker's observed wage history. Intuitively, a proposed decomposition is more likely if four objects are likely under the current wage process: the permanent type, the initial persistent state, the sequence of persistent transitions, and the transitory shocks implied by the observed residual wages.

Formally, let f_ζ, f_{η_0}, f_η, and f_ε denote the densities implied by the current quantile functions for, respectively, the permanent component, the initial persistent component, the one-year transition of the persistent component, and the transitory component. Then the posterior density of a proposed decomposition is proportional to

$$f_\zeta(\zeta_i \mid j_{i0})\, f_{\eta_0}(\eta_{i0} \mid \zeta_i, j_{i0}) \prod_{j > j_{i0}} f_\eta(\eta_{ij} \mid \eta_{i,j-1}, h_j) \prod_{j \in O_i} f_\varepsilon(x_{ij} - \zeta_i - \eta_{ij} \mid h_j),$$

where O_i is the set of ages at which individual i is observed. The first term evaluates how likely the proposed permanent type is. The second evaluates how likely the proposed initial persistent state is, given the permanent type and the age at first observation. The third evaluates the likelihood of the proposed path of persistent states; this product runs over every annual age between the first and last observation, including the intervening years at which wages are not observed in the biennial part of the sample (see Section 2.4). The final term evaluates the likelihood of the transitory shocks that are implied by the proposed permanent and persistent components and the observed residual wages.

These densities are obtained from the quantile functions. If $Q_x(u \mid z)$ denotes the conditional quantile function for component x, then its inverse is the conditional distribution function, whose derivative is the corresponding conditional density. Since the posterior has no closed-form distribution, we sample from it using a Metropolis–Hastings algorithm, which only requires evaluating the expression above up to a normalizing constant.

M-step. Treating the imputed latent paths as if they were observed data, we re-estimate Λ. This means updating the quantile-function coefficients by the corresponding quantile regressions and updating the tail parameters by maximum likelihood on the observations that fall in the tails. The nonlinear and non-normal estimators differ only in the transition equation used for the persistent component; the estimation of the permanent, initial-persistent, and transitory components is common to both.

We run 500 iterations and report the average of the parameter estimates over the final 250 iterations. Averaging over the second half of the chain reduces the simulation noise introduced by using posterior draws of the latent components rather than exact conditional expectations. Appendix B provides the formal objective function, the construction of the posterior density, the Metropolis–Hastings implementation, the densities implied by the estimated quantile functions, and the treatment of the mixed annual-biennial panel.

2.4. Combining Annual and Biennial Data

The PSID changes from annual to biennial observations after the 1997 survey wave. Since the wage process used in the model is annual, we estimate all four specifications as annual processes. For the linear and age-dependent processes, this is straightforward. The empirical moment vector contains the observed age pairs in the data, and the model-implied covariance is available at every lag s. The annual part of the PSID contributes adjacent and longer-lag covariances, while the biennial part contributes the observed longer-lag covariances.

For nonlinear and non-normal processes, the missing intervening-year persistent states in the biennial panel are treated as additional latent variables to be estimated. The E-step draws those unobserved annual persistent states jointly with the permanent, observed-year persistent, and transitory components. The remaining steps of the EM algorithm are unchanged. This procedure uses all observed wage histories while delivering the annual process required by the quantitative model.

2.5. Estimation Results

Table 1 summarizes the estimated processes in the language of the canonical decomposition: the average persistence of the persistent component, and the variances of the permanent component, the persistent innovation, and the transitory component. For the linear process, these are the estimated parameters. For the age-dependent process, they are lifetime averages of the estimated age profiles. The nonlinear and non-normal processes do not have a single AR(1) coefficient or a single additive innovation variance; we therefore report their analogues computed from the estimated quantile functions. Average persistence is the average derivative of the persistent transition with respect to its lagged state, and the innovation variance is the average conditional variance of that transition.

Three patterns stand out. First, the canonical linear process is highly persistent, with average persistence equal to 0.96, and assigns a large role to transitory shocks: $\mathrm{Var}(\varepsilon) = 0.085$, compared with $\mathrm{Var}(\zeta) = 0.046$. Second, allowing age dependence lowers average persistence and reallocates variation from transitory shocks toward persistent innovations: relative to the linear process, average persistence falls from 0.96 to 0.91, $\mathrm{Var}(\nu)$ rises from 0.020 to 0.033, and $\mathrm{Var}(\varepsilon)$ falls from 0.085 to 0.061. Third, the quantile-based processes imply a much larger role for long-lived heterogeneity. The permanent variance is 0.151 under the nonlinear process and 0.209 under the non-normal process, compared with 0.046 in the linear process. At the same time, the transitory variance falls sharply, to 0.027 and 0.040, respectively.[3]

Thus, the richer processes do not simply change the shape of a fixed amount of wage risk. They also change where wage variation is allocated: less is assigned to short-lived transitory shocks and more to permanent or persistent components. This distinction is important for the quantitative analysis. Permanent heterogeneity matters primarily for redistribution across workers with different earning capacities. Persistent shocks arrive during working life and are harder to self-insure when they are durable. Transitory shocks are short-lived and are comparatively easier to smooth through saving and labor supply.[4]

Recent empirical work emphasizes that variances and autocovariances alone provide an incomplete description of earnings dynamics. Earnings changes are sharply peaked, asymmetric, and fat-tailed; higher-order moments vary over the life cycle and across the earnings distribution; and mean reversion can depend on both the size and sign of an earnings change and the worker's position in the distribution (Guvenen et al. 2021; Guvenen, Pistaferri, and Violante 2022). Following this diagnostic approach, we examine marginal distributions, quantile-based dispersion, conditional higher moments, and persistence. The objects are not directly comparable to those in these papers, which primarily study annual earnings rather than hourly wages, but the same diagnostic logic applies. To make the comparisons transparent, every

[3]The linear estimates are close to the canonical wage-process estimates used in related life-cycle work. Kaplan (2012), for example, who estimates a similar process for male hourly wages, reports $\rho = 0.958$, a fixed-effect variance of 0.065, a persistent innovation variance of 0.017, and a transitory variance of 0.081 in his benchmark year-effects specification.

[4]Part of the estimated transitory variation reflects reporting error in PSID wages. Removing an external measurement-error estimate is possible for the Gaussian processes. Since this would treat the four processes asymmetrically, we carry each estimated process into the model as it is estimated. Because transitory shocks are comparatively easy to smooth, this common noise component has little bearing on the welfare comparisons. De Nardi, Fella, and Paz-Pardo (2024), for example, drop their estimated transitory component altogether, noting that it also includes measurement error and that transitory shocks are typically well insured in these models.

TABLE 1. Summary of the estimated wage processes

Process	Avg. persistence	Var(ζ)	Var(ν)	Var(ε)
Linear	0.96	0.046	0.020	0.085
Age-dependent	0.91	0.059	0.033	0.061
Nonlinear	0.92	0.151	0.040	0.027
Non-normal	0.85	0.209	0.030	0.040

NOTES: Average persistence and component variances of the estimated processes, before discretization. Lifetime averages are unweighted means over the working ages. For the linear process, the entries are the estimated parameters (ρ, σ^2_ζ, σ^2_ν, σ^2_ε). For the age-dependent process, persistence and the innovation and transitory variances average the estimated age profiles. For the nonlinear and non-normal processes, persistence is the derivative of the conditional quantile function of the persistent component with respect to its lag, equation (8), averaged over ages, lagged states, and innovation ranks; the innovation variance is the conditional variance of the persistent transition, averaged over ages and lagged states; and the permanent and transitory variances are those of the estimated quantile functions. Table A1 reports the moments of the discretized counterparts used in the quantitative model.

data-versus-model statistic below is computed in the same way on the PSID panel and on a long panel simulated from each estimated process.

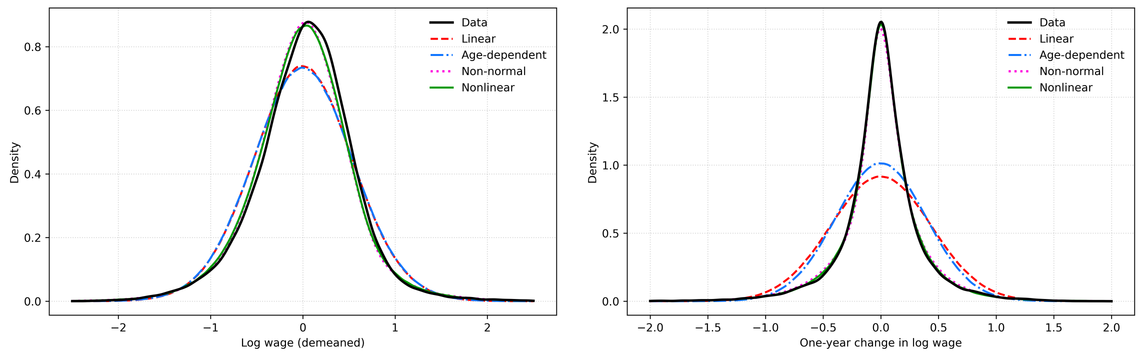
2.5.1. Distributional and Dynamic Fit

Marginal distributions. Figure 2 plots the densities of demeaned residual log wages and of one-year wage changes. In the data, both distributions have a tall peak and fat tails. The nonlinear and non-normal processes track these distributions closely. The linear and age-dependent processes, constrained by normality, miss the peak and place too much mass in the shoulders. The discrepancy is especially visible for one-year changes in panel B, where the Gaussian processes generate a density that is substantially flatter around the mode.

This difference is not merely cosmetic. In a leptokurtic distribution, most observations can be close to zero even when the variance is large, because a disproportionate share of total variation comes from rare, extreme realizations. A Gaussian process can therefore overstate the size of a typical wage change while understating the frequency of very large changes. This distinction provides a first indication of why shock shape can matter for policy: rare, large downward realizations are precisely the states in which a means-tested transfer floor may be most valuable. Their welfare importance, however, also depends on how persistent they are and on which wage component generates them.

Dispersion over the life cycle. Figure 3 plots two measures of wage dispersion by age: the variance of log wages and the quantile-based gap $P_{90} - P_{10}$. Reporting both is informative because normality imposes a fixed relationship between the variance and the distance between the 90th

FIGURE 2. Marginal distributions implied by the estimated processes



A. Log wage levels

B. One-year change in log wage

NOTES: Panel A plots the density of demeaned residual log hourly wages. Panel B plots the density of one-year changes in residual log hourly wages. One-year changes are computed from the annual portion of the sample (1980-1996). The data series is computed from the PSID estimation sample; the model series are computed from long simulated panels generated by each estimated wage process.

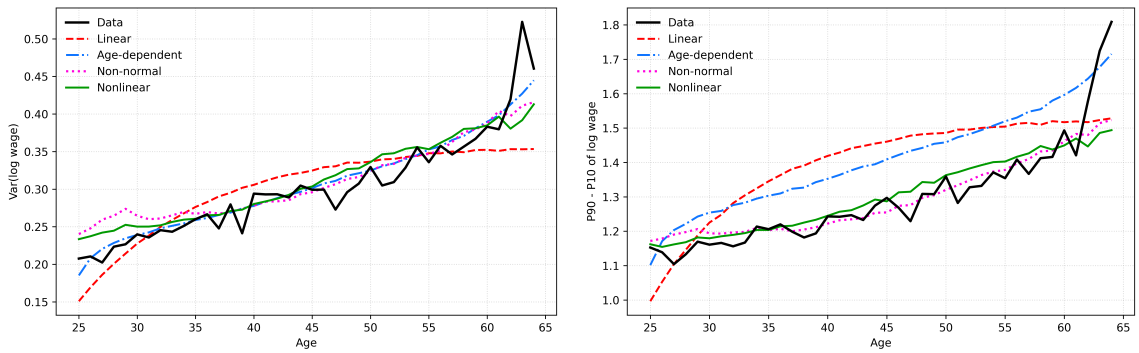
and 10th percentiles, whereas the two measures can move differently under non-normality (Guvenen, Pistaferri, and Violante 2022). Moreover, the variance is especially sensitive to extreme observations, while $P_{90} - P_{10}$ summarizes the width of the central 80 percent of the distribution.

The variance rises steeply over the life cycle in the data, from about 0.21 at age 25 to roughly 0.46 near age 64. The linear process cannot match this profile: its variance is too low early in life and then flattens out after midlife, yielding a concave shape that is at odds with the data. The age-dependent, non-normal, and nonlinear processes all track the rising variance profile more closely. The quantile gap tells a complementary story. The linear and age-dependent processes overstate dispersion in the middle of working life, whereas the nonlinear and non-normal processes remain closer to the data through prime age. Their flexible distributions allow them to match the central quantile spread without imposing the Gaussian relationship between that spread and the variance. No process captures the sharp rise in dispersion observed in the last few years before age 65, when the available data are sparse.

Together, the two panels show that age dependence is important for reproducing the accumulation of wage dispersion over the life cycle, while non-normality is important for matching the distribution of that dispersion between the center and the tails. This distinction can matter for tax and transfer policies because wage movements occurring at different ages have different remaining horizons over which they affect lifetime resources.

Conditional moments. Figure 4 reports quantile-based conditional moments of residual log wages: the standard deviation, the Kelley skewness, and the Crow-Siddiqui kurtosis of the residual wage conditional on the percentile of the lagged residual wage, computed from

FIGURE 3. Wage dispersion over the life cycle



A. Variance of log wages by age

B. $P_{90} - P_{10}$ of log wages by age

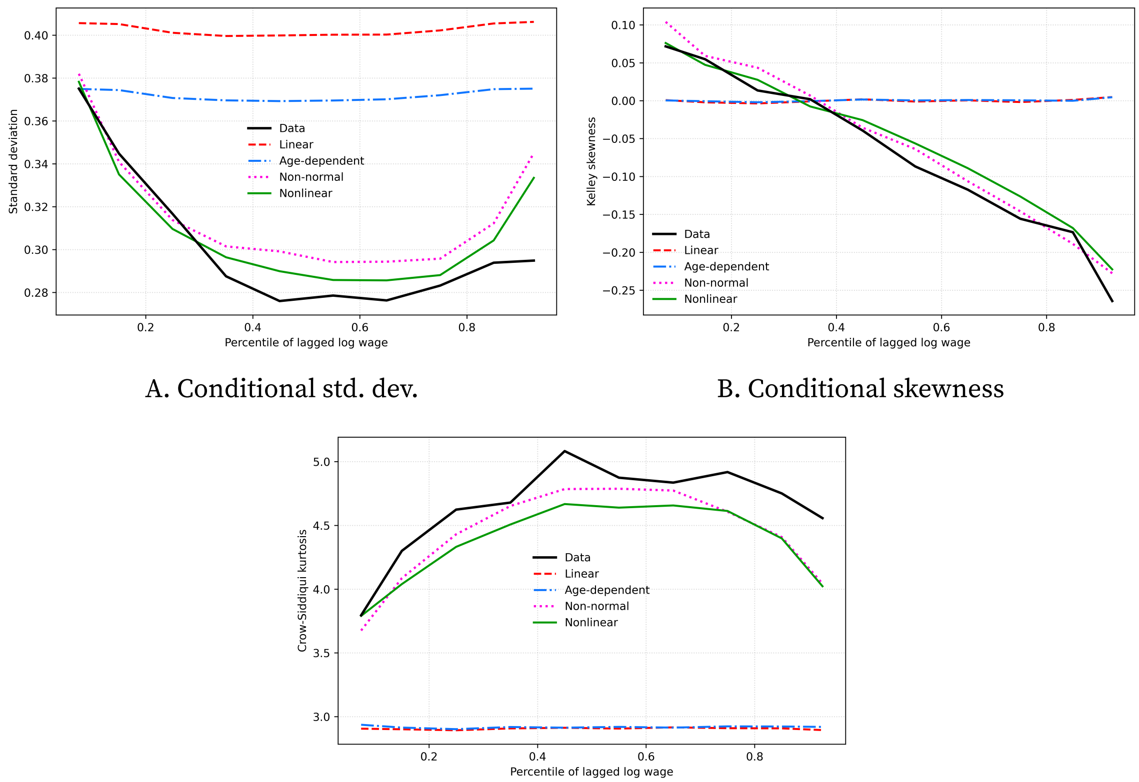
NOTES: Panel A reports the age profile of the variance of residual log-hourly wages. Panel B reports the age profile of the difference between the 90th and 10th percentiles of residual log-hourly wages. The data series is computed from the PSID estimation sample; the model series are computed from long simulated panels generated by each estimated wage process.

consecutive annual observations. The same statistics are computed for the PSID panel and for the simulated panel from each estimated process. Three patterns stand out. Conditional dispersion declines in the lagged percentile, with a modest uptick at the top; conditional skewness declines monotonically with the lagged percentile and turns increasingly negative for workers with high lagged wages; and conditional kurtosis is hump-shaped and well above the Gaussian benchmark (for these quantile-based measures, the Gaussian benchmarks are 0 for the Kelley skewness and about 2.91 for the Crow-Siddiqui kurtosis). These patterns echo the broader evidence that dispersion, asymmetry, and tail risk vary systematically with a worker's position in the earnings distribution (Guvenen et al. 2021; Guvenen, Pistaferri, and Violante 2022).

The nonlinear and non-normal processes capture these patterns, with the negative skewness gradient and excess kurtosis reflecting their flexible, non-Gaussian transition distributions. The linear and age-dependent processes, whose wages are Gaussian by construction, deliver flat conditional skewness at zero and flat quantile kurtosis at the Gaussian value. The negative skewness gradient also indicates that downside risk is not uniformly distributed across current states: workers near the top of the wage distribution face a more left-skewed next-period distribution. Gaussian processes cannot reproduce this state dependence in the likelihood and severity of adverse transitions.

Figure 4 therefore establishes the importance of non-normality, but it does not by itself distinguish sharply between the nonlinear and non-normal processes, both of which fit the conditional higher moments of wages well. The key distinction between them is how the effect of the lagged persistent state varies across current states and innovations, which we examine

FIGURE 4. Conditional moments of log wages given the lagged wage percentile



A. Conditional std. dev.

B. Conditional skewness

C. Conditional kurtosis

NOTES: The figure reports quantile-based conditional moments of residual log wages w_{ij} given the percentile of $w_{i,j-1}$, computed from consecutive annual observations: the conditional standard deviation (panel A), the Kelley skewness (panel B), and the Crow-Siddiqui kurtosis (panel C). The same statistics are applied to the PSID panel and to simulated panels from each estimated process.

next.

Nonlinear persistence and its life-cycle profile. The defining feature of the nonlinear process is that the persistence of the persistent component is not constant. We measure local persistence as the derivative of the conditional quantile function with respect to the lagged persistent state, given by equation (8), which records how strongly last period's position carries over, evaluated at innovation percentile u^{η}_{ij}. Figure 5 plots this object for the benchmark nonlinear process. Panel A averages the surface over ages and plots it against the percentiles of $\eta_{i,j-1}$ and of the innovation. The surface displays the state-dependent resetting documented by Arellano, Blundell, and Bonhomme (2017): persistence is high across most of the state space, but falls sharply when a worker with a low lagged persistent state draws an unusually favorable innovation. In that region the past matters less; the favorable draw partly wipes out a history of bad realizations and pulls the worker back toward the middle of the distribution. Panels B-D show that the resetting pattern shifts over the life cycle. Through the first half of the working life, resetting operates almost exclusively through this upward channel: favorable draws undo low persistent states, while high states survive adverse draws largely intact. Toward the end of the working life, the adverse channel appears as well: an unusually bad draw now also erases much of a high persistent state, while low states hit by further bad draws become more persistent.

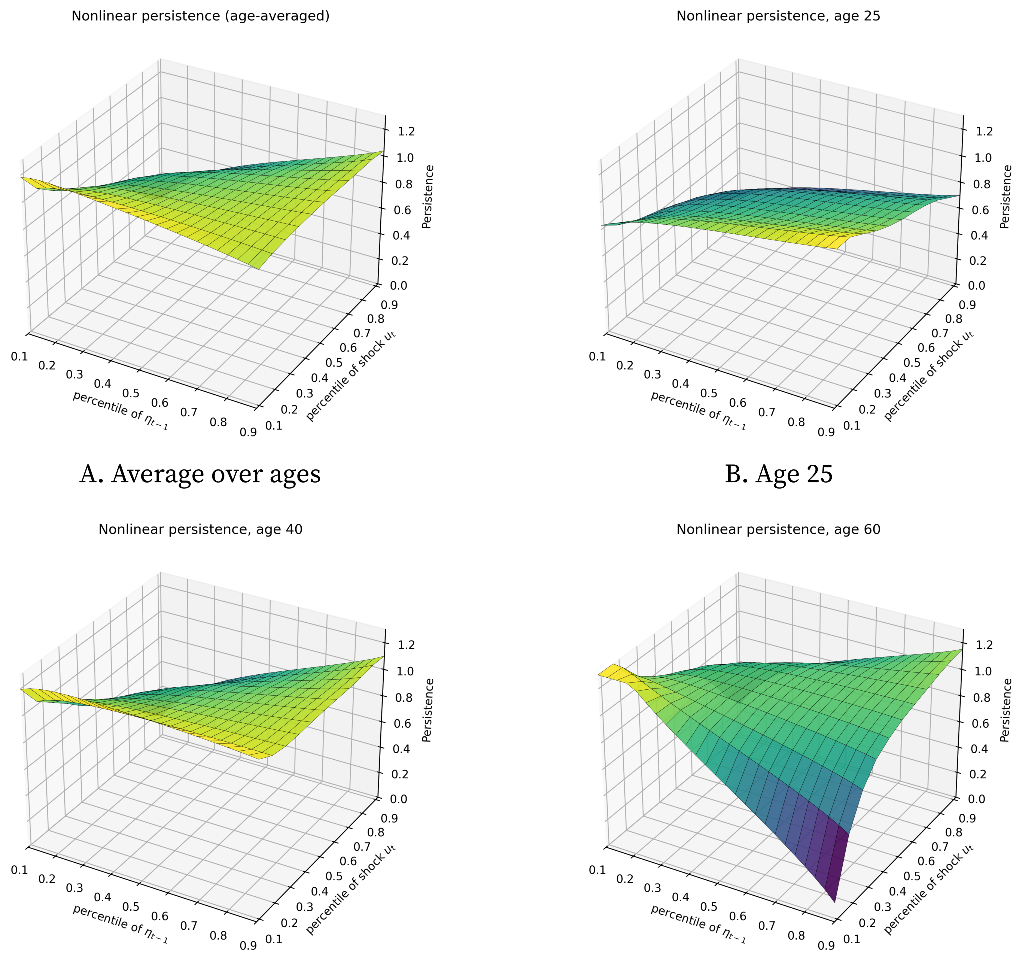
This state-dependent resetting is the principal feature that separates the nonlinear process from the non-normal process. The latter can reproduce skewness, fat tails, and quantile-specific slopes, but its slope with respect to the lagged state does not vary with the level of that state. In the nonlinear process, by contrast, the duration of an unusual realization depends on where the worker started and on the innovation received. The wage process, therefore, partly insures itself: extreme movements have smaller effects on lifetime resources because they are followed by stronger mean reversion. As a result, public insurance can be less valuable under nonlinear persistence than under a process in which the same movements remain uniformly persistent.

The average level of persistence also varies over the life cycle. Figure 6 plots average persistence by age for all four processes. For the nonlinear process, this is $\mathbb{E}[\rho(\eta_{i,j-1}, u^{\eta}_{ij}, j)]$; for the linear process, it is the constant AR(1) coefficient ρ; for the age-dependent process, it is ρ_j; and for the non-normal process, it is the average of $b_1(u^{\eta}_{ij})$ across innovation quantiles. The lifetime averages of these profiles are the persistence entries of Table 1.

The linear process is constant at about 0.96 by construction. The non-normal process has an age-invariant average persistence of about 0.85, although its local slope can vary with the innovation quantile. Only the age-dependent and nonlinear processes display age-varying persistence. Both start low, around 0.71 and 0.78 at age 25, and rise steeply through the first part of the working life; the nonlinear process peaks near 0.97 around age 50 before easing back. This low early-life persistence is a feature that processes with age-invariant average persistence cannot reproduce, and it is consistent with the strong age dependence documented by Karahan and Ozkan (2013). In particular, the linear process treats early-career wage movements as substantially more durable. All else equal, doing so increases their implied effect on lifetime resources and can therefore raise the measured value of public insurance.

Taken together, Figures 2-6 point to two features of shock shape that can have opposing implications for social insurance. Non-normality can lead to rare and potentially severe adverse realizations, which can increase the value of a transfer floor. Nonlinear mean reversion limits the duration of some unusual realizations, reducing their effect on lifetime resources and therefore the need for public insurance. These diagnostics describe the shape and propagation of wage risk, but they do not yet reveal whether the four processes carry the same total amount of wage risk or allocate that risk similarly across components. We examine those differences

FIGURE 5. Nonlinear persistence of the persistent component



A. Average over ages

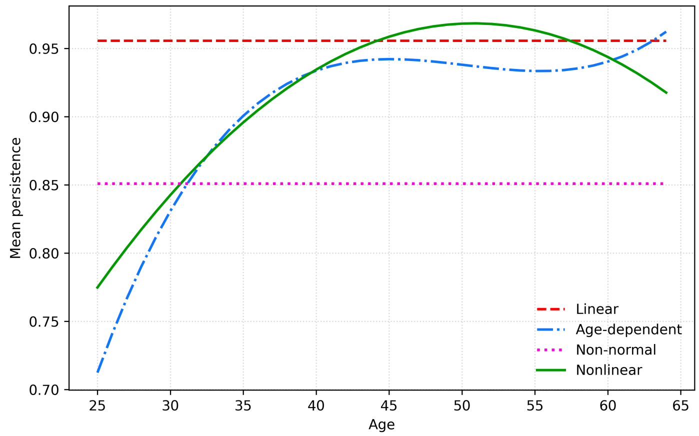
B. Age 25

C. Age 40

D. Age 60

NOTES: The figure plots local persistence for the nonlinear process, as defined in equation (8). Panel A averages local persistence over ages. Panels B–D report the same object at ages 25, 40, and 60. The horizontal axes are the percentiles of the lagged persistent state and the current innovation. Following Arellano, Blundell, and Bonhomme (2017), the lagged persistent state is evaluated at interior percentiles (10-90) and the innovation axis is the estimated quantile grid; the panels share a common vertical scale, at which the age-60 surface is truncated (it peaks at about 1.5 where a low persistent state meets an unusually adverse innovation). Local persistence can exceed one at some states while the process remains mean-reverting on average (Figure 6).

FIGURE 6. Average persistence by age



NOTES: The figure reports average persistence by age for each estimated process. For the nonlinear process, persistence is the average local derivative in equation (8). For the linear process, the AR(1) coefficient is ρ. For the age-dependent process, the age-specific coefficient is ρ_j. For the non-normal process, it is the average slope $b_1(u_{ij}^{\eta})$ across innovation quantiles.

next.

2.5.2. Variance Decomposition

The diagnostics above describe the shape and propagation of wage risk. We next ask how much risk each process carries and where that risk is allocated across the permanent, persistent, and transitory components. Table 2 reports lifetime-average variance shares and the total variance of log wages.[5]

The main contrast is between the Gaussian and quantile-based processes. In the linear and age-dependent processes, the permanent component accounts for only 15–19 percent of total variance, while the transitory component accounts for 20–28 percent. In the nonlinear and non-normal processes, the permanent share is close to one-half, and the transitory share is only 9–10 percent. The total amount of variation also differs: the non-normal process carries noticeably more total variance, 0.39, than the other three processes, whose totals are all close to 0.30.

These differences matter because the components map into different economic mechanisms. The permanent component is drawn before working life and remains fixed, so its dispersion primarily governs redistribution across workers with different permanent earning capacities.

[5] Because the initial persistent state may depend on the permanent type, equation (5), the permanent and persistent components can be correlated. However, the covariance is negligible and is omitted from the table (e.g., the average covariance for the nonlinear process is −0.0002).

TABLE 2. Variance decomposition of the estimated processes

Process	Total var.	s_ζ	s_η	s_ε
Linear	0.30	0.15	0.56	0.28
Age-dependent	0.30	0.19	0.60	0.20
Nonlinear	0.31	0.49	0.42	0.09
Non-normal	0.39	0.54	0.36	0.10

NOTES: The table reports lifetime-average variance shares of the permanent (ζ), persistent (η), and transitory (ε) components, together with the total variance of log wages, computed from the estimated processes before discretization. The η entry is based on the variance of the persistent *state* at each age, which accumulates past innovations, not on the innovation variance $\mathrm{Var}(\nu)$ of Table 1; multiplying the total by the ζ and ε shares reproduces the corresponding variances in Table 1. Lifetime averages are unweighted means over working ages. Shares may not sum exactly to one because of rounding. Table A1 in Appendix D reports the moments of the discretized counterparts used in the quantitative model.

The persistent component evolves during working life but can affect wages for many years, making it difficult to self-insure. The transitory component is short-lived and is therefore comparatively easier to smooth through saving and labor supply. Moving wage variation away from the transitory component and toward permanent and persistent components therefore strengthens, respectively, the redistributive and insurance roles of the tax-transfer system.

3. Life-Cycle Model, Calibration, and Validation

The model period is one year. Agents choose consumption, hours worked, and saving in a single risk-free asset, and discount the future at rate β. They are heterogeneous ex ante in their permanent type and ex post in the persistent and transitory shocks they experience over the life cycle. A government levies taxes and pays transfers and Social Security benefits; a representative firm sets factor prices.

Demographics. A continuum of agents enters the economy each period at age $j = 25$. Let s_{j+1} denote the probability that an agent alive at age j survives to age $j + 1$. Agents live at most until age $J = 99$. Population grows at the constant rate n, so the stationary age distribution μ_j satisfies

$$\mu_{j+1} = \frac{s_{j+1}}{1+n}\mu_j.$$

Agents work until the mandatory retirement age $j_R = 64$ and are retired thereafter, from age 65 on.

Preferences. Agents have time-separable preferences over consumption c and hours worked l. Period utility is

$$u(c, l) = \log c - \psi \frac{l^{1+1/\gamma}}{1 + 1/\gamma},$$

where ψ scales the disutility of work, and γ is the Frisch elasticity of labor supply. During retirement, $l = 0$.

Technology. A representative firm produces output with a constant-returns Cobb-Douglas technology,

$$Y_t = AK_t^{\alpha}(L_t X_t)^{1-\alpha},$$

where K_t is capital, L_t is aggregate labor in efficiency units, A is total factor productivity, and X_t is labor-augmenting technology. The technology level evolves according to

$$X_{t+1} = (1 + g)X_t.$$

For any current-period variable q_t that grows with X_t, let $\widehat{q}_t \equiv q_t/X_t$. Next-period assets are normalized by next period's technology level, $\widehat{a}_{t+1} \equiv a_{t+1}/X_{t+1}$. This normalization applies to consumption, assets, wages, earnings, taxable income, taxes, transfers, Social Security benefits, and government expenditure, as well as to per-capita output and capital. In stationary units, the production function is

$$\widehat{Y}_t = A\widehat{K}_t^{\alpha} L_t^{1-\alpha}.$$

Henceforth, we suppress hats and use the same symbols for variables expressed in stationary units. Capital depreciates at rate δ. Let w denote the wage per efficiency unit and r the net rental rate of capital in stationary units. Factor prices equal marginal products.

Labor productivity. An age-j agent with current wage state $\Omega_j = (\zeta, \eta_j, \varepsilon_j)$ supplies efficiency units

$$e_j(\Omega_j) = \exp\{\kappa_j + \zeta + \eta_j + \varepsilon_j\}.$$

The components $(\zeta, \eta_j, \varepsilon_j)$ follow one of the four wage processes estimated in Section 2.

We solve and calibrate the model separately under each estimated wage process, holding the rest of the environment fixed. Differences in outcomes across economies are therefore attributable to the wage process. In the computation, each process is represented by discretized grids for the wage components and age-specific transition matrices for the persistent component.

Government. Each period, the government consumes G units of goods and runs three programs. First, federal income taxes follow the Heathcote, Storesletten, and Violante (2017) schedule,

$$T_f(I) = I\left[1 - \tau_0 \,(I/\bar{I})^{-\tau_f}\right], \tag{18}$$

where I is taxable income, $\bar{I}$ is mean income, τ_f governs progressivity, and τ_0 sets the tax level.[6] In addition, agents pay a flat local income tax at rate τ_l and a flat capital-income tax at rate τ_k. As in Guner, Lopez-Daneri, and Ventura (2016), taxable income includes labor and capital income, while the capital-income tax term is added separately.

Second, following Guner, Rauh, and Ventura (2026), means-tested transfers map relative income $\tilde{I} = I/\bar{I}$ into a benefit $TR(I) = \bar{I} \cdot tr(\tilde{I})$, with

$$tr(\tilde{I}) = \begin{cases} \exp(b_1 + b_2\tilde{I})\,\tilde{I}^{\,b_3}, & \text{if } \tilde{I} > 0, \\ b_0, & \text{if } \tilde{I} = 0. \end{cases} \tag{19}$$

Income and capital taxes, together with fully taxed accidental bequests, finance transfers and government expenditure.

Third, the government runs a pay-as-you-go Social Security system. Working agents pay a payroll tax at rate τ_{ss} on labor earnings up to a taxable maximum $\bar{y}^{ss}$:

$$T^{ss}(y) = \tau_{ss} \min\{y, \bar{y}^{ss}\}. \tag{20}$$

Retirees receive benefits linked to labor-market histories. Following Kindermann and Krueger (2022), we avoid adding average lifetime earnings as an additional continuous state variable. Instead, retirees are assigned to cells indexed by their permanent type and their last working-year persistent state, (ζ, η_R). For each cell, the model computes the average across cell members of career-average capped earnings (each agent's mean of $\min\{y_j, \bar{y}^{ss}\}$ over the working years, the model's analog of average indexed lifetime earnings) and a benefit function $SS(\zeta, \eta_R)$ maps this base to retirement income.

Household problem. Let $\mathbf{s} = (a, \Omega)$ denote the current state of a working-age agent, where a is assets and $\Omega = (\zeta, \eta, \varepsilon)$ is the current wage state. Let $\mathbf{s}' = (a', \Omega')$ denote next period's state. All variables are in stationary units. Given $\mathbf{s}$ and age $j \le j_R$, the agent solves

$$V(\mathbf{s}, j) = \max_{c,\, a' \ge 0,\, l \in [0,1]} \left\{ \log c - \psi \frac{l^{1+1/\gamma}}{1 + 1/\gamma} + \beta s_{j+1} \mathbb{E}\left[V(\mathbf{s}', j+1) \mid \mathbf{s}\right] \right\}$$

[6] Mean income $\bar{I}$ is the economy-wide average of pre-tax household income; its exact definition in the model is given in Appendix C.

subject to

$$c + (1+g)a' \leq (1+r)a + y - T_f(I) - \tau_l I - \tau_k ra - T^{ss}(y) + TR(I),$$

where

$$y = we_j(\Omega)l, \qquad I = y + ra.$$

Agents enter the economy with zero assets and face the no-borrowing constraint $a' \geq 0$.

Retirees do not work and pay no payroll tax. A retired agent with state (a, ζ, η_R) solves the analogous problem with $l = 0$ and Social Security benefits replacing labor income. The retirement budget constraint is

$$c + (1+g)a' \leq (1+r)a + SS(\zeta, \eta_R) - T_f(I^R) - \tau_l I^R - \tau_k ra + TR(I^R),$$

where

$$I^R = SS(\zeta, \eta_R) + ra.$$

The persistent state η_R is the wage state from the last working year and is retained only to determine Social Security benefits during retirement. A formal definition of a stationary recursive equilibrium is presented in Appendix C.

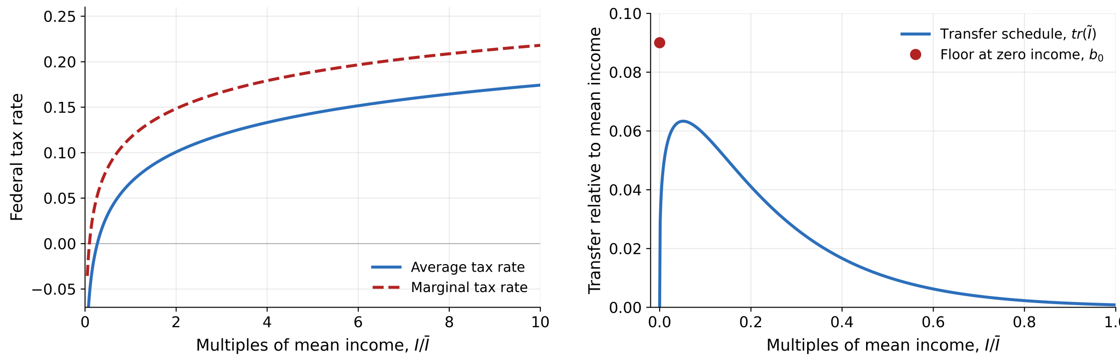
3.1. Calibration and Policy Environment

We calibrate the benchmark economy to U.S. data for the 1980-2019 period, the same window used to estimate the wage processes. The baseline policy environment combines the progressive federal tax of equation (18), the means-tested transfers of equation (19), and the Social Security system described in Section 3. We solve and recalibrate the economy separately for each estimated wage process. Table 3 presents the parameters, organized into three groups: those set outside the model, those calibrated to match aggregate or budget targets, and those estimated using the simulated method of moments (SMM). Appendix D provides details on data sources for different targets. Appendix D, Table A1, reports the corresponding moments of the discretized wage-component grids that are fed into the quantitative model.

The first group is set from the data or the literature. Population and labor-augmenting productivity grow at the average U.S. rates over 1980-2019, $n = 0.94\%$ and $g = 1.81\%$. The Frisch elasticity is set to $\gamma = 1$, in line with the macro labor-supply literature (Chetty et al. 2011; Keane and Rogerson 2015). The capital share is $\alpha = 0.391$. The local income tax is flat at $\tau_l = 5\%$, and the progressivity parameter is $\tau_f = 0.053$, both from Guner, Kaygusuz, and Ventura (2014).[7]

[7] Guner, Kaygusuz, and Ventura (2014) estimate federal income-tax functions using IRS tax-return micro data for 2010. Their estimate of $\tau_f = 0.053$ is close to estimates based on TAXSIM for the post-2010 period, around 0.06-0.07, by Borella et al. (2023). By contrast, Heathcote, Storesletten, and Violante (2017) estimate a larger progressivity

FIGURE 7. Benchmark policy schedules



A. Federal income tax rates

B. Means-tested transfer schedule

NOTES: Panel A plots the average and marginal federal income-tax rates implied by the schedule in equation (18) at the benchmark calibration ($\tau_f = 0.053$). Panel B plots the means-tested transfer schedule in equation (19), expressed relative to mean income. The red point is the transfer floor at zero income, b_0; the blue curve is the positive-income transfer formula. Social Security is not plotted because its institutional parameters are held fixed in the counterfactuals.

The four transfer parameters are taken from Guner, Rauh, and Ventura (2026).

Figure 7 plots the federal tax schedule and the means-tested transfer schedule at the benchmark calibration. Panel A shows that the federal income-tax schedule is progressive but smooth. At mean income, the average federal tax rate is $1 - \tau_0$, which is about 7 percent in the benchmark calibrations; at twice mean income, the average federal tax rate is about 10 percent. Panel B shows that means-tested transfers are concentrated at the bottom of the income distribution. The zero-income transfer is $b_0 = 0.09$ of mean income, or about \$10,800 when mean income is \$120,000 (roughly the US mean household income in 2024). For households with positive income, the transfer schedule rises briefly at very low income levels and then declines rapidly: it peaks at about 6.3 percent of mean income, roughly \$7,600, when income is around 5 percent of mean income. By mean income, transfers are essentially zero.

Social Security is also largely set by institutional parameters. Workers pay the OASI payroll tax rate, $\tau_{ss} = 10.6\%$, on labor earnings up to the taxable maximum, $\bar{y}^{ss} = 2.42\,\bar{y}$, where $\bar{y}$ denotes average labor earnings. Benefits are computed using the Primary Insurance Amount formula of the U.S. Social Security system, with replacement rates of 90%, 32%, and 15%, and bend points equal to 0.21 and 1.29 times average labor earnings. Benefit cells are indexed by a retiree's permanent type and last working-year persistent state rather than by individual lifetime earnings. Specifically, for each retirement cell (ζ, η_R), we compute the cell average of career-average labor earnings, capped period by period at the taxable maximum, and apply the PIA formula to that amount; the resulting benefit is received in retirement by all agents in that

parameter, around 0.18, because their tax function combines taxes and transfers.

cell. Social Security benefits are then multiplied by a scale factor, denoted s^{ss}, that balances the PAYGO pension budget. The payroll tax rate and benefit formula are therefore fixed, while the average replacement rate is an equilibrium outcome. In the benchmark computations, the benefit scale is 0.93-0.95 across the four economies, and the replacement rate at mean earnings is about 44%.

The second group is calibrated so the model reproduces aggregate ratios and revenue targets. The depreciation rate $\delta = 0.031$ matches an investment-to-output ratio of 17.70%. The capital-income tax rate $\tau_k = 0.061$ matches corporate tax collections of 1.8% of GDP. The level parameter τ_0 is set so that revenues from the federal income tax, local income tax, and capital-income tax equal 14.51% of GDP. Payroll taxes and Social Security benefits are excluded from this target and are balanced in the separate Social Security budget.

The third group is estimated by SMM *separately for each wage process*. For each process, we choose the discount factor β and the disutility of work ψ so that the model matches the same two targets: a capital-to-output ratio of 3.01 and mean working-age hours of 1/3. Because every baseline economy matches the same aggregate capital and hours targets, the welfare comparisons are not driven by differences in baseline aggregate calibration.

3.2. Untargeted Model Validation

Each economy is calibrated to match only two aggregate targets, mean hours and the capital-output ratio, so the cross-sectional distributions of consumption and hours are untargeted. We ask which wage process, embedded in the same model, comes closest to these patterns in the PSID.

We focus on three moments, all directly related to insurance. First, we study consumption inequality over the life cycle, $\mathrm{Var}(\log c)$ by age. If wage shocks are poorly insured, consumption dispersion should fan out with age; if households smooth these shocks effectively, the profile should be relatively flat. Second, we measure the pass-through of persistent wage shocks to consumption, ϕ_{pers}. This statistic is especially informative for the value of tax and transfer programs because persistent shocks are harder to self-insure than transitory shocks: they alter lifetime resources and therefore tend to pass more strongly into consumption. We measure ϕ_{pers} using the Blundell, Pistaferri, and Saporta-Eksten (2016) implementation of the Blundell, Pistaferri, and Preston (2008) identification, in which a three-period window of wage changes isolates the lasting component of wage innovations. The implied insurance coefficient is $1-\phi_{\text{pers}}$. Third, we study the dispersion of hours, $\mathrm{Var}(\log h)$, by age, since labor supply is itself an insurance margin.

TABLE 3. Parameter values (benchmark economy)

Parameter	Symbol	Value	Source / target
Set externally			
Life cycle (ages)	$j_R,\ J$	work 25–64, retire 65–99	Model demographics
Survival probabilities	s_j	—	U.S. Life Tables 2005
Population growth	n	0.0094	U.S. data, 1980–2019
Labor-augmenting growth	g	0.0181	U.S. data, 1980–2019
Frisch elasticity	γ	1	Literature
Capital share	α	0.391	FRED
Local income tax	τ_l	0.05	Guner, Kaygusuz, and Ventura (2014)
Progressivity	τ_f	0.053	Guner, Kaygusuz, and Ventura (2014)
Transfers	b_0, b_1, b_2, b_3	0.090, −1.634, −5.504, 0.284	Guner, Rauh, and Ventura (2026)
Payroll tax rate	τ_{ss}	0.106	U.S. OASI rate
Taxable maximum	$\bar{y}^{ss}/\bar{y}$	2.42	U.S. taxable maximum
PIA bend points	—	0.21, 1.29	U.S. Social Security formula
PIA replacement rates	—	0.90, 0.32, 0.15	U.S. Social Security formula
Calibrated to aggregate or budget targets			
Depreciation rate	δ	0.031	I/Y = 17.70%
Capital-income tax	τ_k	0.061	Corporate revenue = 1.8% of GDP
HSV tax level	τ_0	0.936, 0.935, 0.935, 0.933	Non-payroll tax revenue = 14.51% of GDP
Social Security scale	s^{ss}	0.930, 0.931, 0.955, 0.954	PAYGO Social Security budget
Structurally estimated by SMM			
Linear process	(β, ψ)	(0.958, 7.22)	K/Y = 3.01, mean hours = 1/3
Age-dependent process	(β, ψ)	(0.958, 7.24)	K/Y = 3.01, mean hours = 1/3
Non-normal process	(β, ψ)	(0.958, 7.33)	K/Y = 3.01, mean hours = 1/3
Nonlinear process	(β, ψ)	(0.959, 7.47)	K/Y = 3.01, mean hours = 1/3

NOTES: Parameters of the benchmark economy. The first block is set outside the model; the second is calibrated to aggregate and budget targets; and the third is estimated by SMM separately for each wage process, so that every economy matches the same two targets. For the level τ_0 and the Social Security scale s^{ss}, the four reported values correspond, respectively, to the linear, age-dependent, non-normal, and nonlinear economies. Note that $1 - \tau_0$ is the average federal tax rate paid by a household with mean income. TFP is normalized to one. Appendix D details the data sources and target construction.

TABLE 4. Persistent pass-through ϕ_{pers}: data vs. models

	Pooled	25–39	40–54	55–64
PSID data	0.257	0.290	0.326	0.108
Linear	0.558	0.719	0.535	0.388
Age-dependent	0.499	0.617	0.485	0.383
Non-normal	0.405	0.538	0.364	0.234
Nonlinear	0.342	0.468	0.304	0.218

NOTES: Pass-through of a persistent wage shock to consumption; lower values imply more insurance. The implied insurance coefficient is $1 - \phi_{pers}$. The same Blundell, Pistaferri, and Preston (2008) estimator is applied to the PSID and to each model's simulated panel.

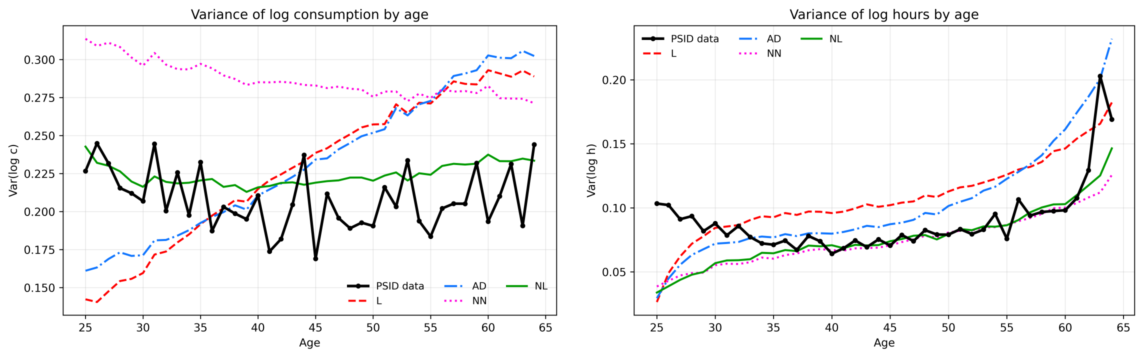
Measurement error. Simulated consumption is generated by the model and is free of any measurement error, while PSID consumption data is not. We therefore add measurement error to simulated consumption using the estimate implied by the short-run reversal in consumption growth in the data, yielding Var(noise) ≈ 0.062. As Blundell, Pistaferri, and Saporta-Eksten (2016) emphasize, this consumption-noise estimate is an upper bound. It is therefore the most favorable correction for matching the level of consumption dispersion in the data. Appendix E reports the sensitivity of the validation results to this correction.

Persistent pass-through. Table 4 reports ϕ_{pers} in the data and in each model, pooled and by age group. The processes line up monotonically: the linear process passes the most persistent risk into consumption (0.558), the nonlinear process passes the least (0.342), and the data value is lower still (0.257). The nonlinear process is closest to the data across all age groups, not only on average. This comparison is important for the value of social insurance. In the nonlinear process, persistent-looking shocks are partly mean-reverting, so less wage risk reaches consumption. The simpler processes, especially the linear one, overstate the extent to which persistent wage risk passes into consumption and therefore tend to overstate the value of public insurance.[8]

Consumption and hours inequality. Figure 8 plots the variance of log consumption and log hours by age. The robust feature in panel A is the shape of consumption inequality. In the data, consumption dispersion is essentially flat over the working life, around 0.20. The linear and age-dependent processes generate a much steeper increase, from roughly 0.15 to 0.30. By contrast, the nonlinear and non-normal processes deliver flat or gently declining profiles. The level is more sensitive to the consumption measurement-error correction: under the

[8] Because the Blundell, Pistaferri, and Preston (2008) method was developed for a canonical permanent-transitory structure, ϕ_{pers} should not be interpreted as the literal structural pass-through of a primitive permanent shock when the wage process is mean-reverting or nonlinear. The key point is that the same empirical lens is applied to the data and to every simulated panel.

FIGURE 8. Consumption and hours inequality over the life cycle



A. Variance of log consumption by age B. Variance of log hours by age

NOTES: Panel A reports the age profile of the variance of log consumption. Panel B reports the age profile of the variance of log hours. The data series are computed from the PSID validation sample; the model series are computed from simulated panels generated by each calibrated economy. Panel A includes the upper-bound consumption measurement-error correction described in the text.

upper-bound correction described above, the nonlinear process also closely matches the level, whereas the non-normal process lies above the data because it assigns more risk to the permanent component.

Panel B shows hours dispersion. The nonlinear and non-normal processes track the data more closely across prime working ages, while the linear process produces excessive dispersion in hours over much of the life cycle. None of the four economies reproduces the sharp decline in hours inequality over the first fifteen working years. As Kaplan (2012) shows, this is a known limitation of frictionless labor-supply models: the intratemporal first-order condition mechanically links hours dispersion to wage dispersion and consumption dispersion. We therefore view this early-life decline as a limitation of the model class rather than as evidence against a specific wage process.

These validation results are consistent with the broader message of De Nardi, Fella, and Paz-Pardo (2020): allowing richer earnings dynamics improves the model's ability to match life-cycle consumption inequality and the pass-through of persistent shocks. They also connect to Guvenen, Ozkan, and Madera (2024), who show that non-Gaussian earnings risk can substantially change measured consumption insurance and welfare relative to a canonical Gaussian process. Our exercise separates these forces in a tax-transfer model with endogenous labor supply: nonlinear mean reversion lowers pass-through and the value of public insurance, while non-normal tail risk raises the value of transfers by creating rare but severe bad states.

Appendix E reports a compact distance-based scorecard and the sensitivity of the consumption-inequality comparison to the consumption measurement-error correction. The nonlinear process ranks first overall and is closest to the data on persistent-shock pass-through; the ranking for the level of consumption dispersion is more sensitive to the consumption-noise adjustment.

4. The Value of Social Insurance

We now ask how much the choice of wage process matters for the welfare evaluation of tax-and-transfer policy. Starting from the baseline economy, we consider three revenue-neutral counterfactuals. The first replaces the progressive federal income tax with a proportional federal tax while keeping means-tested transfers in place (*PT*, $\tau_f = 0$). The second removes means-tested transfers while keeping the progressive federal income tax (*NT*). The third combines the two reforms: it removes federal tax progressivity and means-tested transfers while keeping Social Security in place (*NW*). In each counterfactual, the level parameter τ_0 adjusts so that total revenue from the federal income tax, the local income tax, and the capital-income tax is unchanged from the baseline. Changes in transfer outlays are not rebated back to households. The Social Security tax rate, cap, bend points, and replacement rates are held fixed. The benefit scale adjusts so that the PAYGO pension budget balances separately.

For each wage process and each counterfactual, we compute the consumption-equivalent variation (CEV): the percentage change in lifetime consumption in the baseline economy that makes an agent indifferent between the baseline and the counterfactual. Positive values therefore indicate that the counterfactual raises welfare. Shock histories are held fixed between the baseline and the counterfactual, so each agent faces the same idiosyncratic productivity history in both economies. The CEV is therefore a within-person comparison of policies rather than a comparison of different shock realizations.

Mean welfare effects. Table 5 summarizes the distribution of welfare effects through the mean CEV, the median CEV, and the share of agents who gain. The qualitative pattern is the same across wage processes. Replacing the progressive federal tax with a proportional one raises mean welfare in all four economies. This reform reduces labor-supply distortions while leaving means-tested transfers and Social Security in place. Removing means-tested transfers lowers mean welfare in all four economies because the loss of insurance dominates any efficiency gain from the associated tax adjustment. Combining the two reforms leaves the efficiency gain from lower progressivity partly offset by the loss of transfers: the net effect is positive under the linear, age-dependent, and nonlinear processes, but slightly negative under the non-normal process.

The distributional statistics in Table 5 are also informative. The PT reform has broad support: the median CEV is positive in every economy, and at least 65% of agents gain. The NT reform has negative mean effects, but the median effects are close to zero; removing transfers hurts the

TABLE 5. Welfare effects of the counterfactuals: distribution of gains (CEV, % of lifetime consumption)

	Mean CEV			Median CEV			Share with CEV>0		
Process	PT	NT	NW	PT	NT	NW	PT	NT	NW
Nonlinear	2.01	−1.01	0.72	1.71	−0.13	1.37	76%	47%	62%
Linear	1.91	−0.94	0.63	1.74	−0.36	1.09	77%	39%	61%
Age-dependent	1.84	−1.02	0.49	1.69	−0.35	1.02	76%	39%	59%
Non-normal	1.61	−1.74	−0.47	1.10	−0.29	0.50	65%	44%	53%

NOTES: Mean CEV, median CEV, and the share of agents with CEV > 0, for each counterfactual. PT replaces the progressive federal income tax with a proportional federal tax while keeping means-tested transfers. NT removes means-tested transfers while keeping federal progressivity. NW removes both federal progressivity and means-tested transfers, while Social Security remains in place.

households most exposed to bad wage histories, while some higher-income households gain from the tax adjustment. The NW reform has a positive median CEV in all four economies, but the mean becomes slightly negative under the non-normal process. This divergence between the median and the mean is a first indication that rare but severe losses are especially important under the non-normal process.

Redistribution and downside risk. Table 6 and Figure 9 describe who gains and who loses. Table 6 reports two distributional objects: the welfare gradient across permanent types and the worst individual loss. Figure 9 plots mean CEVs by permanent type under the NW reform; the corresponding PT and NT profiles are reported in Table A4 in Appendix F. In Figure 9, under NW, all processes generate losses for low permanent types and gains for high permanent types, but the nonlinear and non-normal profiles are much steeper than the linear and age-dependent profiles. The non-normal profile is also shifted downward for much of the distribution and delivers the largest losses at the bottom, which explains why its mean CEV is negative even though high permanent types gain substantially.

The welfare gradient is a redistribution object: it measures the CEV of the highest permanent type minus the CEV of the lowest permanent type. It is positive for every reform and every wage process, which means that high permanent types gain relative to low permanent types when federal progressivity, means-tested transfers, or both are reduced. The gradient is especially steep for the nonlinear and non-normal processes. Under the NW reform, for example, the gradient is 20.67 percentage points under the nonlinear process and 23.66 under the non-normal process, compared with 11.26 under the linear process and 12.89 under the age-dependent process. This pattern mirrors the variance decomposition in Table 2: the nonlinear and non-normal processes assign a much larger role to permanent heterogeneity, and permanent heterogeneity is precisely the component against which households cannot self-insure.

TABLE 6. Welfare effects of the counterfactuals: redistribution and worst losses

	Welfare gradient			Worst CEV		
Process	PT	NT	NW	PT	NT	NW
Nonlinear	11.38	8.11	20.67	−4.1	−21.9	−26.9
Linear	5.74	4.81	11.26	−3.5	−14.4	−19.6
Age-dependent	6.47	5.61	12.89	−3.5	−17.0	−21.8
Non-normal	12.94	9.40	23.66	−4.8	−34.3	−39.8

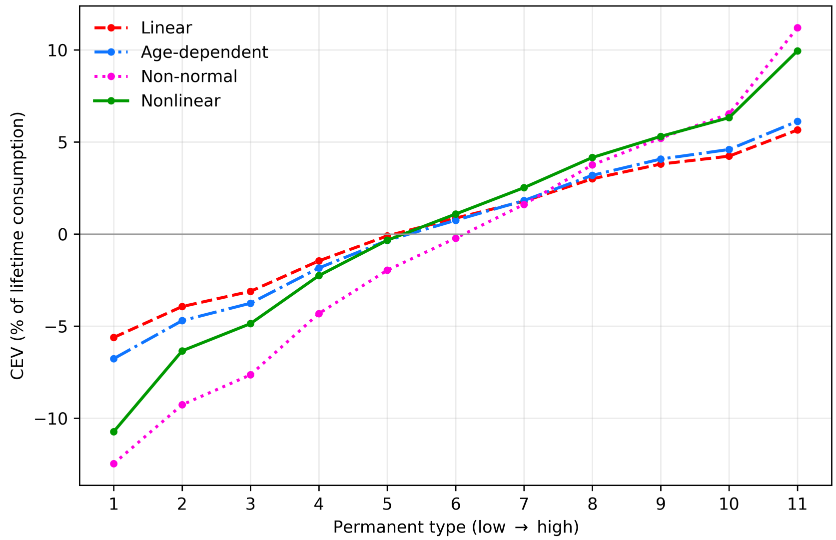
NOTES: The welfare gradient is the CEV of the highest permanent type minus the CEV of the lowest permanent type; the worst CEV is the largest individual loss. Both are reported for each counterfactual. The full breakdown by permanent type is reported in Appendix F, Table A4, and plotted in Figure 9.

Worst-case losses capture a different object: the insurance value of the tax-transfer system for households hit by bad wage histories. The PT reform generates only modest worst losses, between −3.5% and −4.8% of lifetime consumption, because transfers remain in place. By contrast, removing transfers generates large downside losses. The worst loss under NT is −21.9% in the nonlinear economy and −34.3% in the non-normal economy. Under NW, the worst loss reaches −26.9% under the nonlinear process and −39.8% under the non-normal process. Thus, the most severe losses arise when transfers are removed, and they are largest under the process with the fattest tails.

The comparison across wage processes is the central message of these results. The value of reducing tax progressivity is positive in all four economies, but it is not the margin on which the processes differ most. The PT mean CEV ranges from 1.61% under the non-normal process to 2.01% under the nonlinear process. The removal of means-tested transfers generates sharper differences: the mean welfare loss is around 0.9–1.0% under the linear, age-dependent, and nonlinear processes, but 1.74% under the non-normal process. The combined reform is therefore the clearest summary of the competing forces. Under the nonlinear process, NW raises mean welfare by 0.72% of lifetime consumption. Under the non-normal process, the same reform lowers mean welfare by 0.47%. The same tax-transfer reform can therefore change sign when the wage process changes.

These differences are not driven by different aggregate calibration targets: all economies match the same baseline capital-output ratio and average hours. They reflect how each wage process distributes risk across permanent types, persistent shocks, and transitory shocks, and how much of that risk reaches consumption. The next subsection decomposes these channels.

FIGURE 9. Welfare gains by permanent type: removing the welfare state (NW)



NOTES: The figure plots the mean CEV by permanent type under the NW counterfactual (proportional federal tax and no means-tested transfers, with Social Security in place), for each estimated wage process. The eleven permanent types are ordered from lowest to highest and carry population weights 1/21 for types 1–3 and 9–11, and 1/7 for types 4–8. The slope of each line is the welfare gradient reported in Table 6. The corresponding profiles for the PT and NT reforms are reported in Table A4 in Appendix F.

4.1. Amount of Risk versus Shape of Risk

The raw comparison across wage processes combines two forces. First, each process carries a different amount of risk and allocates that risk differently across the permanent, persistent, and transitory components. Second, the processes differ in shape: age dependence, non-normality, and nonlinear persistence. To isolate the second force, we use the nonlinear process as the benchmark and force the other processes to match its component variances. What remains is the effect of shape.

Table 7 reports this decomposition for the nonlinear-minus-non-normal welfare gap. We focus on this comparison because both processes allow rich, non-Gaussian wage risk, but they differ in the key structural feature emphasized in Section 2: the nonlinear process allows persistence to depend on the lagged state, while the non-normal process preserves linear dependence on the lagged persistent component. The first column reports the total gap between the freely estimated nonlinear and non-normal economies. The variance channel reports how much of this gap is closed when the non-normal process is rescaled to have the nonlinear process's component variances. The shape channel is the residual.

The decomposition shows that the amount and allocation of risk are the dominant forces. For the NW reform, the total nonlinear-minus-non-normal gap is 1.19 percentage points, of which

TABLE 7. Decomposing the nonlinear-minus-non-normal welfare gap (mean CEV, pp)

Reform	NN → NL gap	Variance channel	Shape channel
PT	+0.40	+0.43	−0.03
NT	+0.73	+0.51	+0.22
NW	+1.19	+0.99	+0.20

NOTES: The total gap is the mean CEV under the nonlinear process minus the mean CEV under the non-normal process, using the freely estimated processes. The variance channel is the change in the CEV when the non-normal process is rescaled to the nonlinear process's component variances. The shape channel is the residual and captures the effect of the shape of shocks. Columns add to the total gap up to rounding.

0.99 percentage points come from the variance channel. For the NT reform, the total gap is 0.73 percentage points; the variance channel accounts for 0.51 percentage points, while the shape channel accounts for the remaining 0.22 percentage points. Shape therefore matters most when the reform changes insurance directly. For the PT reform, which keeps transfers in place and mainly changes labor-supply distortions, the shape channel is small and slightly negative.

The variance-matching exercises reported in Appendix F make the same point for all three restricted processes at once. When the linear, age-dependent, and non-normal processes are rescaled to carry the nonlinear process's component variances, their NW welfare effects collapse from a range of −0.47 to +0.63 percent of lifetime consumption into a tight band between +0.51 and +0.54 percent, close to the nonlinear benchmark of +0.72 percent; the constrained re-estimated versions remain in a similar range (+0.52 to +0.75). Two lessons follow. First, once the amount and allocation of wage risk are equalized, the welfare evaluations of the same reform nearly coincide across specifications: the variances are the dominant driver. Second, in the rescaling exercise the small remaining gap relative to the nonlinear benchmark is the pure shape effect. Every rescaled process with linear persistence places a value on the welfare state that is roughly 0.2 percentage points higher than the nonlinear process — the imprint of state-dependent mean reversion, which makes bad persistent states less durable and therefore reduces the insurance the tax-transfer system needs to provide.

Appendix F provides the details behind the variance-matching exercises. It reports the full set of CEVs for the freely estimated processes, the rescaled processes, and the constrained re-estimated processes. The appendix also reports the full breakdown of welfare gains by permanent type. The purpose of the appendix is to show that the message in Table 7 is not an artifact of a single rescaling exercise: the amount and allocation of risk explain most of the cross-process welfare gap, while shape has an independent effect concentrated in the insurance comparisons.

4.2. Mechanisms

The welfare mechanisms follow directly from the wage-process evidence in Sections 2 and 3.2. Section 2 showed that the four processes differ not only in higher-order moments, but also in how they allocate wage variation. The nonlinear and non-normal processes assign a much larger role to permanent heterogeneity and a smaller role to transitory shocks than the Gaussian processes. This is why they generate steeper welfare gradients when progressivity and transfers are reduced: the welfare state redistributes across permanent types, and the scope for this redistribution is larger when permanent heterogeneity is larger.

Section 3.2 showed that the processes also differ in how much persistent wage risk reaches consumption. The nonlinear process has lower pass-through of persistent shocks because persistent-looking shocks are partly mean-reverting. After an unfavorable draw, the persistent component tends to move back toward the middle of the distribution. The wage process therefore provides some insurance on its own, reducing the value of means-tested transfers. This is why removing transfers is less costly under the nonlinear process than under an otherwise similar process with linear persistence.

Non-normality works in the opposite direction. Fat-tailed and skewed shocks create rare but severe low-income states. These are precisely the states in which a means-tested transfer floor is most valuable. This explains why the non-normal process generates the largest welfare loss from removing transfers and the largest worst-case losses under NT and NW. It also explains why the combined reform has a negative mean effect only under the non-normal process, even though the median household still gains.

Putting the pieces together, the policy conclusion is not that one feature of the wage process dominates all others. The amount and allocation of risk are first-order: they determine the size of the redistributive and insurance problems faced by the tax-transfer system. Conditional on those variances, shape matters through insurance. Nonlinear mean reversion lowers the value of public insurance by making bad persistent states less durable, while non-normality raises it by increasing the probability and severity of bad states that households cannot smooth privately.

5. Conclusion

Does the specification of wage risk matter for the welfare evaluation of taxes and transfers? This paper shows that it does. We estimate four wage processes on PSID male hourly wages: the canonical linear process, the benchmark nonlinear process with age dependence, non-normality, and state-dependent persistence, and two intermediate processes that isolate age dependence and non-normality. We then embed each process in the same incomplete-markets life-cycle economy with endogenous labor supply, progressive income taxation, means-tested transfers, and an earnings-linked U.S.-style Social Security system.

The validation exercise favors the nonlinear process. It performs best on untargeted consumption and hours moments, especially the pass-through of persistent wage shocks to consumption. The linear and age-dependent processes transmit too much persistent wage risk into consumption. The non-normal process captures the fat-tailed shape of wage changes and the flat profile of consumption inequality, but it assigns more risk to permanent components and implies too much pass-through relative to the nonlinear benchmark.

These differences matter for policy. Removing federal tax progressivity and means-tested transfers, while keeping Social Security in place, raises mean welfare by 0.72% of lifetime consumption under the nonlinear process, but lowers mean welfare by 0.47% under the non-normal process. Thus, the wage process is not just a technical input: it can change the sign and magnitude of the measured value of tax-transfer reform.

The decomposition explains why. Most cross-process differences reflect both the total amount of wage risk and its allocation across permanent heterogeneity, persistent shocks, and transitory shocks. Permanent heterogeneity governs redistribution across permanent types. Persistent shocks govern the insurance value of taxes and transfers. Transitory shocks are easier to smooth through saving and labor supply. Conditional on this amount and allocation of risk, the shape of shocks still matters. Nonlinear mean reversion lowers the value of public insurance because bad persistent states are less durable. Non-normality works in the opposite direction: fat-tailed and skewed shocks create rare but severe low-income states, precisely the states in which means-tested transfers are most valuable.

Finally, two extensions of the current analysis are natural. First, the model has a frictionless labor-supply margin and therefore does not reproduce the early-life decline in hours dispersion emphasized by Kaplan (2012). Incorporating employment frictions, an extensive labor-supply margin, or richer household structure would allow the model to speak more directly to these patterns.[9] A further step is to move from the illustrative reforms studied here to the optimal joint design of progressivity and transfers under nonlinear wage risk.

References

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica* 85 (3): 693–734. 10.3982/ECTA13795.

[9] Hubmer (2018) shows that on-the-job search, separations, endogenous search effort, and human-capital losses in unemployment can generate negative skewness and excess kurtosis in earnings growth.

Blundell, Richard, Luigi Pistaferri, and Ian Preston. 2008. "Consumption Inequality and Partial Insurance." *American Economic Review* 98 (5): 1887–1921. 10.1257/aer.98.5.1887.

Blundell, Richard, Luigi Pistaferri, and Itay Saporta-Eksten. 2016. "Consumption Inequality and Family Labor Supply." *American Economic Review* 106 (2): 387–435. 10.1257/aer.20121549.

Borella, Margherita, Mariacristina De Nardi, Michael Pak, Nicolo Russo, and Fang Yang. 2023. "The Importance of Modeling Income Taxes over Time: U.S. Reforms and Outcomes." *Journal of the European Economic Association* 21 (6): 2237–2286. 10.1093/jeea/jvad053.

Borella, Margherita, Mariacristina De Nardi, and Fang Yang. 2023. "Are Marriage-Related Taxes and Social Security Benefits Holding Back Female Labour Supply?" *The Review of Economic Studies* 90 (1): 102–131. 10.1093/restud/rdac018.

Chetty, Raj, Adam Guren, Day Manoli, and Andrea Weber. 2011. "Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins." *American Economic Review* 101 (3): 471–475. 10.1257/aer.101.3.471.

Conesa, Juan Carlos and Dirk Krueger. 2006. "On the Optimal Progressivity of the Income Tax Code." *Journal of Monetary Economics* 53 (7): 1425–1450. 10.1016/j.jmoneco.2005.03.016.

De Nardi, Mariacristina, Giulio Fella, Marike Knoef, Gonzalo Paz-Pardo, and Raun Van Ooijen. 2021. "Family and Government Insurance: Wage, Earnings, and Income Risks in the Netherlands and the U.S." *Journal of Public Economics* 193: 104327. 10.1016/j.jpubeco.2020.104327.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2020. "Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare." *Journal of the European Economic Association* 18 (2): 890–926. 10.1093/jeea/jvz010.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2024. "Wage Risk and Government and Spousal Insurance." *The Review of Economic Studies* 92 (2): 954–980. 10.1093/restud/rdae042.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2025. "Wage Risk and Government and Spousal Insurance." *Review of Economic Studies* 92 (2): 954–980. 10.1093/restud/rdae042.

Ferriere, Axelle, Philipp Grübener, Gaston Navarro, and Oliko Vardishvili. 2023. "On the Optimal Design of Transfers and Income Tax Progressivity." *Journal of Political Economy Macroeconomics* 1 (2): 276–333. 10.1086/725034.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2016. "Redistribution and Social Insurance." *American Economic Review* 106 (2): 359–86. 10.1257/aer.20111550.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2012. "Taxation and Household Labour Supply." *The Review of Economic Studies* 79 (3): 1113–1149. 10.1093/restud/rdr049.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2014. "Income Taxation of U.S. Households: Facts and Parametric Estimates." *Review of Economic Dynamics* 17 (4): 559–581. 10.1016/j.red.2014.01.003.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2023. "Rethinking the Welfare State." *Econometrica* 91 (6): 2261–2294. 10.3982/ECTA19921.

Guner, Nezih, Martin Lopez-Daneri, and Gustavo Ventura. 2016. "Heterogeneity and Government revenues: Higher taxes at the top?" *Journal of Monetary Economics* 80: 69–85. 10.1016/j.jmoneco.2016.05.002.

Guner, Nezih, Christopher Rauh, and Gustavo Ventura. 2026. "Means-Tested Transfers in the U.S.: Facts and Parametric Estimates." *Review of Economic Dynamics* 61: 101348. https://doi.org/10.1016/j.red.2026.101348.

Guvenen, Fatih, Fatih Karahan, Serdar Ozkan, and Jae Song. 2021. "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?" *Econometrica* 89 (5): 2303–2339. `10.3982/ECTA14603`.

Guvenen, Fatih, Serdar Ozkan, and Rocio Madera. 2024. "Consumption Dynamics and Welfare under Non-Gaussian Earnings Risk." *Journal of Economic Dynamics and Control* 169: 104945. `https://doi.org/10.1016/j.jedc.2024.104945`.

Guvenen, Fatih, Luigi Pistaferri, and Giovanni L. Violante. 2022. "Global Trends in Income Inequality and Income Dynamics: New Insights from GRID." *Quantitative Economics* 13 (4): 1321–1360. `https://doi.org/10.3982/QE2260`.

Gálvez, Julio and Gonzalo Paz-Pardo. 2026. "Richer Earnings Dynamics, Consumption and Portfolio Choice over the Life Cycle." *Journal of Financial Economics* 176: 104206. `10.1016/j.jfineco.2025.104206`.

Heathcote, Jonathan, Fabrizio Perri, and Giovanni L. Violante. 2010. "Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States, 1967–2006." *Review of Economic Dynamics* 13 (1): 15–51. `10.1016/j.red.2009.10.010`.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L. Violante. 2017. "Optimal Tax Progressivity: An Analytical Framework." *The Quarterly Journal of Economics* 132 (4): 1693–1754. `10.1093/qje/qjx018`.

Hubmer, Joachim. 2018. "The Job Ladder and its Implications for Earnings Risk." *Review of Economic Dynamics* 29: 172–194. `10.1016/j.red.2018.01.003`.

Huggett, Mark. 1996. "Wealth Distribution in Life-cycle Economies." *Journal of Monetary Economics* 38 (3): 469–494. `10.1016/S0304-3932(96)01291-3`.

Kaplan, Greg. 2012. "Inequality and the Life Cycle." *Quantitative Economics* 3 (3): 471–525. `10.3982/QE200`.

Karahan, Fatih and Serdar Ozkan. 2013. "On the Persistence of Income Shocks over the Life Cycle: Evidence, Theory, and Implications." *Review of Economic Dynamics* 16 (3): 452–476. `10.1016/j.red.2012.08.001`.

Keane, Michael and Richard Rogerson. 2015. "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective." *Annual Review of Economics* 7 (1): 89–117. `10.1146/annurev-economics-080614-115601`.

Kindermann, Fabian and Dirk Krueger. 2022. "High Marginal Tax Rates on the Top 1 Percent? Lessons from a Life-Cycle Model with Idiosyncratic Income Risk." *American Economic Journal: Macroeconomics* 14 (2): 319–366. `10.1257/mac.20150170`.

Storesletten, Kjetil, Christopher I. Telmer, and Amir Yaron. 2004. "Consumption and Risk Sharing over the Life Cycle." *Journal of Monetary Economics* 51 (3): 609–633.

Turek, Konrad, Matthijs Kalmijn, and Thomas Leopold. 2021. "The Comparative Panel File: Harmonized Household Panel Surveys from Seven Countries." *European Sociological Review* 37 (3): 505–523. `10.1093/esr/jcab006`.

Appendix A. Sample Construction

This appendix describes the construction of the estimation sample. We use the Panel Study of Income Dynamics (PSID), harmonized through the Comparative Panel File (Turek, Kalmijn, and Leopold 2021), for 1980–2019. The PSID interviewed households annually up to and including the 1997 wave, and every two years thereafter. Because income and hours refer to the previous calendar year, the annual waves cover income and hours through 1996. We restrict attention to the nationally representative Survey Research Center (SRC) sample, dropping the Survey of Economic Opportunity (SEO) oversample of low-income families. We keep households headed by men aged 25 to 64 and drop observations with missing information on age, education, race, or state of residence.

Variable construction. The analysis uses hourly wages for males aged 25 to 64, calculated as annual real labor income divided by total annual hours worked. Nominal quantities are deflated using the CPI, with 1968 as the base year. Education and ethnicity are the CPF's harmonized classifications: four ISCED-based education levels (below secondary; some high school; high-school diploma up to some college; college degree or more) and five ethnicity categories (white, Black, Asian, American Indian, and other).

Sample selection. Beyond the demographic restrictions above, we keep individuals who worked between 520 and 5,200 hours in the calendar year. We drop observations whose nominal hourly wage falls below one-half of the applicable state minimum wage, and observations with miscoded hourly wages (at or above $999). We also remove observations with implausibly large wage changes, which are likely to reflect measurement error. For each individual, we compute changes in log wages between consecutive observed waves (one-year changes in the annual part of the panel, two-year changes in the biennial part) and form the product of each change with the preceding one. This product is large and negative when a sharp movement is immediately reversed. We drop the middle observation whenever this product falls in the bottom 0.25 percent of its distribution within the survey year.

First-stage residualization. Before estimating the shock processes, we purge log wages of observable heterogeneity. On the full 1968–2019 cleaned sample, we regress log wages on an indicator for whether the wife has positive earnings, fixed effects for education, ethnicity, state, year, household size, and the number of children, together with year-by-education interactions. We denote the residualized log wage from this first-stage regression by ω_{ij}. Although the wage-process estimation uses the 1980–2019 sample, the first-stage residualization uses the cleaned 1968–2019 panel to estimate the demographic and time controls more precisely.

Frequency and panel structure. We date observations by survey wave rather than by the interview date, so interviews conducted in the calendar year after their wave remain assigned to the wave year. Income and hours refer to the calendar year preceding the wave. The estimation sample covers the 1980–2019 waves. From the residualized wages in this window, we estimate the deterministic age profile κ_j by fitting a fourth-order Hermite polynomial in age; Figure 1 plots the estimated profile. We then define the residual wage net of the deterministic age profile as

$$x_{ij} \equiv \omega_{ij} - \kappa_j,$$

and use x_{ij} to identify the permanent, persistent, and transitory components. The covariance-based estimators (linear and age-dependent) use all spells of at least five consecutive annual or three consecutive biennial observations; the stochastic-EM estimators (nonlinear and non-normal) use balanced, non-overlapping windows of exactly five annual or three biennial observations. When a worker's history yields more than one non-overlapping window, the windows enter the estimation as separate units. The age profile is estimated on all observations in the window, before these spell requirements are imposed. The resulting sample is a panel of wage residuals indexed by age and time.

Appendix B. Estimation Details for Non-normal and Nonlinear Wage Processes

This appendix provides formal details on the stochastic EM algorithm used to estimate the non-normal and nonlinear wage processes. Section 2.3 gives an overview of the algorithm; here we describe the complete-data criterion, the posterior distribution of the latent components, the construction of densities from the conditional quantile functions, and the treatment of the mixed annual-biennial panel.

For individual i, let $\mathcal{O}_i$ denote the set of ages at which the residual wage is observed, let $j_{i0} = \min \mathcal{O}_i$ denote the age at first observation, and let $j_{iT} = \max \mathcal{O}_i$ denote the age at last observation. We write

$$x_i = (x_{ij})_{j \in \mathcal{O}_i}, \qquad x_{ij} = \omega_{ij} - \kappa_j.$$

Let

$$\mathcal{A}_i = \{j_{i0}, j_{i0} + 1, \dots, j_{iT}\}$$

denote the complete sequence of annual ages between the first and last observations. In the annual part of the sample, $\mathcal{A}_i = \mathcal{O}_i$. In the biennial part, $\mathcal{A}_i$ also contains the intervening ages at which wages are not observed.

The latent variables consist of the permanent component ζ_i, the annual path of the persistent component $(\eta_{ij})_{j \in \mathcal{A}_i}$, and the transitory components $(\varepsilon_{ij})_{j \in \mathcal{O}_i}$ at observed wage ages. At every

$j \in \mathcal{O}_i$, they satisfy the adding-up restriction

$$x_{ij} = \zeta_i + \eta_{ij} + \varepsilon_{ij}.$$

Complete-data criterion and EM representation. Let Λ collect the quantile-function coefficients evaluated at the grid points, together with the lower- and upper-tail parameters for the permanent, initial persistent, subsequent persistent, and transitory components. For a candidate parameter vector $\widetilde{\Lambda}$, let

$$R_i(\zeta_i, \eta_i, \varepsilon_i, x_i, \mathcal{O}_i; \widetilde{\Lambda})$$

denote the complete-data criterion for individual i. It consists of four sets of contributions: one for the permanent component, one for the initial persistent state, one for the subsequent one-year persistent transitions, and one for the transitory components,

$$\begin{aligned} R_i(\cdot; \widetilde{\Lambda}) &= R_{\zeta,i}(\widetilde{\Lambda}) + R_{\eta_0,i}(\widetilde{\Lambda}) \\ &\quad + \sum_{\substack{j \in \mathcal{A}_i \\ j > j_{i0}}} R_{\eta,ij}(\widetilde{\Lambda}) + \sum_{j \in \mathcal{O}_i} R_{\varepsilon,ij}(\widetilde{\Lambda}). \end{aligned} \tag{A1}$$

For the central part of each conditional distribution, these terms contain the usual quantile-regression check losses. For a generic outcome y, conditioning variables z, and basis vector $B_x(z)$, the contribution at quantile τ_l is

$$\rho_{\tau_l}\left(y - B_x(z)' a^x(\tau_l)\right), \qquad \rho_\tau(v) = v\left(\tau - \mathbf{1}\{v < 0\}\right).$$

The relevant dependent and conditioning variables are ζ_i conditional on j_{i0}, $\eta_{i0} \equiv \eta_{i,j_{i0}}$ conditional on (ζ_i, j_{i0}), η_{ij} conditional on $(\eta_{i,j-1}, h_j)$, and ε_{ij} conditional on h_j. The criterion also includes the negative log-likelihood contributions used to estimate the lower and upper Laplace-tail parameters.

If all latent components were observed, the population parameters would solve

$$\Lambda^\star = \arg\min_{\widetilde{\Lambda}} \mathbb{E}\left[R_i(\zeta_i, \eta_i, \varepsilon_i, x_i, \mathcal{O}_i; \widetilde{\Lambda})\right]. \tag{A2}$$

By the law of iterated expectations, the population criterion can equivalently be written as

$$\mathbb{E}\left[R_i(\cdot; \widetilde{\Lambda})\right] = \mathbb{E}_{x_i, \mathcal{O}_i}\left[\mathbb{E}\left[R_i(\cdot; \widetilde{\Lambda}) \mid x_i, \mathcal{O}_i\right]\right]. \tag{A3}$$

At iteration t, the EM criterion evaluates the conditional distribution of the latent variables at the current parameter estimate $\widehat{\Lambda}^{(t)}$ while optimizing over the candidate parameter vector $\widetilde{\Lambda}$:

$$\mathcal{Q}_N\left(\widetilde{\Lambda} \mid \widehat{\Lambda}^{(t)}\right) = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}_{\widehat{\Lambda}^{(t)}}\left[R_i(\cdot;\widetilde{\Lambda}) \mid x_i, \mathcal{O}_i\right]. \tag{A4}$$

The stochastic EM algorithm approximates each conditional expectation in (A4) with a simulated latent path.

Posterior distribution of the latent components. Given $\widehat{\Lambda}^{(t)}$, the stochastic E-step draws the permanent component and the annual persistent path from their posterior distribution conditional on the observed wage history. It is convenient to write this posterior over (ζ_i, η_i) and recover the transitory component at each observed age from

$$\varepsilon_{ij} = x_{ij} - \zeta_i - \eta_{ij}, \qquad j \in \mathcal{O}_i.$$

This representation imposes the adding-up restrictions exactly and is equivalent to drawing all three components jointly subject to those restrictions.

Suppressing the parameter argument, define the posterior kernel

$$\begin{aligned}\pi_i(\zeta_i, \eta_i; \Lambda) = {} & f_\zeta(\zeta_i \mid j_{i0}; \Lambda) f_{\eta_0}\left(\eta_{i0} \mid \zeta_i, j_{i0}; \Lambda\right) \\ & \times \prod_{\substack{j\in\mathcal{A}_i \\ j>j_{i0}}} f_\eta\left(\eta_{ij} \mid \eta_{i,j-1}, h_j; \Lambda\right) \\ & \times \prod_{j\in\mathcal{O}_i} f_\varepsilon\left(x_{ij} - \zeta_i - \eta_{ij} \mid h_j; \Lambda\right).\end{aligned} \tag{A5}$$

The first line contains the density of the permanent component and the conditional density of the persistent component at entry into the estimation sample. The second line contains all one-year transitions of the persistent component between the first and last observed ages. The final line evaluates the transitory density at the value implied by the observed residual wage and the proposed permanent and persistent components.

The posterior distribution used in iteration t is therefore

$$p\left(\zeta_i, \eta_i \mid x_i, \mathcal{O}_i; \widehat{\Lambda}^{(t)}\right) = \frac{\pi_i(\zeta_i, \eta_i; \widehat{\Lambda}^{(t)})}{\int \pi_i(\zeta, \eta; \widehat{\Lambda}^{(t)})\, d\zeta\, d\eta}.$$

The nonlinear and non-normal estimators differ only in the density used for the subsequent persistent transitions. For the nonlinear process, this density is implied by the tensor-product quantile function in equation (4). For the non-normal process, it is implied by the additive quantile function in equation (12). The densities of the permanent, initial persistent, and transitory components are constructed in the same way in both estimators.

Densities implied by the quantile functions. The densities in (A5) are obtained from the estimated conditional quantile functions. For a conditional quantile function $Q_x(u \mid z)$ that is strictly increasing in u, its inverse is the conditional distribution function:

$$F_x(v \mid z) = Q_x^{-1}(v \mid z).$$

Consequently,

$$f_x(v \mid z) = \frac{\partial Q_x^{-1}(v \mid z)}{\partial v} = \left[\frac{\partial Q_x(u \mid z)}{\partial u} \right]^{-1} \Bigg|_{u = Q_x^{-1}(v \mid z)}. \tag{A6}$$

For example, the one-year transition density of the persistent component is

$$f_\eta \left(\eta_{ij} \mid \eta_{i,j-1}, h_j \right) = \frac{\partial Q_\eta^{-1} \left(\eta_{ij} \mid \eta_{i,j-1}, h_j \right)}{\partial \eta_{ij}}. \tag{A7}$$

The interpolation and tail specification in equation (9) make these densities available in closed form. To see this, let

$$q_l^x(z) \equiv \widehat{Q}_x(\tau_l \mid z)$$

denote the fitted conditional quantile at grid point τ_l. For an ordered fitted quantile grid, the corresponding density is

$$f_x(v \mid z) = \begin{cases} \lambda_L^x \tau_1 \exp\{\lambda_L^x [v - q_1^x(z)]\}, & v \leq q_1^x(z), \\ \dfrac{\tau_{l+1} - \tau_l}{q_{l+1}^x(z) - q_l^x(z)}, & q_l^x(z) < v \leq q_{l+1}^x(z), \quad l = 1, \ldots, L-1, \\ \lambda_U^x (1 - \tau_L) \exp\{-\lambda_U^x [v - q_L^x(z)]\}, & v > q_L^x(z). \end{cases} \tag{A8}$$

Thus, linear interpolation of the quantile function produces a constant density between adjacent fitted quantiles, while the two Laplace-tail extensions produce decaying lower and upper tails exponentially. The lower tail integrates to probability τ_1, the upper tail integrates to probability $1 - \tau_L$, and the central intervals carry the probability mass between adjacent quantile grid points.

Metropolis-Hastings step. The posterior distribution has no convenient closed-form representation because the permanent component enters every observed wage, the persistent states are linked across ages, and the transitory components are tied to the other components through the adding-up restrictions. We therefore use a Metropolis–Hastings sampler.

Let $\chi_i = (\zeta_i, \eta_i)$ denote the sampler state. Given a current state $\chi_i^{(m)}$, suppose the sampler proposes a candidate $\chi_i^\star$ from a proposal density $q_i(\chi_i^\star \mid \chi_i^{(m)})$. The candidate is accepted with probability

$$\alpha_i = \min\left\{1, \frac{\pi_i(\chi_i^\star; \widehat{\Lambda}^{(t)}) q_i(\chi_i^{(m)} \mid \chi_i^\star)}{\pi_i(\chi_i^{(m)}; \widehat{\Lambda}^{(t)}) q_i(\chi_i^\star \mid \chi_i^{(m)})}\right\}. \tag{A9}$$

Because the normalizing constant of the posterior cancels from this ratio, the sampler requires only the kernel in (A5). We use a coordinate-wise Gaussian random-walk proposal, so the proposal densities also cancel from the acceptance ratio; each E-step runs 200 sweeps over the coordinates of χ_i and retains the final draw. Once a draw $(\widehat{\zeta}_i^{(t)}, \widehat{\eta}_i^{(t)})$ has been obtained, the transitory components at observed ages are recovered as

$$\widehat{\varepsilon}_{ij}^{(t)} = x_{ij} - \widehat{\zeta}_i^{(t)} - \widehat{\eta}_{ij}^{(t)}, \qquad j \in \mathcal{O}_i.$$

M-step. With one simulated complete-data path per individual, the stochastic approximation to (A4) is

$$\widehat{\mathcal{Q}}_N^{(t)}(\widetilde{\Lambda}) = \frac{1}{N}\sum_{i=1}^{N} R_i\left(\widehat{\zeta}_i^{(t)}, \widehat{\eta}_i^{(t)}, \widehat{\varepsilon}_i^{(t)}, x_i, \mathcal{O}_i; \widetilde{\Lambda}\right).$$

The M-step updates the parameters according to

$$\widehat{\Lambda}^{(t+1)} = \arg\min_{\widetilde{\Lambda}} \frac{1}{N}\sum_{i=1}^{N} R_i\left(\widehat{\zeta}_i^{(t)}, \widehat{\eta}_i^{(t)}, \widehat{\varepsilon}_i^{(t)}, x_i, \mathcal{O}_i; \widetilde{\Lambda}\right). \tag{A10}$$

At each grid point τ_l, this amounts to four quantile regression models. The permanent-component regression uses $\widehat{\zeta}_i^{(t)}$ as the dependent variable and the entry-age basis in equation (7). The initial-persistent regression uses $\widehat{\eta}_{i0}^{(t)}$ and the tensor-product basis in $(\widehat{\zeta}_i^{(t)}, j_{i0})$ from equation (5). The transitory regression uses $\widehat{\varepsilon}_{ij}^{(t)}$ and the age basis in equation (6).

For the subsequent persistent transitions, the nonlinear estimator regresses $\widehat{\eta}_{ij}^{(t)}$ on the tensor-product basis in $(\widehat{\eta}_{i,j-1}^{(t)}, h_j)$ from equation (4). The non-normal estimator instead imposes the additive basis $(1, \widehat{\eta}_{i,j-1}^{(t)}, h_j, h_j^2)$ from equation (12). The lower- and upper-tail parameters are updated by maximizing the likelihood contributions of observations in the corresponding tails. Because the component levels are identified separately only up to location, we apply the same location normalization at each M-step. All other steps are common to the nonlinear and non-normal estimators.

We initialize the algorithm from a preliminary set of parameter values and run 500 stochastic EM iterations. To reduce the simulation noise generated by using posterior draws in place of analytical conditional expectations, we report the average of the parameter estimates over the final 250 iterations.

Mixed annual and biennial observations. The wage process used in the quantitative model is annual, while the PSID is observed biennially after the 1997 survey wave. For the covariance-based estimators of the linear and age-dependent processes, no wage or state imputation is needed. The empirical moment vector contains only age pairs observed in the data, and the model-implied covariance is evaluated at the corresponding annual lag.

For the stochastic EM estimators, the persistent component is defined at every age in $\mathcal{A}_i$, including ages at which wages are not observed. These missing annual persistent states are treated as additional latent variables in the posterior sampler. Suppose, for example, that residual wages are observed at ages j and $j+2$, but not at age $j+1$. The sampler draws $\eta_{i,j+1}$ jointly with ζ_i, η_{ij}, $\eta_{i,j+2}$, and the remaining persistent path. The persistent-density contribution for this segment is

$$f_\eta\left(\eta_{i,j+1} \mid \eta_{ij}, h_{j+1}\right) f_\eta\left(\eta_{i,j+2} \mid \eta_{i,j+1}, h_{j+2}\right).$$

The observed wages at ages j and $j+2$ contribute the transitory-density factors

$$f_\varepsilon\left(x_{ij} - \zeta_i - \eta_{ij} \mid h_j\right) f_\varepsilon\left(x_{i,j+2} - \zeta_i - \eta_{i,j+2} \mid h_{j+2}\right).$$

There is no transitory-density contribution at age $j+1$, because no wage is observed at that age. The unobserved transitory component can be integrated out and contributes a factor of one to the posterior. In the M-step, the two sampled one-year transitions through $\eta_{i,j+1}$ enter the persistent-component quantile regressions, whereas the transitory-component quantile regressions use only observed wage ages.

The likelihood therefore incorporates every annual persistent transition implied by the model while attaching transitory shocks only to observed wage observations. This procedure uses the full annual and biennial wage histories and delivers the annual transition process required by the quantitative model.

Appendix C. Stationary Recursive Equilibrium

Let Φ_j denote the stationary distribution of agents over the age-j state space $\mathcal{S}_j$. Aggregate capital and labor in efficiency units are

$$K = \sum_{j=25}^{J} \mu_j \int_{\mathcal{S}_j} a \, d\Phi_j(\mathbf{s}),$$

$$L = \sum_{j=25}^{j_R} \mu_j \int_{\mathcal{S}_j} e_j(\Omega) l_j(\mathbf{s})\, d\Phi_j(\mathbf{s}).$$

Output and factor prices in stationary units are

$$\begin{aligned} Y &= AK^{\alpha} L^{1-\alpha}, \\ w &= (1-\alpha) A \left(\frac{K}{L}\right)^{\alpha}, \\ r &= \alpha A \left(\frac{K}{L}\right)^{\alpha-1} - \delta. \end{aligned}$$

Aggregate consumption is

$$C = \sum_{j=25}^{J} \mu_j \int_{\mathcal{S}_j} c_j(\mathbf{s})\, d\Phi_j(\mathbf{s}),$$

and the stationary aggregate resource constraint is

$$Y + (1-\delta)K = C + G + (1+n)(1+g)K.$$

Accidental bequests are fully taxed by the government. With the timing of the household problem above, their stationary value is

$$\mathcal{B} = \frac{1}{1+n} \sum_{j=25}^{J} \mu_j (1 - s_{j+1})(1+r) \int_{\mathcal{S}_j} a_j'(\mathbf{s})\, d\Phi_j(\mathbf{s}),$$

where $s_{J+1} = 0$. Bequests are measured at the start of the following period, inclusive of the return on the assets of the deceased.

For compactness, define age-j taxable income by

$$I_j(\mathbf{s}) = \begin{cases} we_j(\Omega) l_j(\mathbf{s}) + ra, & \text{if } j \le j_R, \\ SS(\zeta, \eta_R) + ra, & \text{if } j > j_R. \end{cases}$$

The general government budget, which excludes Social Security, is

$$G = \sum_{j=25}^{J} \mu_j \int_{\mathcal{S}_j} \left[T_f\big(I_j(\mathbf{s})\big) + \tau_l I_j(\mathbf{s}) + \tau_k ra - TR\big(I_j(\mathbf{s})\big) \right] d\Phi_j(\mathbf{s}) + \mathcal{B}.$$

Thus, government expenditure is financed by federal, local, and capital-income taxes and by fully taxed accidental bequests, net of means-tested transfers. Payroll-tax collections and

Social Security benefits are kept in a separate budget, which satisfies

$$\sum_{j=25}^{j_R} \mu_j \int_{\mathcal{S}_j} T^{ss}\Big(we_j(\Omega)l_j(\mathbf{s})\Big)\, d\Phi_j(\mathbf{s}) = \sum_{j=j_R+1}^{J} \mu_j \int_{\mathcal{S}_j} \frac{SS(\zeta, \eta_R)}{(1+g)^{j-j_R}}\, d\Phi_j(\mathbf{s}),$$

where the de-escalation factor $(1+g)^{j-j_R}$ reflects that benefits are fixed in real terms at retirement and expressed here in stationary units.

Given the policy parameters and a revenue requirement $\bar{R}$, a stationary recursive competitive equilibrium consists of household decision rules, factor prices (w, r), a federal tax level τ_0, mean income $\bar{I}$, aggregate quantities (K, L, Y, C), government consumption G, and age-specific distributions $\{\Phi_j\}_{j=25}^{J}$ such that households optimize, the representative firm maximizes profits, the capital and labor markets clear, the distributions evolve according to the household policies, survival probabilities, and wage-state transitions and are stationary, the aggregate resource constraint holds, mean income is consistent with the distributions, total revenue from the federal, local, and capital-income taxes equals $\bar{R}$, government consumption is determined residually by the general government budget, and Social Security payments equal payroll-tax collections. In the baseline, $\bar{R}$ is set so that this revenue equals 14.51% of GDP; in the counterfactuals, $\bar{R}$ equals the baseline's realized revenue level. In the model, mean income is measured as $\bar{I} = wL + rK$, pre-tax household income inclusive of the payroll contribution.

Appendix D. Calibration Targets and Data Sources

This appendix documents the data sources for each calibration target in Section 3.1, states the Social Security benefit formula, and describes the data windows used in the validation exercise in Section 3.2.

Externally set parameters. Population grows at $n = 0.94\%$, the average U.S. rate over 1980–2019. Labor-augmenting productivity grows at $g = 1.81\%$, the average growth rate of U.S. real GDP per capita over the same period, computed from FRED data; along the model's balanced growth path, output per capita grows at rate g, so per-capita output growth is the model-consistent calibration target. Survival probabilities are taken from the U.S. Life Tables for 2005 (*National Vital Statistics Reports*, Vol. 58, No. 10, 2010). The Frisch elasticity is $\gamma = 1$, in line with the macro labor-supply literature. The capital share $\alpha = 0.391$ is the complement of the labor share of income in FRED (series `LABSHPUSA156NRUG`). The flat local income tax $\tau_l = 5\%$ and the progressivity parameter $\tau_f = 0.053$ are from Guner, Kaygusuz, and Ventura (2014). The four transfer parameters (b_0, b_1, b_2, b_3) are the parametric estimates of Guner, Rauh, and Ventura (2026).

Calibrated parameters. The depreciation rate δ is derived from the steady-state capital-accumulation identity, $\delta = I/K - [(1+n)(1+g) - 1]$, to match an investment-to-output ratio of 17.70%, the 1980–2019 average from BEA data on GDP and Gross Private Domestic Investment. The capital-income tax rate τ_k matches corporate tax collections of 1.8% of GDP, the 1980–2019 average from OMB Historical Table 2.3. The level parameter τ_0 is set so that combined federal, local, and capital-income tax revenue meets the 14.51%-of-GDP target, the 1980–2019 average of the corresponding U.S. tax collections relative to GDP, computed from FRED data. The Social Security scale s^{ss} balances the PAYGO pension budget.

SMM targets. The discount factor β and the disutility-of-work parameter ψ are estimated by SMM, separately for each wage process, to match (i) a capital-to-output ratio of 3.01, the 1980–2019 average from BEA data on GDP and the Net Stock of Fixed Assets, and (ii) mean working-age hours of one-third of the time endowment.

Social Security benefit formula. For a benefit base z (the cell average of career-average capped earnings described in Section 3.1), the Primary Insurance Amount is

$$PIA(z) = 0.90 \min\{z, b_1\} + 0.32 \max\{0, \min\{z, b_2\} - b_1\} + 0.15 \max\{0, z - b_2\},$$

with bend points $b_1 = 0.21\,\bar{y}$ and $b_2 = 1.29\,\bar{y}$, where $\bar{y}$ is average labor earnings. Benefits are given by $s^{ss} \cdot PIA(z)$, with the scale s^{ss} balancing the PAYGO budget at the fixed payroll rate $\tau_{ss} = 10.6\%$.

Validation data windows. The untargeted moments in Section 3.2 use the following PSID windows. Consumption moments use 1999–2019, the biennial period in which the PSID collects a broad measure of household expenditures; the non-durable consumption basket aggregates food, transportation, telephone, clothing, trips, and recreation expenditures, deflated and residualized in the same first stage as wages. Hours and wage moments use 1980–2019. The consumption pass-through ϕ_{pers} is estimated on the biennial 1999–2019 panels of residualized wages and consumption. These windows define the "PSID validation sample" referenced in the figure notes.

Table A1 reports the full distribution of each latent component (its variance, skewness, and kurtosis) on the discretized grids used in the quantitative model. For the persistent component, we report the distribution at labor-market entry, η_0. Entry is the only age at which the persistent component is a primitive object of the estimation: at any later age, η_j is an accumulation of past innovations, so its distribution mixes the shape of the shocks with the history of the process, and its dispersion and persistence by age are already reported in Figures 3 and 6.

TABLE A1. Moments of the estimated components

Process	Permanent ζ			Initial persistent η_0			Transitory ε		
	Var	Skew	Kurt	Var	Skew	Kurt	Var	Skew	Kurt
Linear	0.043	0.00	2.56	0.019	0.00	2.77	0.078	0.00	2.56
Age-dependent	0.055	0.00	2.56	0.079	0.00	2.77	0.041–0.086	0.00	2.56
Nonlinear	0.128	+0.31	3.30	0.144	−1.12	3.94	0.018–0.021	+1.10	6.19
Non-normal	0.179	+0.60	3.50	0.193	−1.19	4.34	0.022–0.031	+0.14	6.52

NOTES: Moments are computed from the *discretized grids* the quantitative model uses, not from the smooth estimated processes. η_0 denotes the persistent component at labor-market entry (age 25): its variance, skewness, and kurtosis are those of the initial distribution from which the persistent state is drawn. Under the linear process, this is the (discretized) Gaussian innovation distribution; under the age-dependent process, it is the separately estimated initial distribution; under the nonlinear and non-normal processes, it is the estimated initial-conditions distribution. Discretization collapses the far tails inward, pulling variance and especially kurtosis below their textbook values: a Gaussian shows a grid kurtosis of about 2.56 on the 11-point grids used for ζ and ε, and about 2.77 on the 18-point grid used for η, rather than 3.0. The same binning scheme is applied to all four processes, so the cross-process differences reflect the underlying processes rather than the grid. $\text{Var}(\varepsilon)$ is reported as a range when the transitory variance is age-dependent. The pre-discretization variances of the estimated processes are reported in Table 1.

Appendix E. Validation Scorecard and Consumption-Noise Sensitivity

This appendix summarizes the untargeted validation exercise in Section 3.2 and reports the sensitivity of the consumption-inequality comparison to the measurement-error correction.

Scorecard. Table A2 summarizes the fit of each calibrated economy to the validation moments. For the life-cycle profiles, the distance is a sample-size-weighted root-mean-square deviation between the model and data profiles over ages 25–64, with age cells weighted by the square root of their observation counts. For ϕ_{pers}, the distance is the absolute difference between the pooled model and data estimates. The nonlinear process ranks first on all three moments. The ranking for ϕ_{pers} and $\text{Var}(\log h)$ is not affected by the consumption measurement-error correction. The ranking for the level of $\text{Var}(\log c)$ depends on the correction, as discussed below.

Sensitivity. Only one result is sensitive to the consumption measurement-error correction: the level of consumption dispersion. Because the estimate is an upper bound, we vary the amount of noise added to simulated consumption, from none to the full correction. With no consumption noise, the non-normal process is closest on $\text{Var}(\log c)$ (0.026 versus 0.050 for the nonlinear process). With half of the upper-bound noise, the nonlinear process is already closest (0.024 versus 0.052). With the full upper-bound correction, the nonlinear process is closest (0.024 versus 0.082 for the non-normal process). The shape of the consumption profile is much less sensitive: only the nonlinear and non-normal processes generate the flat profile

TABLE A2. Validation scorecard: distance to PSID (rank in parentheses)

Process	$\mathrm{Var}(\log c)$	ϕ_{pers}	$\mathrm{Var}(\log h)$	Rank-sum
Nonlinear	0.024 (1)	0.085 (1)	0.025 (1)	**3**
Non-normal	0.082 (4)	0.148 (2)	0.025 (2)	8
Age-dependent	0.055 (2)	0.241 (3)	0.030 (3)	8
Linear	0.058 (3)	0.301 (4)	0.032 (4)	11

NOTES: Smaller distance is better. For life-cycle profiles, the distance is a sample-size-weighted root-mean-square deviation between the model and data profiles over ages 25–64, with age cells weighted by the square root of their observation counts. For ϕ_{pers}, the distance is the absolute difference between the pooled model and data estimates. Rank-sum adds the three per-moment ranks. The $\mathrm{Var}(\log c)$ column is evaluated using the upper-bound consumption measurement-error correction.

observed in the data. For this reason, the main text emphasizes the pass-through of persistent shocks, the shape of consumption inequality, and the hours moments, rather than relying only on the level of consumption dispersion.

Appendix F. Robustness: Isolating the Role of Shape

This appendix provides the details behind the decomposition in Section 4.1. The raw comparison across wage processes combines three forces: the total amount of wage variance, the allocation of that variance across the permanent, persistent, and transitory components (Table 2), and the shape of the shocks. By shape, we mean features such as skewness, kurtosis, age dependence, and state-dependent persistence. To isolate the role of shape, we need to hold fixed both the amount of risk and its allocation across components.

Variance-matching exercises. We use two complementary exercises. In both, the nonlinear process is the anchor and is left exactly as estimated. The linear, age-dependent, and non-normal processes are then modified so that they carry the nonlinear process's component variances.

The first exercise is the decomposition reported in Table 7. For each restricted process, we rescale each wage component so that its variance matches the corresponding variance in the nonlinear process. This rescaling preserves standardized features of the process, such as skewness, kurtosis, age profiles, and persistence patterns, while changing the amount and allocation of risk. Let W^{NL} denote the mean CEV under the nonlinear process, W^{p} the mean CEV under another freely estimated process p, and $W^{p,R}$ the mean CEV when process p is rescaled to the nonlinear component variances. The total nonlinear-minus-p gap is

$$W^{NL} - W^{p}.$$

We decompose this gap as

$$W^{NL} - W^{p} = \underbrace{\left(W^{p,R} - W^{p}\right)}_{\text{variance channel}} + \underbrace{\left(W^{NL} - W^{p,R}\right)}_{\text{shape channel}}.$$

The variance channel measures how much the welfare effect changes when process p is forced to carry the nonlinear process's component variances. The shape channel is the remaining gap between the nonlinear process and the variance-matched version of process p. Table 7 in the main text applies this decomposition to the non-normal process.

The second exercise is a constrained re-estimation. We re-estimate each restricted process subject to the requirement that its component variances match those of the nonlinear process. For the linear process, the three component variances are pinned down and the AR(1) persistence parameter is re-optimized. For the age-dependent process, the cubic age profiles are re-optimized subject to equality constraints on the integrated component variances. For the non-normal process, component variances are implied by the fitted quantile functions rather than chosen directly as free parameters. We therefore implement the constrained exercise through a shape-preserving rescaling of the fitted quantile functions and then recalibrate the economy. For the non-normal process, the "re-estimated" row therefore differs from the "rescaled" row only through the recalibration of (β, ψ), not through a distinct constrained estimation.

Results. Table A3 reports mean CEVs for the freely estimated processes, the rescaled processes, and the constrained re-estimated processes. The nonlinear row is the anchor and is unchanged throughout.

The first lesson is that the amount and allocation of risk explain a large share of the welfare differences across processes. This is most transparent for the non-normal process. In the free estimates, the non-normal process implies a much lower CEV than the nonlinear process for the reforms that remove transfers. Once the non-normal process is rescaled to carry the nonlinear process's component variances, its welfare implications move substantially toward the nonlinear benchmark.

The second lesson is that shape still matters. Comparing the nonlinear and non-normal processes at the same component variances isolates the role of state-dependent mean reversion, because both processes allow non-Gaussian shocks but differ in whether persistence depends on the lagged persistent state. Under the nonlinear process, bad persistent realizations are more mean-reverting, so the wage process partly insures itself. This reduces the welfare loss from removing means-tested transfers relative to a process with linear persistence. Non-normality works in the opposite direction: fat-tailed and skewed innovations generate rare but severe bad states, precisely the states in which a means-tested transfer system is most

TABLE A3. Mean CEV under the variance-matching robustness checks (%)

Process	PT	NT	NW
Nonlinear (anchor)	2.01	−1.01	0.72
Linear — free	1.91	−0.94	0.63
Linear — rescaled	1.71	−0.91	0.51
Linear — re-estimated	2.03	−1.08	0.64
Age-dependent — free	1.84	−1.02	0.49
Age-dependent — rescaled	1.76	−0.94	0.54
Age-dependent — re-estimated	1.81	−0.77	0.75
Non-normal — free	1.61	−1.74	−0.47
Non-normal — rescaled	2.04	−1.23	0.52
Non-normal — re-estimated	2.03	−1.22	0.52

NOTES: CEVs are expressed as percentages of lifetime consumption. PT replaces the progressive federal income tax with a proportional federal tax while keeping means-tested transfers. NT removes means-tested transfers while keeping federal progressivity. NW removes federal progressivity and means-tested transfers while keeping Social Security in place. In the rescaled and re-estimated rows, each restricted process is forced to carry the nonlinear process's component variances. The nonlinear process is the anchor and is unchanged throughout.

valuable.

The third lesson is that the redistributive gradients in the free estimates are closely connected to the dispersion of permanent types. Table A4 shows that the nonlinear and non-normal processes generate much steeper welfare gradients across permanent types than the linear and age-dependent processes. This is consistent with Table 2, where the nonlinear and non-normal processes assign a larger share of wage variation to the permanent component.

Tail discretization. The quantitative model uses discretized versions of the estimated wage processes. Discretization compresses the extreme tails, thereby reducing measured kurtosis relative to the smooth estimated processes. This matters most for the nonlinear and non-normal processes, whose fat tails are one of the objects of interest. The results in Table A3 should therefore be interpreted as conservative with respect to the welfare role of non-normality. A finer discretization of the tails would likely strengthen the tail-risk channel, but a complete check would require re-running the full pipeline: discretizing the wage processes more finely, recalibrating each economy, and recomputing the welfare counterfactuals.

Welfare gains by permanent type. Table A4 reports the full breakdown of the free-estimate CEVs by permanent type. These are the numbers summarized in the main text by the mean CEV and the welfare gradient between the highest and lowest permanent types, and plotted in Figure 9.

TABLE A4. Welfare gains by permanent type (CEV, %)

		Permanent Types										
	Mean	1/21	1/21	1/21	1/7	1/7	1/7	1/7	1/7	1/21	1/21	1/21
	Proportional Taxation											
Nonlinear	2.01	-2.36	-1.42	-0.92	-0.02	0.92	1.74	2.67	3.84	4.75	5.68	9.02
Linear	1.91	-0.71	-0.21	0.08	0.75	1.36	1.85	2.35	3.11	3.59	3.95	5.03
Age-dependent	1.84	-1.07	-0.50	-0.18	0.53	1.18	1.75	2.33	3.20	3.76	4.20	5.39
Non-normal	1.61	-2.69	-2.13	-1.69	-0.80	0.23	1.14	2.22	3.64	4.80	5.99	10.24
	No Means-tested Transfers											
Nonlinear	-1.01	-7.15	-4.17	-3.32	-1.84	-0.99	-0.46	-0.02	0.40	0.61	0.69	0.97
Linear	-0.94	-4.10	-3.08	-2.61	-1.76	-1.11	-0.69	-0.33	0.07	0.35	0.41	0.71
Age-dependent	-1.02	-4.79	-3.51	-2.95	-1.91	-1.18	-0.71	-0.29	0.15	0.44	0.50	0.81
Non-normal	-1.74	-8.43	-6.16	-5.14	-3.00	-1.85	-1.12	-0.46	0.19	0.46	0.57	0.97
	Proportional Taxation and No Means-tested Transfers											
Nonlinear	0.72	-10.72	-6.34	-4.86	-2.25	-0.34	1.09	2.52	4.16	5.30	6.32	9.95
Linear	0.63	-5.61	-3.93	-3.11	-1.45	-0.10	0.86	1.78	3.00	3.80	4.23	5.65
Age-dependent	0.49	-6.76	-4.70	-3.75	-1.84	-0.35	0.75	1.82	3.19	4.07	4.59	6.12
Non-normal	-0.47	-12.46	-9.27	-7.64	-4.31	-1.96	-0.22	1.61	3.76	5.21	6.52	11.21

NOTES: Free-estimate CEVs by permanent type from the benchmark calibration of each process. The eleven permanent types are ordered from lowest to highest and carry population weights 1/21, 1/21, 1/21, 1/7, 1/7, 1/7, 1/7, 1/7, 1/21, 1/21, 1/21. PT replaces the progressive federal income tax with a proportional federal tax while keeping means-tested transfers. NT removes means-tested transfers while keeping federal progressivity. NW removes federal progressivity and means-tested transfers while keeping Social Security in place.