Taxes and Transfers with Nonlinear Wage Dynamics*

Nezih Guner† and Emre Enes Yavuz ‡

December 1, 2022

Preliminary - Please Do Not Circulate

Abstract

Following the methodology developed by Arellano, Blundell and Bonhomme (2017), we use the Panel Study of Income Dynamics to estimate a shock process for hourly wages that is both nonlinear, i.e., not AR(1), and not normal, i.e., with shocks drawn from flexible probability distributions. With the estimated nonlinear wage process, unusual shocks wipe out good or bad productivity histories, which can not be captured by a standard AR(1) specification. The persistence of shocks increases over the life cycle. We then build a standard life-cycle model with idiosyncratic income risk, incomplete markets, and endogenous labor supply decisions and simulate the model economy with standard and new (nonlinear and non-normal) wage shock to study the value of social insurance. We find that when wage dynamics are nonlinear and non-normal, social insurance (progressive taxes and means-tested transfers) is less valuable for poor households but more valuable for richer ones. The estimated wage dynamics show less persistence than the standard specification, and there is more income mobility along the life cycle, which increases the value of the welfare state for the rich and lower for the poor.

*We thank Manuel Arellano, Josep Pijoan Mas, Matthias Doepke, Mariacristina De Nardi for helpful comments.

†Universitat Autonoma de Barcelona, ICREA and BSE. E-mail: ngunermail@gmail.com

‡Northwestern University. E-mail: eey@u.northwestern.edu

1 Introduction

The life-cycle models with idiosyncratic earnings or wage shocks are one of the main building blocks of modern macroeconomics.[1] The shocks in these models are typically represented by a combination of permanent, persistent, and transitory components. The permanent and transitory shocks are drawn from mean-zero normal distributions, while persistent shocks follow an AR(1) process, with innovations also coming from a mean-zero normal distribution.

Recent evidence, however, shows that earnings dynamics are better represented by less restrictive, nonlinear shocks (Arellano, Blundell and Bonhomme, 2017; **?**). Arellano, Blundell and Bonhomme (2017) estimate earnings dynamics where components of shocks are drawn from distributions are estimated without restrictive distributional form assumptions. In particular, the persistent shock is estimated with a nonlinear dependence on the past with a flexible distributional form. As a result, income shocks are both non-linear, i.e., they are not restricted to AR(1), and they do not come from a normal distribution. We refer to shocks with these features as *nonlinear* and use *linear* shocks to indicate those under the standard assumptions.

Arellano, Blundell and Bonhomme (2017) show that the persistence of the earnings in the data depends on the size of a given shock and the current position of the individual receiving the shock in the earnings distribution. In particular, the persistence is lower when an unusual shock hits an individual, i.e., when a high-earner gets a bad shock or a low-earner a good one. The low persistence in these cases implies that the history of shocks matters less, and unusual shocks can wipe out a history of good or bad shocks and reset the process. This pattern creates positive earnings risk at the bottom of the earnings distribution and negative earnings risk at the top.

In this paper, we follow the methodology in Arellano, Blundell and Bonhomme (2017) to estimate a nonlinear shocks process for male *hourly wages* (earnings per hours worked), using data from the Panel Study of Income Dynamics (PSID). We also estimate a linear

[1] This literature builds on works by Aiyagari (1994); Bewley (1986); **?**); Huggett (1996).

process using the same data and sample restrictions. The nonlinear estimation uses annual PSID data for the 1968-1997 period and biannual data for the post-1997 period to estimate yearly shock.

We find evidence for both nonnormality in all three (permanent, persistent, and transitory) components of wages and nonlinearity in the persistent component. First, the distributions of wages and their changes are leptokurtic, with higher peaks and fatter tails than what a normal distribution would generate. The linear process misses these features due to restrictive normality assumption. Second, wage shocks exhibit nonlinear persistence, i.e., unusual shocks wipe out good or bad productivity histories. We also show that this *resetting* effect is stronger at the top of the wage distribution. In contrast, the persistence is governed by a single parameter in the linear process and does not allow nonlinear lag dependence. Finally, although these features are present for all ages, their strength changes over the life cycle. In particular, persistence increase with age; it is about 0.65 at age 25 and increases monotonically to around 0.95 at around 45. By construction, there is no such age dependence in the linear process.

Furthermore, we find that linear and nonlinear wage shocks imply very different sources of wage risk. In both processes, roughly half of the total variance in log wages comes from persistent shocks. The importance of permanent and transitory shocks, however, are very different. The contribution of permanent shocks is much more significant for the nonlinear process. They account for about 40% of the total variance in log wages, while their contribution is only 14% with a linear process. In contrast, the contribution of transitory shocks is much more important with linear shocks.

We next study the implications of these differences for the value of social insurance. We build a standard life-cycle model with idiosyncratic wage risk, incomplete markets, and endogenous labor supply decisions. In the model economy, there are three sources of insurance against productivity shocks: labor supply responses, savings, and government-provided social insurance. The government taxes households with a progressive tax schedule. As a result, tax liabilities decline more than the drop in income when a household faces an adverse shock.

There are also means-tested transfers to a household whose income falls below a threshold. We calibrate the model with linear and non-linear productivity processes, forcing it to match the same set of targets. To highlight the role of nonlinearity, for model simulations, we use a linear process that has the same variance decomposition of wage risk as the one implied by the non-linear process.

We then investigate the insurance value of taxes and transfers. First, we replace progressive taxes with a proportional income tax. This is a revenue-neutral reform where the proportional tax rate generates the same level of government tax revenue as in the benchmark economy. Second, we eliminate the means-tested transfers and lower income taxes accordingly. Finally, we move to a proportional tax system and eliminate the means-tested transfers. We report welfare gains and losses for each experiment, calculated as the percentage change in each period's consumption that would leave the agent indifferent between the counterfactual economy and the benchmark.

When we reduce redistribution, households at the bottom of the income distribution in the initial benchmark lose, while those at the top gain. But how much they gain and lose depends critically on the underlying productivity process. For poor households, losses are more significant when the productivity shocks are linear. These shocks imply very high persistence, and, as a result, poor households have little chance of moving away from the bottom of the income distribution. The nonlinear productivity process has lower persistence for them when they are hit by unusually good shocks. As a result, poor households can become not-so-poor next period, which reduces the value of the welfare state for them. By the same token, the value of progressive taxation is higher when the wage process is nonlinear for richer households. As rich households have higher chances of falling in the income distribution, a proportional income tax provides smaller gains under the nonlinear process.

Our paper closely relates to the large literature that studies optimal tax and transfer policies in life-cycle economies with idiosyncratic shocks. A group of papers within this literature focuses on the optimal progressivity of income taxes. One insight that emerges from this literature is that the revenue or welfare-maximizing level of progressivity depends critically on the nature of labor productivity shocks at the top of the income distribution. If being a top earner is a relatively transitory state (e.g. Kindermann and Krueger, forthcoming), then it is optimal to tax top earners at a high rate since their labor supply reaction will be limited. On the other hand, if being a top earner is a relatively permanent state (e.g. Guner, Lopez-Daneri and Ventura, 2016) or requires human capital accumulation (e.g. Heathcote, Storesletten and Violante, 2017; Badel, Huggett and Luo, 2020), there is less room to increase taxes at the top. Furthermore, Golosov, Troshkin and Tsyvinski (2016) show that capturing higher moments (like kurtosis) of income distribution implies higher optimal labor and capital taxes at the top. Another, more recent, set of paper study reforms of the welfare state with a focus on universal basic income programs, e.g., Luduvice (2021); Conesa, Li and Li (2021); Daruich and Fernández (2020); Guner, Kaygusuz and Ventura (2021). Within this literature, De Nardi, Fella and Paz-Pardo (2020) estimate nonlinear wage dynamics using UK data following a similar approach to this paper. They study the UK's 2016 Universal Credit benefit reform and contrast the implications of linear and nonlinear wage processes. Our results on the value of social insurance are also related to De Nardi, Fella and Paz-Pardo (2020), who introduce earnings risk with these features into a life-cycle model and show that earnings fluctuations with nonlinear shocks generate lower welfare costs than the linear ones.

The rest of the paper is organized in the following way. Section 2 specifies the wage processes and the estimation strategies, Section 3 and 4 provides the estimation results for nonlinear and linear processes, Section 5 provides the details of the economic environment, Section 5.1 gives the functional forms and the parameter values of the economic model, Section 6 explains the analysis and provides the results and Section 7 concludes the paper.

2 Productivity Process and Estimation

We assume that log wages is the sum of four components. Let ω_{it} denote log wages of individual i at age t, given by,

$$\omega_{it} = \kappa_t + \theta_i + \eta_{it} + \epsilon_{it}. \tag{2.1}$$

The first component κ_t is deterministic, only a function of age, and common to all individuals. The remaining three components are stochastic. The first one, θ_i, is a permanent shock that individuals draw at the beginning of life-cycle and it stays fixed. The second one, η_{it}, is a persistent shock that follows a Markovian process, i.e., its distribution is conditional on the last period's realization (η_{it-1}). Lastly, ϵ_{it} is a transitory shock and drawn independently every period.

Linear Process The distributions of permanent and transitory shocks in the linear process are drawn from mean-zero normal distributions, i.e.,

$$\theta_i \sim \mathcal{N}(0, \sigma_\theta), \quad \epsilon_{it} \sim \mathcal{N}(0, \sigma_\epsilon). \tag{2.2}$$

The persistent shock follows an AR(1) process,

$$\eta_{it} = \rho\eta_{it-1} + \nu_{it}, \quad \nu_{it} \sim \mathcal{N}(0, \sigma_\nu), \quad \eta_{i(-1)} = 0, \tag{2.3}$$

where the persistence is governed by a single parameter, ρ. As a result, the persistence is the same for any values of the innovation, ν_{it}, or the lagged variable, η_{it-1}. Furthermore, the conditional distribution is assumed to be normal, and only its mean depends on the lagged value. Therefore conditional variance is the same for all values of the lagged variable. The distribution is symmetric, i.e., skewness is zero, and all higher moments, e.g., kurtosis, only depend on the variance parameter. Also, the distributions are assumed to be the same for all ages.

Nonlinear Process The nonlinear process is defined by quantile functions for each shock. The permanent, θ_i, and the temporary, ϵ_i, shocks are draws from,

$$\theta_i = Q_\theta(u_i^\theta), \quad u_i^\theta \sim U(0,1) \quad \text{and} \quad \epsilon_{it} = Q_\epsilon(u_{it}^\epsilon | age_t), \quad u_{it}^\epsilon \sim U(0,1), \tag{2.4}$$

where $Q_\theta(u_i^\theta)$ and $Q_\epsilon(u_{it}^\epsilon|age_t)$ are quantile functions, and $U(0,1)$ is the uniform distribution over the unit interval. Any distribution can be defined using its quantile function. For example, for the linear specification, the quantile functions would be just the inverse of the normal cumulative density function with the corresponding variance parameters.[2] We do not assume any parametric distributional form. Instead, we estimate the quantile function with a set of quantile regressions. The distribution of transitory shock, (ϵ_{it}), is allowed to depend on age.

The persistent shocks follow a Markovian process, defined by the following conditional quantile function,

$$\eta_{it} = Q_\eta(u_{it}^\eta | \eta_{it-1}, age_t), \quad u_{it}^\eta \sim U(0,1), \tag{2.5}$$

where we allow nonlinear dependence on η_{it-1} and age. Similarly, instead of relying on a known parametric distribution, such as normal distribution, we estimate the quantile function with a set of quantile regressions. The details of the estimation algorithm are in Section 2.2.

Since there are no parametric assumptions, the nonlinear process is much richer than the linear one. The persistence does not have to be linear, the shape and the moments of distributions are not restricted, and age dependency is allowed.

[2]For example, the linear persistent shock defined in equation (2.3) can be specified as

$$\eta_{it} = \rho\eta_{it-1} + \sigma_\nu \Phi^{-1}(u_{it}^\eta), \quad u_{it}^\eta \sim U(0,1),$$

where $\Phi(\cdot)$ is c.d.f. of standard normal distribution.

2.1 Data and Sample Selection

We use the data from the Panel Study of Income Dynamics (PSID) from 1968 to 2015. The data is annual until 1997 and biannually afterward. The PSID consists of two samples, one representing low-income families (SEO) and the other for the U.S. population. We focus on non-SEO households with married male household heads between ages 25 and 64. We drop observations with missing information for age, education, and state of residence. The wage rate is calculated by dividing annual real labor income by the annual number of hours worked. As it is standard practice in the literature, (e.g. Heathcote, Perri and Violante, 2010), we do not use an observation if the hourly wage rate is less than half of the minimum wage in a given state. Also, if there is an unusual jump in income, the observation is not used since the jump most likely comes from a measurement error. We keep all wage observations that we can link for 5 consecutive periods for annual and 3 consecutive periods for the biennial part of the PSID and construct a panel of wages indexed by age and time. The Appendix A provides further details.

In a first-stage regression, we regress male log wages on a set of state dummies, year dummies that are interacted with education and race categories, and cohort dummies. Cohort and race dummies capture the individual heterogeneity in productivity beyond the initial value of persistent shock in the model, while year dummies remove fluctuations in wages common to all individuals in a given year. We then regress the residual wages from this regression on a polynomial of ages to estimate the deterministic age profile, κ_t. Figure 1 plots the resulting wage profile. The deviations between residualized wages, ω_{it}, and κ_t are then used to estimate permanent, persistent, and temporary shocks.

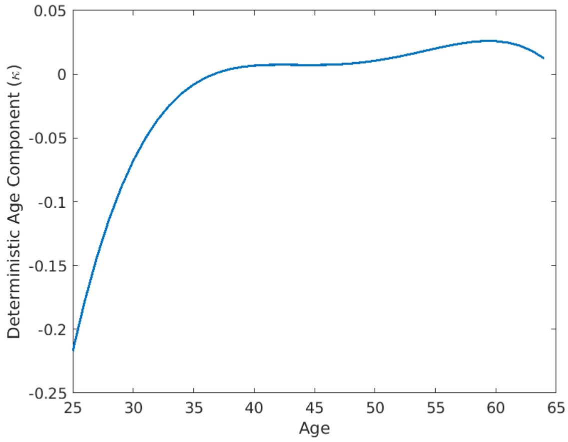


Figure 1: Deterministic age profile

2.2 Estimation

We approximate quantile functions with a set of polynomials. The quantile function of the persistent shock is specified as

$$Q_\eta(u_{it}^\eta|\eta_{it-1}, age_{it}) = \sum_{k=0}^{K_\eta} a_k^\eta(u_{it}^\eta)\varphi_k(\eta_{it-1}, age_{it}), \tag{2.6}$$

where t denotes the year, φ_k is a polynomial indexed by k, and $a_k^\eta(u_{it}^\eta)$'s are coefficients of polynomials for each quantile u_{it}^η. In practice, we use all interactions of constant and orthogonal polynomials of η_{it-1} and age_{it} up to 3rd degree, hence $K_\eta = (3+1)\times(3+1) = 16$.

The quantile function of the initial value of persistent shock is given by,

$$Q_{\eta 0}(u_{i0}^\eta|age_{i0}) = \sum_{k=0}^{K_{\eta 0}} a_k^{\eta 0}(u_{i0}^\eta)\varphi_k(age_{i0}), \tag{2.7}$$

where age_{i0} is the age in the first period of the panel data and $\varphi_k(age_{i0})$'s are constant and orthogonal polynomials of age up to a third degree, hence $K_{\eta 0} = 4$. Quantile functions of permanent and temporary shocks are similarly defined as

$$Q_\epsilon(u_{it}^\epsilon|age_{it}) = \sum_{k=0}^{K_\epsilon} a_k^\epsilon(u_{it}^\epsilon)\varphi_k(age_{it}), \tag{2.8}$$

and

$$Q_\theta(u_i^\theta) = a^\theta(u_i^\theta), \tag{2.9}$$

with $K_\epsilon = 3$ and $Q_\theta(u_i^\theta)$ does not have any polynomial because its distribution does not depend on any other variable.

We take a grid of quantiles $u \in \{u_0, u_1, \ldots, u_L\} \subset (0,1)$ and estimate parameters, $a_k(u)$'s of each quantile functions for the grid points. In practice, we set $L = 11$ and take an equidistant grid covering the unit interval. We approximate the function with linear splines for any points that are not on the grid. To avoid flat tails, we fit an exponential tail to both the upper and lower tails of the distribution. Therefore the complete specification of the approximated quantile function is given by,

$$\hat{Q}_v(\cdot, u_{it}) = \begin{cases} \sum_{k=0}^{K_v} a_k(u_1)\varphi_k(\cdot) + \frac{\ln(u_{it}/u^1)}{\lambda_1^v} \quad \text{if} \quad u_{it} \leq u^1 \\ \\ \sum_{k=0}^{K_v} \frac{u_{it}-u^l}{u^{l+1}-u^l}[a_k(u^{l+1}) - a_k(u^l)]\varphi_k(\cdot) + a_k(u^l)\varphi_k(\cdot) \\ \hfill \text{if} \quad u^l < u_{it} \leq u^{l+1} \\ \\ \sum_{k=0}^{K_v} a_k(u^L)\varphi_k(\cdot) - \frac{\ln((1-u_{it})/(1-u^L))}{\lambda_L^v} \quad \text{if} \quad u_{it} > u^L \end{cases}, \tag{2.10}$$

for each component, $v \in \{\theta, \eta, \epsilon\}$ and $\cdot$ is a placeholder for their corresponding argument of quantile function, λ_1^v, λ_L^v are parameters of exponential tails. We follow a quantile regressions-based stochastic EM algorithm provided in Arellano, Blundell and Bonhomme (2017).

We would like to run quantile regressions to estimate the parameters of polynomials and the maximum likelihood estimator for the exponential tails, however, the components of the wage process are not observable. Only their sum, the residual wages ω_{it}'s, is observable. Let Λ denote a set of all the parameters i.e., $\Lambda \equiv \{a_k^v(u_i), \lambda_1^v, \lambda_L^v\}$ for all $u_i, i = 1, \ldots, L$, and for all components $v \in \{\theta, \eta, \epsilon\}$, then the true parameter values solves a set of optimization problems coming from moment conditions of quantile regressions,

$$\Lambda = \arg\min_{\tilde{\Lambda}} \mathbb{E}\left[R(\theta_i, \eta_i, \epsilon_i, \omega_i, \tilde{\Lambda})\right], \tag{2.11}$$

where $R(\cdot)$ consists of moment condition for quantile regressions and maximum likelihood estimator for exponential tails, $\tilde{\Lambda}$ is optimization argument, $\eta_i \equiv \{\eta_{it}\}_{t=0}^{T}, \epsilon \equiv \{\epsilon_{it}\}_{t=0}^{T}, \omega_i \equiv \{\omega_{it}\}_{t=0}^{T}$ are vectors of variables over time. It is not possible to minimize the sample counterpart of the objective function because the components of the wage process , i.e., $\theta_i, \eta_i, \epsilon_i$, are not observable. Applying the law of iterated expectations will give,

$$\Lambda = \arg\min_{\tilde{\Lambda}} \mathbb{E}_{\omega}\left[\mathbb{E}_{\theta_i,\eta_i,\epsilon_i,|\omega_i,\tilde{\Lambda}}\left[R(\theta_i,\eta_i,\epsilon_i,\omega_i,\tilde{\Lambda})\right]\right], \tag{2.12}$$

where the inner expectation is taken with the distribution of the unobservable variables i.e., θ, η, ϵ conditional on the observable variables, residual wages ω. This conditional distribution depends on the parameters because it is defined by quantile functions specified in equation 2.10.

The dual role of parameters is visible in equation 2.12, in conditional expectation and the moment conditions. The idea of EM is fixing the parameter values for the first role and solving the optimization for the second role because it would be easy to do that and use the resulting value in the first step with iterations until convergence. It is essentially a minimization algorithm that converges to a local minimum.

The stochastic EM algorithm starts with a set of parameter guesses, say $\hat{\Lambda}^{(0)}$ follows the iteration of following two steps with s being the iteration index,

1. *E step:* Given a set of parameters $\hat{\Lambda}^{(s)}$ in iteration step s, simulate M many sample of unobservable variables with MCMC sampling methods,

   $$\theta_{im}, \eta_{im}, \epsilon_{im} \sim f(\theta,\eta,\epsilon|\omega,\hat{\Lambda}^{(s)}), \quad \text{for} \quad i=1,\ldots,N, m=1,\ldots,M$$

   where $f(\theta,\eta,\epsilon|\omega,\hat{\Lambda}^{(s)})$ is the probability density function of unobservable conditional on the data and can be obtained using their quantile functions in equation 2.10[3].

[3] Inverse of a quantile function is cumulative distribution function and its derivative gives the probability density function. As an example, for the persistent component η_{it}, it is given by

$$f(\eta_{it}|\eta_{it-1}, age_{it}) = \frac{\partial Q_{\eta}^{-1}(\eta_{it}|\eta_{it-1}, age_{it})}{\partial \eta_{it}}.$$

2. *M step:* Given the sample of unobservable variables, we can evaluate the sample counterparts of the moment conditions in equation 2.12. In other words, the inner expectation is approximated with MCMC sampling. We update the parameter values by solving the optimization problem,

$$\hat{\Lambda}^{s+1} = \arg\min_{\tilde{\Lambda}} \frac{1}{NM} \sum_{i=0}^{N} \sum_{m=0}^{M} R(\theta_{im}, \eta_{im}, \epsilon_{im}, \omega_i, \tilde{\Lambda}).$$

   Go back to *E step* with a new set of parameters and keep iterating for until a convergence.

Because of the sampling error in the *E step*, the parameter values in the iteration after converging around the true parameter values keep fluctuating around them. Arellano, Blundell and Bonhomme (2017) show that the resulting series of iterations converge to a stationary distribution around the truth. We set $M = 1$ and iterate $S = 300$ times. As the final estimates, we take the average of estimated parameters in the last 150 iteration to reduce the noise introduced by the sampling error in the stochastic EM algorithm.

2.2.1 Combining Annual and Biannual Data

The PSID started to be collected annually but switched to biennial in 1997. We combine both parts of the PSID to estimate an annual process. This is not a problem for the permanent and transitory components $(\theta_i, \epsilon_{it})$, because their distributions are i.i.d over time, and there is no past dependence. However, the distribution of the persistent component, η_{it}, is Markovian, and it depends on the past period in an annual timeline, but the data is not available in the biennial part.

Our approach to dealing with unobservable latent variables can be easily applied to the persistent component in the missing periods in the biennial part. In the annual part, we can observe the residual wages ω_{it} for all periods $t = 0, 1, \ldots, T$ and the latent variables are $\theta_i, \eta_{it}, \epsilon_{it}$ for all annual time periods $t = 0, 1, \ldots, T$. In the *E step*, the p.d.f. of the unobservable variables is given by,

$$f(\theta_i, \eta_{i0}, \eta_{i1}, \ldots, \eta_{iT} | \omega_{i0}, \omega_{i1}, \ldots, \omega_{iT}, age_{it}) \propto \prod_{t=0,1,\ldots,T} f(\epsilon_{it}|age_{it}) \times \prod_{t=1,2\ldots,T} f(\eta_{it}|\eta_{it-1}, age_{it}) \times f(\eta_{i0}|age_{it}) \times f(\theta_i),$$

with $\epsilon_{it} = \omega_{it} - \eta_{it} - \theta_i$ and each p.d.f function on the right-hand side is specified by its corresponding quantile function [4]. Using these p.d.f functions, we simulate unobservable latent variables annually in the *E step*, and the rest of the algorithm follows as it was described.

2.2.2 Linear Shocks

We estimate the linear process defined in the equations 2.2 and 2.3 with the generalized method of moments (GMM) using variances and covariances of wage residuals across periods as moment conditions. The covariances of wage shocks are calculated by pooling all waves of the PSID. The covariances between age t and $t+s$ is given by,

$$\widehat{cov}(\omega_t, \omega_{t+s}) = \frac{1}{N_{(t,t+s)} - 1} \sum_{i=1}^{N_{(t,t+s)}} \omega_{it}\omega_{it+s}, \tag{2.13}$$

where $N_{(t,t+s)}$ is the number of observation pairs available in the data for age t and age $t+s$.

The covariance term implied by the linear specification is given by,

$$cov(\omega_t, \omega_{t+s}) = \rho^s \frac{1-\rho^{2(t+1)}}{1-\rho^2}\sigma_\nu + \sigma_\theta + 1(s=0)\sigma_\epsilon, \tag{2.14}$$

where $1(s=0)$ is an indicator function for $s=0$.

We define vectors moments for data, $\widehat{M}$, and the linear model, M, and minimize the distance between them, i.e.,

$$\min_{\sigma_\theta, \sigma_\nu, \sigma_\epsilon} (\widehat{M} - M)A(\widehat{M} - M)', \tag{2.15}$$

where the diagonal elements of A is set to square root of number observations, $N_{(t,t+s)}$, to corresponding moment condition.

[4] See footnote 3

3 Results: Nonlinear Shocks

In the estimations, we use data after 1980 since there has been a substantial increase in income inequality since then. We report the results with pre-1980 data in Appendix C. In this section, we focus on two novel aspects of nonlinear shocks: i) the resetting effects of unusual shocks through nonlinear persistence and ii) changes in persistence along the life cycle.

The persistence of the nonlinear shocks is calculated as;

$$\rho(\eta_{it-1}, u_{it}, age_{it}) \equiv \frac{\partial Q_{\eta}(u_{it}|\eta_{it-1}, age_{it})}{\partial \eta_{it-1}}, \tag{3.1}$$

which measures the effect of past realization (η_{it-1}) on current realization at different quantiles of the current period's innovation, u_{it}.

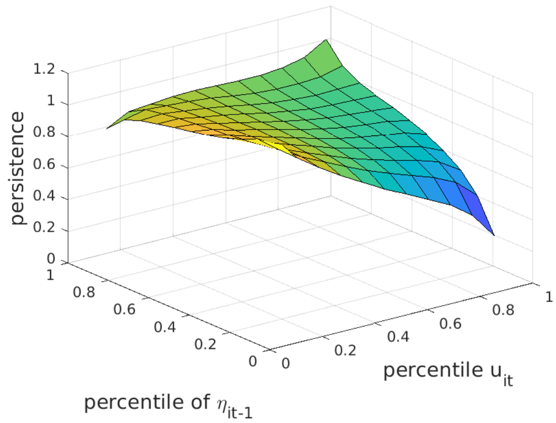
Figure 2 shows the estimated nonlinear persistence of η_{it} process averaged over all ages (Panel a) and for ages 25, 40, and 60 (Panels b-d). We plot the persistence for different percentiles of η_{it-1}, i.e., position in the distribution of last period's realization, and the percentiles of u_{it}, the innovation, i.e., its high values means the worker is hit by better shocks and end up in the high values of η_{it}.

The nonlinear persistence has a saddle shape, i.e., it is lower when η_{it-1} is low and u_{it} is high, and vice versa. This means that when a worker with high productivity (high η_{it-1}), is hit by a bad shock (low u_{it}), i.e., ends up in the lower tail of the conditional distribution of η_{it} the effect of the past is lower. Similarly, if a worker with low productivity is hit by a good shock, the past matters less because of low persistence. In other words, unusual shocks have the property of deleting/resetting the effect of the past.

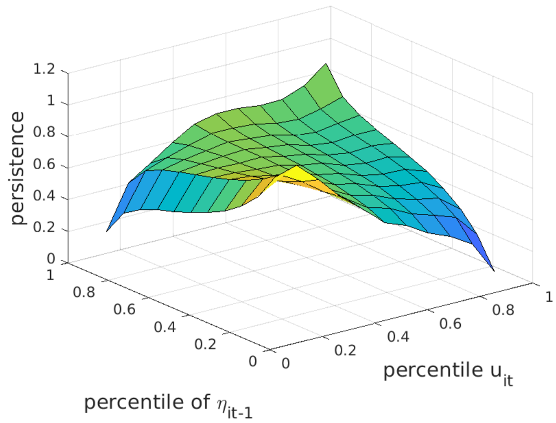
The nonlinear persistence plots for different ages in Figure 2 show that the wiping history effect of unusual shock is somewhat different for different ages. Where the saddle shape is very pronounced at age 25, the persistence becomes flatter later. Therefore the wiping history effect of the unusual shock is more pronounced at earlier ages.

The overall persistence level is also changing over the lifecycle. Figure 3 he average

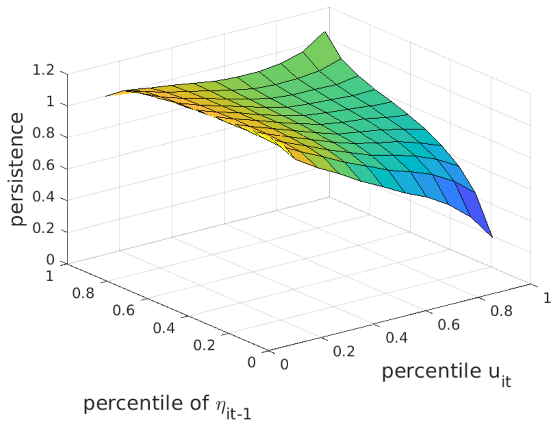
Figure 2: Nonlinear persistence for different ages



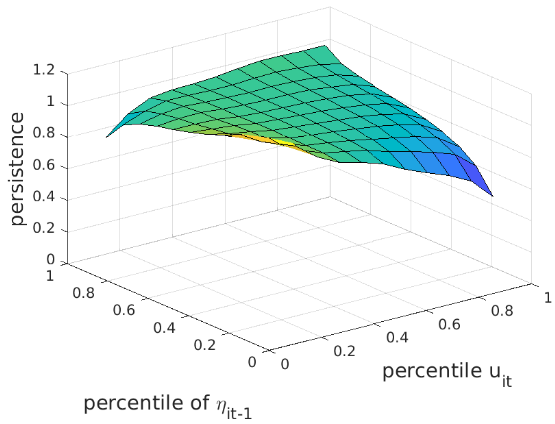
(a) Nonlinear persistence (Average)



(b) Age 25



(c) Age 40



(d) Age 60

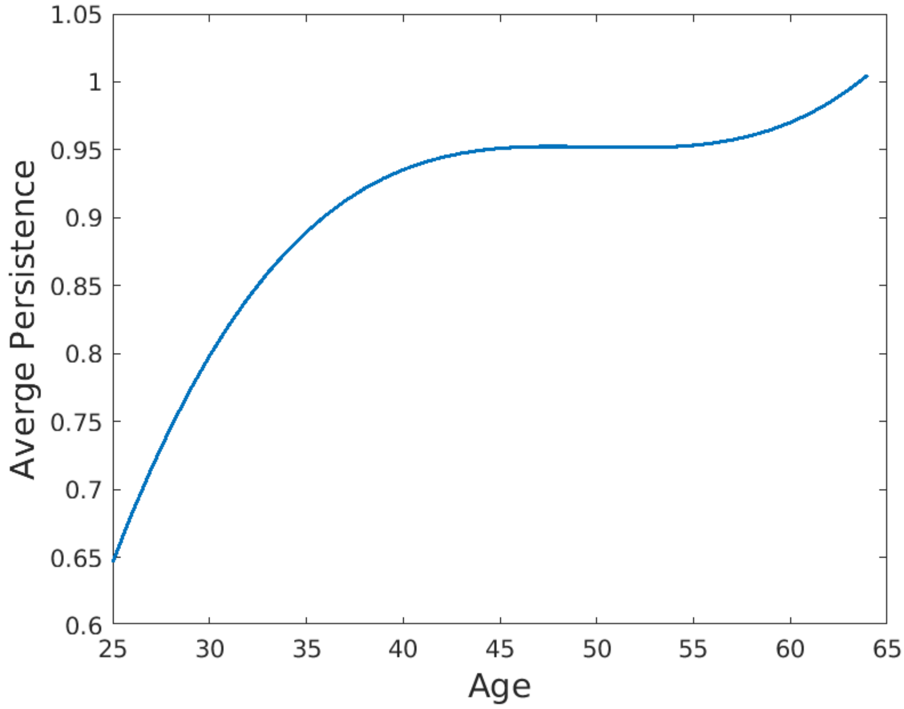


Figure 3: Average persistence by age

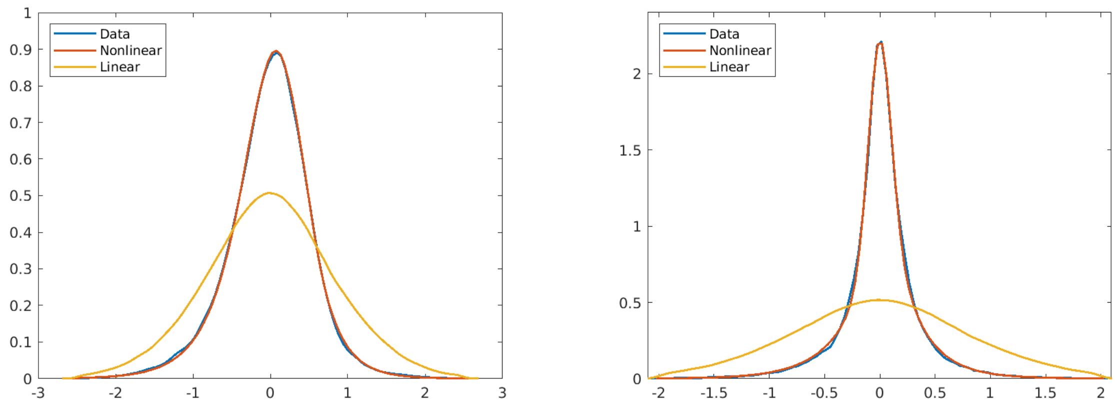
persistence, $\mathbb{E}(\rho(\eta_{it-1}, u^{\eta}_{it}))$ at each age. The average persistence is as low as around 0.65 initially at age 25, then it rises sharply until it flattens after around age 40. The lower persistence of unusual shocks pushes down the average persistence for earlier ages. [5]

4 Results: Nonlinear vs. Linear Shocks

In this section, we compare the estimation results of linear and nonlinear wage processes. Figure 4 illustrates the marginal distribution of estimated productivity both for levels and one-year changes. In the data, both wages and wage changes have a leptokurtic shape, with a higher peak and fatter tails. The nonlinear process can match the distribution perfectly, while the linear process fails to do so due to restricted normality assumption.

[5] Karahan and Ozkan (2013) document that the persistence of income has an inverted-U shape, starting quite low at the beginning of the life cycle.

Figure 4: Marginal Distributions implied by the estimated processes



(a) Wage Levels

(b) One Year Change in Wage

Figure 5a plots the variance of log wages over the life cycle implied by both processes and the data. For the data part, we calculate the variance of each age for each year of the PSID and take the average over the years. Although the nonlinear process initially has a slightly higher variance than the data, it later follows the data pretty closely and captures the trend change around age 50. Notice that these variances in the data are included in the moment conditions of the linear process, so the estimation minimizes the distance by construction. However, this is not the case for nonlinear process and it still performs better.

Figure 5b plots a quantile based variance measure given by,

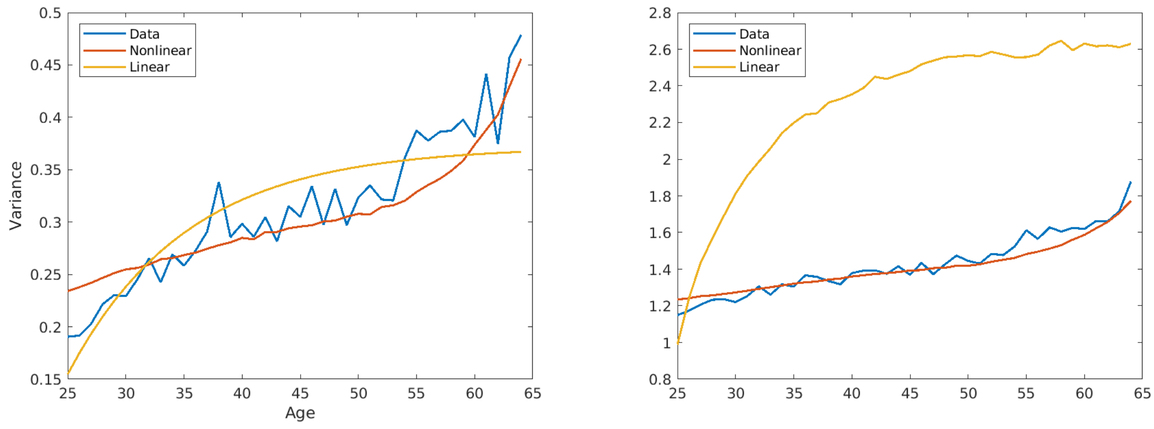
$$\sigma_Q \equiv Q(1-u) - Q(u) \quad \mathit{for} \quad u \in (0, 0.5), \tag{4.1}$$

which reports the distance between a high and a low quantile. Again, the nonlinear process fits the data very well, but the linear process is completely off, mainly because normal distribution has thin tails.

Finally, we report conditional moments. In Figure 6, we plot variance, skewness, and kurtosis for estimated processes and the data conditional of the quantiles of past value. We calculate moments using simulated histories of productivity.[6]

[6]Although such moments can be calculated exactly for the linear model, we provide these estimates for the sake of validation. Any fluctuations reflect estimation errors.

Figure 5: Variance of productivity implied by estimated processes over life cycle

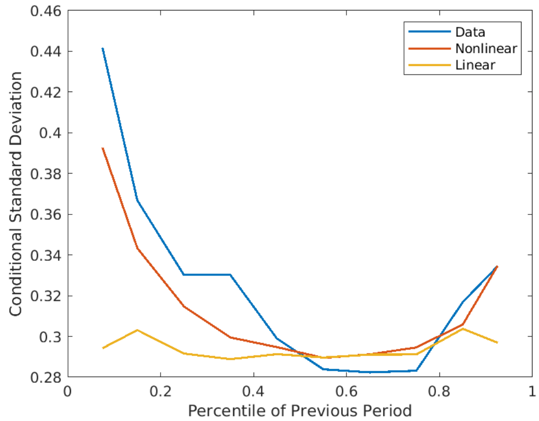


(a) Variance of log wages by age

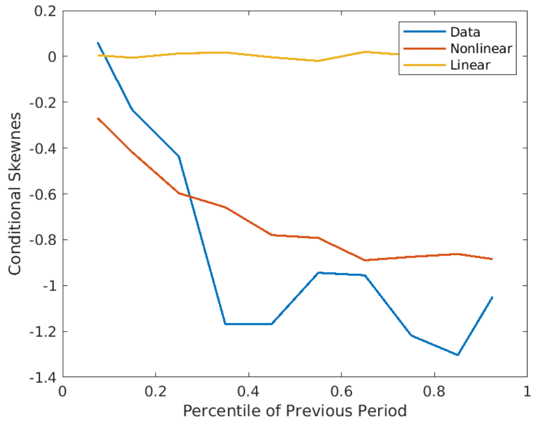
(b) Quantile-based variance measure by age

Figure 6a shows the standard deviation, which is U-shaped. Both low and high-productive workers are subject to more diverse risk than medium-level productive workers. We can see that the nonlinear process is able to capture the pattern. The linear process has constant variance by definition, the fluctuations are because of estimation error. Figure 6b shows that there is negative skewness for almost all quantiles and it is decreasing. Again while the nonlinear process can capture it, the linear process has zero skewness by normality assumption. The more productive workers are subject to much more negative risk than low productive workers, which can affect social insurance mechanisms in the economy. Finally, Figure 6c illustrates the conditional kurtosis has an inverted U shape, and similarly, it is captured by the nonlinear process and constant by definition in the linear model. Overall the message is the nonlinear process can capture higher moments of the data, not only second moments like the linear process.

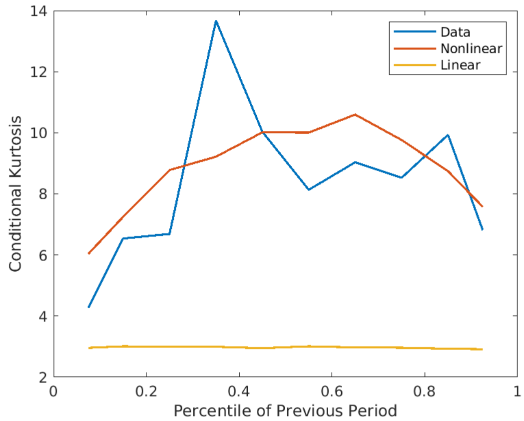
The first row of Table 1 shows the estimated parameters for the linear specification. In Table 2, we compare the variance decomposition of estimated processes across components. In both processes, roughly half of the variance comes from persistent components, but the contribution of temporary and permanent components is substantially different. In the nonlinear process, 40.3% of the variance comes from permanent components, whereas in the linear one, 32.4% comes from the transitory part.



(a) Conditional Standard Deviation



(b) Conditional Skewness



(c) Conditional Kurtosis

Figure 6: Moments of predicted wages conditional on the past

Table 1: Parameters of linear processes

	ρ	σ_ν	σ_α	σ_ϵ
Estimated	0.962	0.019	0.046	0.109
Constructed	0.952	0.020	0.126	0.030

Table 2: Variance decomposition of estimated processes

	$var(\omega_{it})$	$var(\eta_{it})$	$var(\alpha_i)$	$var(\epsilon_{it})$
Nonlinear	0.312	0.156	0.126	0.030
	100%	50.0%	40.3%	9.7%
Linear	0.338	0.181	0.046	0.109
	100%	53.8%	13.7%	32.4%

In the analysis of the quantitative model, we use an alternative linear process with the same variance composition to focus on the effects of nonlinearity and nonnormality instead of different sources of risk. To construct this alternative process, we set the variance of normal distributions of temporary and permanent components to the variances in the nonlinear process. For the persistent part, we fit an AR(1) process (eq. 2.3) by matching the total variance of the nonlinear one exactly and minimizing the distance to the age profile of variance. Table 1 includes the constructed linear process. The constructed linear process performs similarly in matching data moments that we discussed so far in this section.

5 Model

We build a standard life-cycle model with idiosyncratic income risk, incomplete markets, and endogenous labor supply decisions (e.g. Kaplan, 2012; Guner, Lopez-Daneri and Ventura, 2016; ?). The economy is stationary, and the model period is one year. Each period, individuals decide how much to save on a single risk-free asset and how much to work. Individuals are ex-ante heterogeneous in their productivity through permanent shocks and ex-post due to idiosyncratic productivity shocks they face throughout the life cycle.

A government taxes individuals and provides transfers. There are three different taxes on household incomes. First, households face a progressive tax schedule on their total income, which approximates the Federal Income taxes in the U.S. Second, there is an additional flat tax on total income, which captures local taxes in the U.S. Finally, there is an additional flat tax on capital income, which approximates the corporate income tax. Means-tested government transfers are represented by a nonlinear transfer scheme that maps household income into transfers received. A social security system also pays a constant benefit after retirement and is financed through a flat payroll tax on the labor income. The social security system is in balance in each period. A representative firm with constant returns to scale production function determines factor prices.

Technology The output is produced by a representative firm that hires capital and labor. It operates a constant return-to-scale technology given by

$$Y = F(K, L) = AK^{\alpha}(LX)^{1-\alpha},$$

where L is aggregate labor input in efficiency units, and A and X parametrizes Hicks-Neutral and labor augmenting technology level, respectively. X grows at a constant rate g to capture the GDP growth. The parameter α measures the elasticity of output with respect to capital. Each period capital depreciates at a constant rate δ.

Preferences Every period, a continuum of agents enters the economy. Conditional on being alive at t, they survive to period $t+1$ with a probability s_t and can live up to T_{max} periods. Population grows at a constant rate n. The demographic structure is stationary, and at any given time, the population share of agents at age t, denoted by μ_t, is determined by the recursion, $\mu_{t+1} = \frac{s_t}{1+n}\mu_t$.

The preferences of the agents are defined on the consumption (c_t) and working time (l_t).

Agents maximize their lifetime utility

$$\mathbb{E}\sum_{t=0}^{N}\beta^t\left(\prod_{i=1}^{t}s_i\right)\left[\log(c_t)-\phi\frac{l_t^{1+\frac{1}{\gamma}}}{1+\frac{1}{\gamma}}\right],$$

where β is the discount factor.

The price of labor per productivity unit is denoted by w, which is common to all individuals. Productivity endowment of an individual i at age t is given by $\omega(\Omega_{it})$, where $\Omega_{it}=(\theta_i,\eta_{it},\varepsilon_{it})$. The first component, θ_i, is a permanent shock drawn in the beginning of the life cycle, and it sticks to the agent through the life cycle. The second one, η_{it}, is a persistent shock, while ε_i is a transitory one. Agents enter the model with zero asset holdings and can work until the mandatory retirement age t_R. After retirement, the productivity endowment is set to zero. Each period, agents make decisions on consumption, labor hours, and saving on a risk-free asset (capital).

The budget constraint of agent i at age t with Ω_{it} is given by

$$c_t+a_{t+1}\leq a_t+I_t-\tau_p w\omega(\Omega_{it})l_t-T(I_t)+TR(I_t), \tag{5.1}$$

where

$$I_t=w\omega(\Omega_{it})l_t+ra_t+B_t.$$

In equation 5.1, a_t is an asset holding at age t, r is the one-period return of the risk-free asset, and l_t denotes the labor hours. B_t is the social security payment which is zero before the retirement age, and then it equals to a fixed benefit for all agents. τ_p is a flat tax on the labor income to finance social security benefits. $T(I_t)$ and $TR(I_t)$ are the total amount of the income taxes paid and the transfers received by an individual as a function of their total pre-tax income.

To define the dynamic programming problem associated with individuals' lifetime maximization problem, we first transform variables to remove the effects of secular growth and indicate transformed variables with symbol $(\hat{.})$. An individual's state is given by the pair $x=(\hat{a},\Omega)$, $x\in\mathbf{X}$, where $\hat{a}$ are current (transformed) asset holdings and Ω are the productivity shocks. The set $\mathbf{X}$ is defined as $\mathbf{X}\equiv[0,\bar{a}]\times\mathbf{\Omega}$, where $\bar{a}$ stands for an upper bound on (normalized) asset holdings. We denote total taxes at state $(x, t\ )$ by $T(x, t)$. Consequently, optimal decision rules are functions for consumption $c(x, t)$, labor $l(x, t)$, and period asset holdings $a(x, t)$ that solve the following dynamic programming problem:

$$V(x, t) = \max_{(\hat{l}, \hat{a}')} u(\hat{c}, t) + \beta s_{t+1} E[V(\hat{a}', \Omega', t+1)|x] \tag{5.2}$$

subject to

$$\hat{c} + \hat{a}'(1+g) \leq \hat{a}(1+\hat{r}) + (1-\tau_p)\hat{w}\omega(\Omega, t)l + T\hat{R}_j(x, t) - \widehat{T}(x, t), \tag{5.3}$$

and

$$\hat{c} \geq 0, \quad \hat{a}' \geq 0, \quad \hat{a}' = 0 \ \text{ if } j = N, \ V(x, T_{\max} + 1) \equiv 0 \tag{5.4}$$

Labor Productivity Shocks The labor market productivity of an age-t individual is given by

$$\log \omega_{it} = \kappa_j + \theta + \eta_{it} + \varepsilon_{it}. \tag{5.5}$$

Individuals face wage processes that we estimate in Section 2. The linear wage process is defined by three mean zero normal distributions with standard deviations $\sigma_\theta, \sigma_\varepsilon$, and σ_v and a persistence parameter ρ. When the wage process is non-linear, individuals draw θ_i and ε_{it} from non-parametric cumulative probability distributions, defined by their quantile functions, while the persistence component follows a Markovian process, defined by a conditional quantile function.

5.1 Benchmark Economy

We calibrate the benchmark economy with the constructed linear and nonlinear wage processes estimated in Section 5. We first set several parameters to their data counterparts or to commonly-used values in the literature. The population growth in the US between 1960 and 2015 was 0.43% so we set $n = 0.043$, based on 1960-2015 period.[7] Similarly, we set $g = 0.0202$, the average growth rate of real GDP per capita fro the 1960-2015 period.[8] The survival probabilities are set according to the U.S. Life Tables for the year 2005.[9] For preferences, we set the intertemporal elasticity of substitution for labor supply, γ, to 1. The early estimates of γ, such as MaCurdy (1981); Altonji (1986), were much smaller. There has emerged, however, a consensus that these early estimates were biased downwards, see, e.g., Imai and Keane (2004); Domeij and Flodén (2006); Chetty (2012). For aggregate labor supply elasticity, Keane and Rogerson (2015) suggest values of γ in excess of 1. Finally, we set $\alpha = 0.379$, the share compensation in GDP.[10] The upper panel in Table 3 shows all the parameters that are chosen based on a priori information.

Then we are left us with three parameters to calibrate, ϕ, β, and δ. We choose ϕ so that on average hours worked is 1/3. Following the standard practice in the literature, β is selected to match the average capital-to-output ratio of 3.27 for the 1950-2015 period.[11] Finally, we set $\delta = 0.07$ such that the model economy matches the investment to GDP ratio of 17.36%, again the average value for the 1950-2015 period.[12] While we allow these parameters to differ depending on the underlying productivity process, the calibrated parameters are identical for β and δ, and there is only a minor difference in disutility from work, ϕ. The lower panel in Table 3 shows the calibrated parameters.

The parametric functional form for the federal income taxes is given by the following

[7]The annual population growth rate is calculated based on FRED data series SPPOPGROWUSA (Population Growth for the United States, Percent Change at Annual Rate, Annual, Not Seasonally Adjusted), https://fredhelp.stlouisfed.org.

[8]The real GDP per capita growth rate is calculated based on FRED data series SA939RX0Q048SBEA_PCH (Real gross domestic product per capita, Percent Change, Annual, Seasonally Adjusted Annual Rate), https://fredhelp.stlouisfed.org.

[9]National Vital Statistics Reports, Volume 58, Number 10, 2010.

[10]We use FRED data series SLABSHPUSA156NRUG (Share of Labour Compensation in GDP at Current National Prices for United States, Ratio, Annual, Not Seasonally) Adjusted), https://fredhelp.stlouisfed.org

[11]The calculations are based on NIPA Tabls.1.1, 1.2, 1.1.2, and 1.1.5. We use fixed assets as the notion of capital corresponding to the model environment.

[12]The investment-output ratio, I/Y, is computed as the grosdover GDP, where both figures are taken from the National Accounts. We use the average value for 2010-2015, 0.206, as the target.NIPA Table 1.1.5.

Table 3: Parameter values (Common to all specifications unless stated otherwise)

Pop. growth (n)	1.043	U.S Data
Labor eff. growth (g)	2.02	U.S. Data
Frisch Elasticity (γ)	1	Literature
Capital Share (α)	0.379	U.S. Data
Income tax (τ_l)	0.05	Guner, Lopez-Daneri and Ventura (2016)
Capital income tax (τ_k)	0.057	Calibrated, see the text
Social security tax (τ_p)	0.12	Calibrated, see the text
Depreciation rate (δ)	0.02	Calibrated to match $I/K = 0.17$
Discount factor (β)	0.97	Calibrated to match $K/Y = 3.27$
Labor disutility (ϕ), Nonlinear Model	7.5	Calibrated to match hours to 1/3
Labor disutility (ϕ), Linear Model	7.2	Calibrated to match hours to 1/3
Tax Function Parameters		
Tax function level (λ)	0.911	Guner, Kaygusuz and Ventura (2014)
Tax function curvature (τ)	0.053	
Transfer Function Parameters		
b_0	0.121	Guner, Rauh and Ventura (2020)
b_1	-2.122	
b_2	-4.954	
b_3	0.044	

expression,

$$T(I) = I \times t(\tilde{I}) \text{ with } \quad t(\tilde{I}) = 1 - \lambda \tilde{I}^{-\tau},$$

where $\tilde{I} = I/\bar{I}$ is income relative to the mean income in the economy, $\bar{I}$. This functional form is used in the literature by Bénabou (2002); Heathcote, Storesletten and Violante (2014) and others. The tax function is characterized by two parameters, τ controls the progressivity, and λ controls the general level of federal income taxes. Benchmark parameter values are taken from Guner, Kaygusuz and Ventura (2014), estimated using the micro-data from Internal Revenue Service for the year 2000 (Statistics of Income Public Use Tax File). The estimation is done for all households. The estimates of parameters are $\tau = 0.053$ and $\lambda = 0.911$, which imply that a household around mean income faces an average tax rate of about 8.9% and a marginal tax rate of 13.73%. At five times the mean household income level in the IRS data (about $265,000 in 2000 U.S. dollars), the average and marginal rates amount to 16.35% and 20.78%, respectively.

Guner, Kaygusuz and Ventura (2014) show that the local tax rates are essentially flat, ranging from 4% around mean income to about 5.3% at the top one percent of household income. Therefore, we $\tau_l = 0.05$. We use τ_k to proxy the U.S. corporate income tax. We set $\tau_k = 0.057$ to reproduce the observed level of tax collections out of corporate income taxes after the major reforms of 1986. Such tax collections averaged about 1.91% of GDP for 1995-2008 period. Finally, we calculate $\tau_p = 0.12$, as the (endogenous) value that generates an earnings replacement ratio of about 53%.[13]

Following Guner, Rauh and Ventura (2020), transfers are also represented by a parametric function, the total amount of transfers received by an agent with income I is given by $TR(I) = \bar{I} \times tr(\tilde{I})$ where $\tilde{I}$ is relative income with respect to mean income and $tr(\tilde{I})$ is the amount of transfers normalized by the mean income, $\bar{I}$. The functional form of $tr(\tilde{I})$ is given by

$$tr(\tilde{I}) = \begin{cases} e^{b_1} e^{b_2 \tilde{I}} \tilde{I}^{b_3} & \text{if} \quad \tilde{I} > 0 \\ b_0 & \text{if} \quad \tilde{I} = 0 \end{cases} \tag{5.6}$$

[13]This is the value of the median replacement ratio in the mid-2000s for 64-65-year-old retirees, according to Biggs and Springstead (2008).

Transfer functions is characterized by four parameters. b_0 has a clear interpretation, it controls the amount of transfer paid to agents with zero income. b_1 and b_2 governs the levels and the convexity of the function. Lastly, b_3 controls the asymptotic behavior of the function close to the zero income level.

Inequality Table 4 shows the distribution of earnings in the model and the data. The model economy, with a linear or a non-linear wage process, does a good job of generating the earnings inequality observed in the data. Both versions have a hard time generating the share of earnings going to the top 1% of earners.

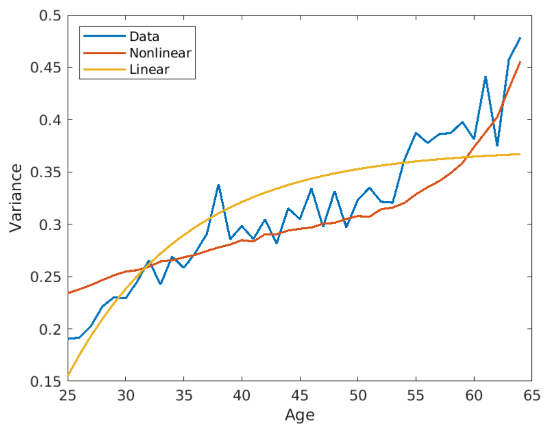
Figure 7a shows the variance of log wages, hours, and earnings along the life cycle. Except for initial ages, the variance of log wages implied by the nonlinear process follows the data very closely, while the concave shape implied by the linear process is at odds with the data. In Figure 7b, the variance log hours implied by the model is not able to capture the slight U-shaped pattern in the data with a linear or nonlinear process, but the model outcome with the non-linear process is able to capture overall shape better after around age 30. Finally, in Figure 7c, the variance of log earnings evolves much closer to the data with the non-linear process, but the level is higher due to the higher variance of log hours.

6 Value of Social Insurance

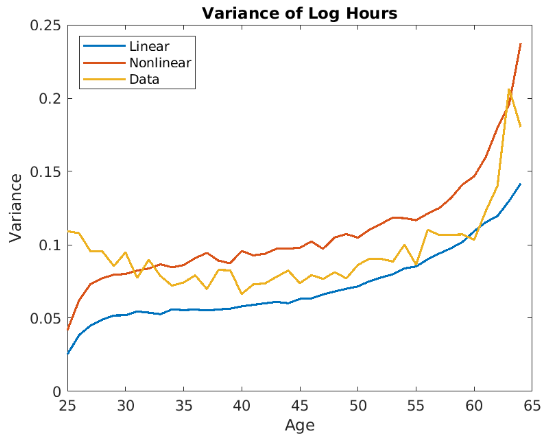
Do different income processes matter for how much households value social insurance? To answer this question, we compare the benchmark economy with alternatives economies without taxes and transfers:

1. We first replace progressive income taxes with a proportional income tax. The proportional tax rate generates the same level of government tax revenue as in the benchmark economy, about 9% for both specifications of wage shocks.

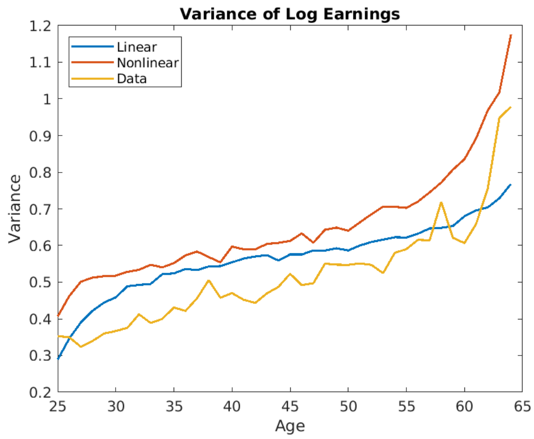
2. We then eliminate the means-tested transfers and lower income taxes by increasing λ



(a) Variance of log wages



(b) Variance of log hours



(c) Variance of log earnings

Figure 7: Variance of log wages, hours and earnings over life-cycle

Table 4: Earnings Distribution

	PSID	**Nonlinear**	**Linear**
0 - 20%	6.2%	5.4%	5.5%
40-20%	11.7%	11.8%	10.7%
60-40%	16.2%	17.0%	16.0%
80-60%	22.1%	23.5%	23.4%
80-100%	43.7%	42.25%	44.3%
Top 10%	28.7%	26.4%	27.7%
Top 5%	18.6%	16.1%	16.8%
Top 1%	7.0%	4.8 %	4.8%
Gini	0.37	0.37	0.39

to achieve revenue neutrality. For both specifications of wage shocks, λ increases from 0.91 to 0.92. This implies that, on average, income taxes increase by 1% points for all households.

3. Finally, we introduce a proportional tax and eliminate the means-tested transfers. Now the revenue-neutrality requires a 7.5% proportional tax for both specifications of wage shocks.

For each experiment, we find welfare gains and losses for each individual, calculated as the percentage change in each period's consumption that would leave the individual indifferent between the new economy and the benchmark. Table 5 shows the results where we report average welfare gains and losses for everyone and households in different points of the income distribution points in the benchmark economy. Table 6 shows the same statistics by the permanent types (θ_i).

On average, households prefer an economy with a proportional income tax. The mean welfare gain is about 0.68% with nonlinear wage shocks and 0.35% with constructed-linear wage shocks. The gains are mainly due to larger output that results from lower taxes on high-productivity households (the aggregate output increase by 8% in both economies). When we eliminate the means-tested transfer, there is also a welfare gain with nonlinear shocks. But with linear shock, there is a slight welfare loss. The difference between the two specifications is sharper when we move to a proportional tax to eliminate the welfare state. With nonlinear shocks, there is a sizable welfare gain of about 0.51%, while with linear shocks, there is a welfare off of about 0.15%.

Table 5: Welfare Gains by Position in Benchmark Income Distribution

		Quantiles of Lifetime Earnings in Benchmark								
	Mean	Bottom 10%	Bottom 20%	20% - 40%	40% - 60%	60% - 80%	Top 20 %	Top 10%	Top 5%	Top 1%
	Proportional Taxes Taxation									
Nonlinear	0.68	-2.40	-2.33	-1.46	-0.03	1.80	5.40	7.06	8.51	10.90
Linear	0.35	-2.27	-2.53	-2.15	-0.66	1.51	5.57	7.32	8.85	11.53
	No Means-tested Transfers									
Nonlinear	0.18	-5.49	-3.38	0.25	1.09	1.39	1.54	1.57	1.62	1.62
Linear	-0.02	-7.90	-4.97	0.18	1.23	1.62	1.84	1.92	1.97	2.08
	Proportional Taxation and No Means-tested Transfers									
Nonlinear	0.51	-9.93	-7.16	-1.54	1.03	3.24	6.99	8.68	10.18	12.57
Linear	-0.15	-13.29	-9.73	-2.51	0.57	3.30	7.63	9.46	11.05	13.88

Table 6: Welfare Gains by Permanent Types

		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	0.68	-2.58	-2.40	-2.29	-1.61	-0.69	0.20	1.33	2.76	3.66	4.62	8.03
Linear	0.35	-2.21	-2.46	-2.22	-1.79	-1.04	-0.12	0.74	2.04	3.71	4.73	6.93
	No Means-tested Transfers											
Nonlinear	0.18	-5.70	-2.71	-1.43	-0.47	0.43	0.86	1.10	1.33	1.39	1.48	1.59
Linear	-0.02	-7.32	-3.82	-2.12	-1.06	0.20	0.77	1.12	1.49	1.69	1.73	1.84
	Proportional Taxation and No Means-tested Transfers											
Nonlinear	0.51	-10.48	-6.39	-4.64	-2.61	-0.51	0.96	2.40	4.12	5.09	6.15	9.66
Linear	-0.15	-12.48	-8.08	-5.56	-3.75	-1.29	0.44	1.79	3.61	5.56	6.64	8.99

When we reduce redistribution, either by eliminating progressivity or means-tested transfers, households at the bottom of the income distribution in the initial benchmark lose, while those at the top gain. But how much they gain and lose depends critically on the underlying productivity process. When shocks are nonlinear, the households at the top of the income distribution gain *less* from a move to less redistribution. The households at the top 1% of income distribution in the initial steady state, for example, gain 10.90% with a move to a proportional income tax and 12.57% when transfers are also eliminated. The gains with linear shock are higher, 11.53% and 13.88%, respectively. Differential gains do not just happen for those at the very top of the income distribution; it is also the case for households in the top 40% of the income distribution in the initial steady state. With nonlinear dynamics, richer households have a higher chance of falling from the top of the income distribution, making progressivity more valuable for them.

In contrast, with nonlinear shocks, poor households lose *less* from a move to less redistribution. They potentially face higher taxes and do not enjoy means-tested transfers in the new steady state. But they are also more likely to move up in the income distribution and benefit from a proportional tax. As a result, the welfare losses for households at the bottom 10% of the income distribution are 9.93% with nonlinear shocks, while they are 13.29% with linear ones.

7 Conclusions

In this paper, we use data from the Panel Study of Income Dynamics to estimate a shock process for hourly wages that is both nonlinear, not AR(1), and not normal, with shocks drawn from nonparametric probability distributions. We assume that wage shocks are represented by permanent, persistent, and transitory components and show that both nonnormality and nonlinearity are present in all three components. In particular, wage shocks exhibit nonlinear persistence, i.e., unusual shocks wipe out good or bad productivity histories. We also show that the persistence of shocks increases over the life cycle. Finally, we found that the variance components of estimated wage shocks are pretty different for nonlinear and linear processes. In both processes, roughly half of the total variance comes from the persistent component. The contribution of the permanent part is larger for the nonlinear process (40%) and smaller for the linear one(14%). As a result, the two processes imply very different sources of risk.

We then build a standard life-cycle model with idiosyncratic income risk, incomplete markets, and endogenous labor supply decisions and simulate the model economy with standard and new wage dynamics to study the insurance value of taxes and transfers. First, we replace progressive taxes with proportional income taxes. The proportional tax rate generates the same level of government tax revenue as in the benchmark economy. Second, we eliminate the means-tested transfers and lower income taxes accordingly. Finally, we move to a proportional tax system and eliminate the means-tested transfers. When wage dynamics are nonlinear and nonnormal, social insurance (progressive taxes and means-tested transfers) is less valuable for poorer and richer households. Standard wage shocks with an AR(1) process imply high persistence, while the nonlinear productivity process has lower persistence. As a result, poor households have a higher chance of becoming not-so-poor next period, and hence the value of the welfare state is smaller for them. Similarly, the value of progressive taxation is higher for richer households. As they have a higher chance of falling in the income distribution, they value redistribution more under a nonlinear process.

References

Aiyagari, S. Rao. 1994. "Uninsured Idiosyncratic Risk and Aggregate Saving*." *The Quarterly Journal of Economics*, 109(3): 659–684.

Altonji, Joseph G. 1986. "Intertemporal Substitution in Labor Supply: Evidence from Micro Data." *Journal of Political Economy*, 94(3): S176–S215.

Arellano, Manuel, Richard Blundell, and Stéphane Bonhomme. 2017. "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework." *Econometrica*.

Badel, Alejandro, Mark Huggett, and Wenlan Luo. 2020. "Taxing Top Earners: a Human Capital Perspective." *The Economic Journal*, 130(629): 1200–1225.

Bénabou, Roland. 2002. "Tax and Education Policy in a Heterogeneous-Agent Economy: What Levels of Redistribution Maximize Growth and Efficiency?" *Econometrica*, 70(2): 481–517.

Bewley, Truman. 1986. "Stationary monetary equilibrium with a continuum of independently fluctuating consumers." *Contributions to mathematical economics in honor of Gérard Debreu*, 79.

Biggs, A. G., and G. R. Springstead. 2008. "Alternate measures of replacement rates for social security benefits and retirement income." *Soc Secur Bull*, 68(2): 1–19.

Chetty, Raj. 2012. "Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply." *Econometrica*, 80(3): 969–1018.

Conesa, Juan Carlos, Bo Li, and Qian Li. 2021. "Universal Basic Income and Progressive Consumption Taxes." Stony Brook University, Department of Economics Department of Economics Working Papers.

Daruich, Diego, and Raquel Fernández. 2020. "Universal Basic Income: A Dynamic Assessment." National Bureau of Economic Research Working Paper 27351.

De Nardi, Mariacristina, Giulio Fella, and Gonzalo Paz-Pardo. 2020. "Wage Risk and Government and Spousal Insurance." National Bureau of Economic Research Working Paper 28294.

Domeij, David, and Martin Flodén. 2006. "The labor-supply elasticity and borrowing constraints: Why estimates are biased." *Review of Economic Dynamics*, 9(2): 242–262.

Golosov, Mikhail, Maxim Troshkin, and Aleh Tsyvinski. 2016. "Redistribution and Social Insurance." *American Economic Review*, 106(2): 359–386.

Guner, Nezih, Christopher Rauh, and Gustavo Ventura. 2020. "Means-Tested Transfers in US: Facts and Parametric Estimates." mimeo.

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.

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.

Guner, Nezih, Remzi Kaygusuz, and Gustavo Ventura. 2021. "Rethinking the Welfare State." Centre for Economic Policy Research Working Paper.

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.

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni L Violante. 2014. "Optimal Tax Progressivity: An Analytical Framework." National Bureau of Economic Research.

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

Huggett, Mark. 1996. "Wealth distribution in life-cycle economies." *Journal of Monetary Economics*, 38(3): 469–494.

Imai, Susumu, and Michael P. Keane. 2004. "Intertemporal Labor Supply and Human Capital Accumulation*." *International Economic Review*, 45(2): 601–641.

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

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.

Keane, Michael, and Richard Rogerson. 2015. "Reconciling Micro and Macro Labor Supply Elasticities: A Structural Perspective." *Annual Review of Economics*, 7(1): 89–117.

Kindermann, Fabian, and Dirk Krueger. forthcoming. "High Marginal Tax Rates on the Top 1%? Lessons from a Life Cycle Model with Idiosyncratic Income Risk." *American Economic Journal: Macroeconomics*.

Luduvice, André Victor Doherty. 2021. "The Macroeconomic Effects of Universal Basic Income Programs." Federal Reserve Bank of Cleveland Working Papers.

MaCurdy, Thomas E. 1981. "An Empirical Model of Labor Supply in a Life-Cycle Setting." *Journal of Political Economy*, 89(6): 1059–1085.

A Data Selection and Variable Construction

The dataset contains all available PSID data, which starts from 1968 and covers until 1997 annually and then biennially until the year 2015. PSID consists of two main parts, namely Survey of Economic Opportunity (SEO) sample and the Survey Research Center (SRC) sample. SEO focuses on the poor population. We focus on the SRT part of the PSID which is designed to represent US population. We use the following list of variables: age, sex, marital status, race, education, employment status, labor hours, labor income. There are family based data files for each year and an individual based data file which contains some variables for all individuals who are part of the PSID from its beginning. We use the individual based dataset if some variables are missing in the family based dataset.

A.1 Variable Construction

Employment Status Between 1968-1975, there are 6 categories for employment status; Employed, Unemployed, Retired, Housewife, Student, Other. After 1975, two more categories are added, namely, "Only temporarily laid off" which was codified as Employed before and "Permanently disabled" which consisted in the same category with Retired. We reduced them to 6 categories above to make them comparable across all the data. Wife's employment status is available from 1994 and it has 8 categories.

Race It is reduced to three categories; white, black and other. If different races are mentioned in the questionnaire in a given year, we assigned him/her to other category. If the respondents switch between races in different years, we assign them into other category.

Education We create three broad categories of education levels and reduced education variables to these three categories for all years. The categories are

- **No High School:** 0-11th grades (includes those are with no schooling or illiterate people);

- **High School:** 12th grades or high school graduates;
- **College:** College dropouts, college graduates, or advanced/professional degree graduates.

For some years, there are two educational variables exist one is measured in years and the other is measured in the last completed education level. We used both variables and took the maximum education level as the true one. Also if there are different education levels are mentioned in different years we took the maximum education level.

Earnings Until 1994 there is only one variable for labor income. From 1995, the labor income excludes the farm and business income and they are reported separated. We consider half of the farm income and all of the business income as labor income

Wages We create hourly wage variable by dividing the total labor income by the total labor hours. Minimum wages for each state are taken from the website of Department of Labor (https://www.dol.gov/whd/state/stateMinWageHis.htm). If there is no minimum wage law in the state, we set it to zero. If there are multiple minimum wage rates are applies for a given state and a given year, we took the minimum one. Minimum wages are used for selection of the data.

Wife variables We set all wife variables as missing if there is no wife since in some observations mistakenly wife variables are not missing although there is no wife.

Price Index (CPI) Taken from Bureau of Labor Statistics (https://data.bls.gov/cgi-bin/surveymost?bls). We used to construct real variables.

A.2 Selection

Our data selection criteria consist of following steps. Before our selections, there are 155,964 year/indivial observations. We keep only male household heads hence 36,751 observations belongs to either female head or a non-head member of family are droppped.

We droped if education or race information is missing. In total, respecively 5712 and 119 obrservations are dropped because of missing information. We focues on working age, between 25 and 65. 24,039 observations dropped for this reason.

We kept individuals who worked between 520 and 5200 hours during the calender year, this resulted 6,726 of them are dropped. Also we drop, if the nominal wage is less than half the corresponding minimum wage for that year. (1702 dropped).

We drop observations with large jumps in the yearly income. To identify them, we generate the first difference for logs earnings and the interaction between first difference and lagged first difference. The interaction term would be negative and large in absolute value for big jumps which is more likely to because of measurement error. We drop observations in the year of jump, and we define "large jump" having interaction term in the first 0.25 percentile in a given year.

B Estimation of Linear Process

We calculate covariances of wage residuals by pooling all waves of the PSID. The covariances between age t and $t+s$ is given by,

$$\widehat{cov}(\omega_t, \omega_{t+s}) = \frac{1}{N_{(t,t+s)} - 1} \sum_{i=1}^{N_{(t,t+s)}} \omega_{it}\omega_{it+s}, \tag{B.1}$$

where $N_{(t,t+s)}$ is the number of observation pairs available in the data for age t and age $t+s$.

The covariance term implied by the linear model is given by,

$$cov(\omega_t, \omega_{t+s}) = \rho^s \frac{1-\rho^{2(t+1)}}{1-\rho^2}\sigma_\nu + \sigma_\theta + 1(s=0)\sigma_\epsilon, \tag{B.2}$$

where $1(s = 0)$ is an indicator function for $s = 0$.

We define vectors moments, $\widehat{M}$ and M for data and the linear model and minimize the distance between them, weighting by the squareroot of number of observations used to estimate the data moments.

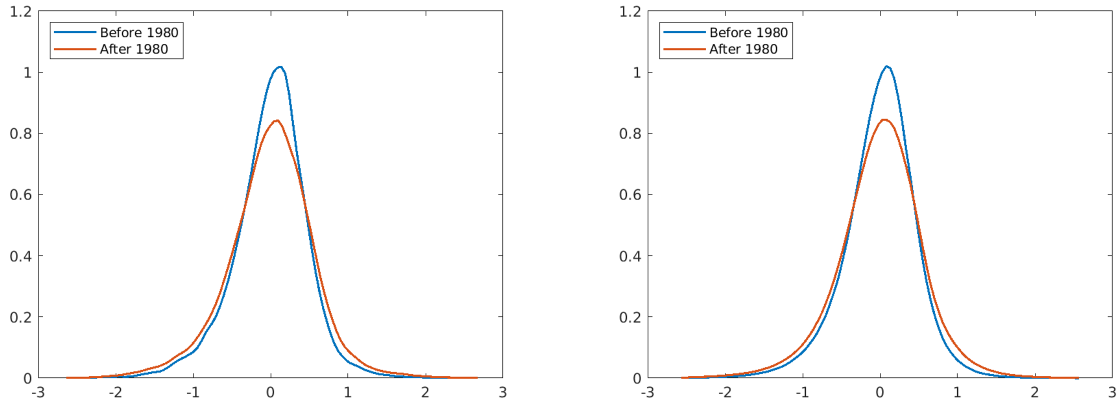
$$\min_{\sigma_\theta,\sigma_\nu,\sigma_\epsilon} (\widehat{M} - M)A(\widehat{M} - M)', \tag{B.3}$$

where the diagonal elements of A is set to square root of number obrservations, $N_{(t,t+s)}$, to corresponding moment condition.

C Comparison of before 1980 and after 1980 estimations

Wage Residuals First we compare the distributions and dynamics of wage residuals before estimation a nonlinear process. Figure 8a compares the marginal distribution of wage residuals. The distribution for after 1980 has more dispersion and seems to have fatter tails, i.e. higher kurtosis. Figure 8b shows that this feature is captured by estimated nonlinear process.

The first panel of Table 7 is for first three central moments, i.e. variance, skewness and kurtosis. We can see the the variance is higher after 1980 for all ages, which is consistent with increasing inequality. If we look at the skewness, there is slightly less negative skewness after 1980 over all, however this is driven by younger ages between 25 -34. For older ages the there is more negative skewness. The change in kurtosis is also heterogeneous across age groups. While it decreases for ages 25-34, it increases substantially for ages 35-44 and stay at similar level for older ages 45-64. We will get the standard errors of estimates in order to test whether these differences are statistically significant.



(a) Wage Residuals in the Data

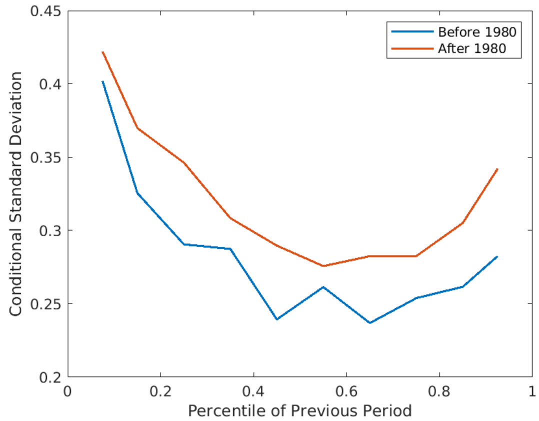
(b) Estimated Nonlinear Process

Figure 8: Marginal distribution of data and nonlinear process

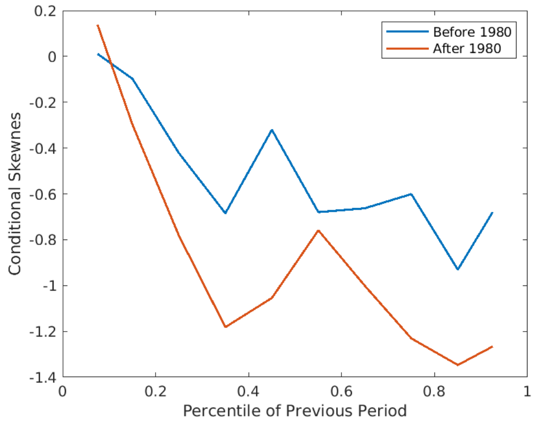
The third panel of Table 7 gives nonlinearity measures for the wage residuals in the data. We estimate the quantile function of wage residuals conditional on the its lag value with quantile regressions. Then we can calculate the measure given in the equation **??** for both lower and upper tail. For lower tail although there is not much difference overall, it goes down for ages 25-35 and up for ages 45-64. For upper tail nonlinearity goes up for ages 25-34 and 45-64, it goes down slightly for ages 35-44.

We can check the moments in the Table 7 conditional on the percentile of lag distribution. Figure 9 illustrates the conditional versions of the moments in the first panel of the Table 7. We can see that the qualitative relationship did not change and only the general levels were shifted.

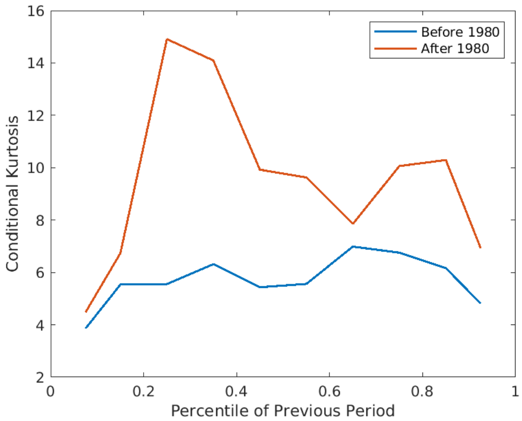
Lastly we check the nonlinear persistence of the estimated persistent part for before and after 1980. Figure 10 plots the nonlinear persistence of estimated persistent part. Overall persistence did not change much. When we investigate nonlinearity further for age groups, nonlinearity measure is reported in the last panel of Table 8. We observe that nonlinearity increased for ages 25-34 at upper tail and decreased for ages 35-44 at lower tail.



(a) Conditional Standard Deviation

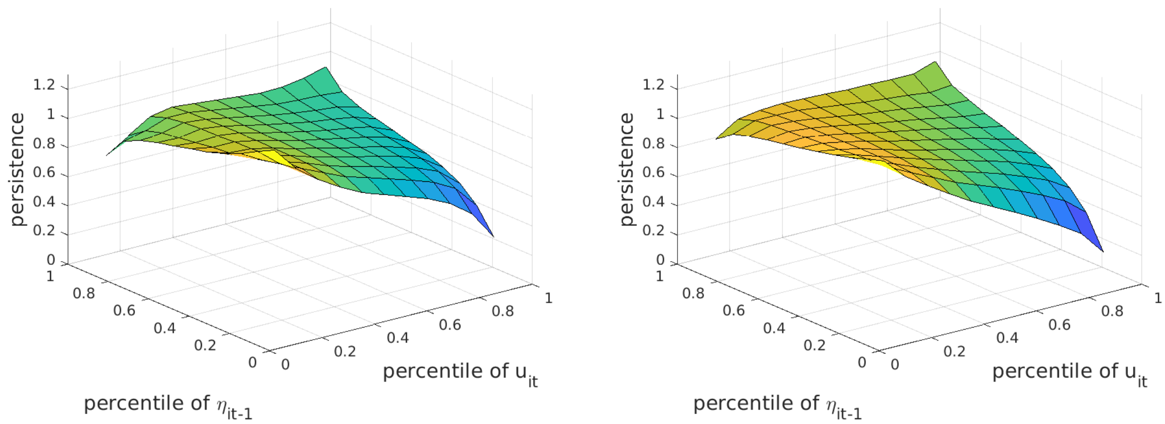


(b) Conditional Skewness



(c) Conditional Kurtosis

Figure 9: Marginal distribution of data and nonlinear process



(a) Before 1980

(b) After 1980

Figure 10: Nonlinear Persistence on Persistent Component

Table 7: Moments of Marginal Distribution of Wage residuals and nonlinearity measure

Age Group	All Ages		25 - 34		35 - 44		45 - 64	
Time Period	Before 1980	After 1980	Before 1980	After 1980	Before 1980	After 1980	Before 1980	After 1980
	Moments							
Variance	0.2248	0.3190	0.1967	0.2427	0.2211	0.2987	0.2490	0.3701
Skewness	-0.5225	-0.4343	-0.8115	-0.5450	-0.4736	-0.7202	-0.3571	-0.4708
Kurtosis	5.8322	7.2729	7.3237	5.3498	5.1623	10.5307	4.4511	5.8084
	Nonlinearity Measure							
Lower Tail	0.6750	0.7049	0.8035	0.6885	0.7195	0.6836	0.6040	0.6672
Upper Tail	0.4664	0.4041	0.5935	0.6911	0.4067	0.3722	0.2415	0.3366
Average	0.5707	0.5545	0.6985	0.6898	0.5631	0.5279	0.4227	0.5019

Age Group	All Ages		25 - 34		35 - 44		45 - 64	
Time Period	Before 1980	After 1980	Before 1980	After 1980	Before 1980	After 1980	Before 1980	After 1980
	Moments							
Variance	0.1295	0.1573	0.1061	0.1238	0.1162	0.1462	0.1548	0.1947
Skewness	-1.0751	-1.2495	-1.4185	-1.5199	-1.5193	-1.3253	-0.8216	-1.0379
Kurtosis	16.4919	13.9082	17.2890	22.7561	32.3547	14.1497	8.1824	9.9800
	Nonlinearity Measure							
Lower Tail	1.0089	0.9783	1.0682	1.0939	1.0879	1.0607	0.8236	0.8878
Upper Tail	0.1513	0.2305	0.4131	0.6004	0.3470	0.0074	0.0047	0.0651
Average	0.5801	0.6044	0.7407	0.8472	0.7174	0.5340	0.4141	0.4764

Table 8: Moments of marginal distribution of estimated persistent part and nonlinearity measures

D Discretization Nonlinear Process

- We first explain the details as if we have only annual data and we estimate annual process.
- After we reshape the data, we a panel of wage residuals and ages.

$$W = [\omega_{it}]_{i=1,\ldots,N;t=1,\ldots,T} \quad Age = [age_{it}]_{i=1,\ldots,N;t=1,\ldots,T}, \tag{D.1}$$

  where W and Age are $N \times T$ matrices, i is individual index and t is the period of the data.

- Our productivity process has three components:

$$\omega_{it} = \theta_i + \eta_{it} + \epsilon_{it}. \tag{D.2}$$

- Quantile functions are assumed to be polynomials,

$$Q^{\eta}_{it}(u|\eta_{it-1}, age_{it}) = \sum_{k=0}^{K_Q} a^Q_k(u)\varphi_k(\eta_{it-1}, age_{it}). \tag{D.3}$$

  where t is the year of the data and φ_k is a polynomial at degree k[14].

- Similarly for other quantile functions,

$$Q^{\eta_{i0}}(u|age_{i0}) = \sum_{k=0}^{K_\eta} a^{\eta_{i0}}_k(u)\varphi_k(age_{i0}), \quad Q^{\epsilon}(u|age_{it}) = \sum_{k=0}^{K_\epsilon} a^{\epsilon}_k(u)\varphi_k(age_{it}), \quad Q^{\theta_i}(u) = a^{\theta}(u). \tag{D.4}$$

- *E step:*
  - Given the parameter values, i.e. $a_k(u)$'s for the quantiles u in our quantile grid, we can calculate densities of η_{it} and θ_{it} conditional on data and use that to draw a history of η_{it} and θ_{it}.
  - So we have the following matrices by size $N \times T$,

$$\Theta = [\theta_i]_{i=1,\ldots,N} \quad Eta = [\eta_{it}]_{i=1,\ldots,N;t=1,\ldots,T}, \tag{D.5}$$

[14] In practice we use products of probabilist's Hermite polynomials.

- Given those the ϵ_{it}'s are given by the residuals,

$$Eps = [\epsilon_{it}]_{i=1,\ldots,N;t=1,\ldots,T} = W - \Theta - Eta. \tag{D.6}$$

- *M step:*
  - Given the histories of θ_i,η_{it} and ϵ_{it} we can update the parameters of quantile functions with quantile regressions.
  - For example, we regress η_{i0} on the polynomial of age_{i0} for the quantiles in the quantile grid. Notice that age_{i0} can be any age, it is just the age of the individuals when they appear in the data first period.
- The EM algorithm gives the final estimates of parameters.
- We simulate histories of θ_i,η_{it} and ϵ_{it} like in the M step using the final estimation. We use them for create discretized process for the model.

Discretization:

- For discretization of θ_i and ϵ_{it} we need a grid and a probability distribution for each age because they do not have Markovian dependence. However for η_{it} we need a transition matrix for each age too because it is a Markovian process.
- For the grids, we take a set of bins where each bins contains 10% of the draws except it is finer for the tails. In particular first bin has lowest 1%, 2nd to 5th bin they have 2% and the middle 8 bins contain 10% of the draws.
- Let N be the number of grids and $\{z_1, \ldots, z_N\}$ be the grid, then we set each z_n to the median value of its bin and the probability mass for that would be equal to the share of the draws the bin contains.
- We do this for each age by selecting the draws for each age using Age matrix.
- For the Markovian process, we construct the transition matrix for each age by counting transitions across bins between periods.

- De Nardi et al. showed this discretization does a good job for nonlinear process.