NBER WORKING PAPER SERIES

CHANGING INCOME RISK ACROSS THE US SKILL DISTRIBUTION: EVIDENCE FROM A GENERALIZED KALMAN FILTER

J. Carter Braxton
Kyle F. Herkenhoff
Jonathan L. Rothbaum
Lawrence Schmidt

Working Paper 29567
http://www.nber.org/papers/w29567

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2021, Revised December 2023

We thank Anmol Bhandari, Stephane Bonhomme, Mariacristina De Nardi, Giulio Fella, Fatih Guvenen, Jonathan Heathcote, Loukas Karabarbounis, Jeremy Lise, Hannes Malmberg, Elena Manresa, Ellen McGrattan, Robert Moffitt, Jo Mullins, Michael Nattinger, Gonzalo Paz Pardo, Fabrizio Perri, Sergio Salgado, Gianluca Violante and numerous seminar participants for helpful comments. Any opinions and conclusions expressed herein are those of the author(s) and do not necessarily represent the views of the Federal Reserve System or the U.S. Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product (Data Management System (DMS) number: P-7503840, Disclosure Review Board (DRB) approval number: CBDRB-FY23-SEHSD003-071 and CBDRB-FY24-SEHSD003-009.) The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Changing Income Risk across the US Skill Distribution: Evidence from a Generalized Kalman Filter
J. Carter Braxton, Kyle F. Herkenhoff, Jonathan L. Rothbaum, and Lawrence Schmidt
NBER Working Paper No. 29567
December 2021, Revised December 2023
JEL No. E24,J3,J6

ABSTRACT

For whom has earnings risk changed, and why? We answer these questions by combining the Kalman filter and EM-algorithm to estimate persistent and temporary earnings for every individual at every point in time. We apply our method to administrative earnings linked with survey data. We show that since the 1980s, persistent earnings risk rose by 20% for both employed and unemployed workers and the scarring effects of unemployment doubled. At the same time, temporary earnings risk declined. Using education and occupation codes, we show that rising persistent earnings risk is concentrated among high-skill workers and related to technology adoption.

J. Carter Braxton
University of Wisconsin
jcbraxton@wisc.edu

Kyle F. Herkenhoff
University of Minnesota
Department of Economics
4-101 Hanson Hall
1925 Fourth Street South
Minneapolis, MN 55455
and IZA
and also NBER
kfh@umn.edu

Jonathan L. Rothbaum
Social, Economic, and Housing Statistics
US Census Bureau
4600 Silver Hill Road
Washington, DC 20233
jonathan.l.rothbaum@census.gov

Lawrence Schmidt
Sloan School of Management
Massachusetts Institute of Technology
100 Main Street
Cambridge, MA
ldws@mit.edu

Prior work has shown that persistent earnings shocks are often not well insured (e.g., Blundell, Pistaferri, and Preston (2008)), and hence understanding *how* and *why* the dispersion in persistent and temporary earnings risk has evolved over time is critical for individual welfare and policy design. We make progress on this agenda by developing a filtering method that estimates parameters of an income process and recovers persistent and temporary earnings for every individual at every point in time. We apply our method to earnings records from the Social Security Administration (SSA) linked to the Current Population Survey's Annual Social and Economic Supplement (CPS). We find that the variance of shocks to persistent earnings, i.e., persistent earnings risk, among employed workers rose by over 20% since the 1980s, while temporary earnings risk fell by a similar magnitude. Over the same time period, the average decline in persistent earnings per year of unemployment doubled, and combined earnings risk – which incorporates layoff risk and earnings risk among the unemployed – also increased by nearly 20%. The tractability of our approach allows us to estimate the evolution of persistent and temporary risk among finely partitioned subsets of our data and explore the *why* portion of our research question. Using the demographic and occupation information from our linked SSA-CPS sample, we find that the increase in persistent earnings risk is concentrated among high-skill workers and related to technology adoption. Finally, we find that the observed increase in persistent earnings risk since the 1980s generates sizable welfare losses.

This paper makes three contributions. First, we show how the Kalman filter and an Expectation Maximization (EM) algorithm can be used to estimate the parameters of a flexible but easily interpretable model of income dynamics and recover estimates of earnings shocks for each individual in every period. As is consistent with much of the income process literature, we write down a low-dimensional representation of individual earnings as the sum of latent persistent and temporary components.[1] Using the EM algorithm, we derive updating equations for the parameters of the income process, which resemble generalized least squares regressions. To provide intuition, consider a simple example in which z_{it} is persistent earnings and the parameter F governs the degree of persistence (i.e., $z_{it} = Fz_{it-1} + \epsilon_{it}$ where ϵ_{it} is mean-zero, white noise). The Kalman filter yields estimates of the persistent state $\hat{z}_{it}$ and the EM updating equation for F is simply the slope coefficient in a regression of $\hat{z}_{it}$ on $\hat{z}_{it-1}$. These closed form updating equations allow the model to easily handle income process parameters that depend linearly on a potentially large number of observables (e.g., age, employment status, education, and occupation are easy to incorporate since they are simple interaction terms

[1] While we work with the canonical example in which persistent earnings follow a persistent AR(1) process throughout, our approach can naturally extend to incorporate additional linear dynamics—including individual fixed effects in earnings levels and growth rates, as well as moving average components—while remaining tractable.

in the updating regressions). The output of our combined Kalman filter-EM procedure yields (quasi-)maximum likelihood estimates of the parameters governing income risk and a panel of persistent and temporary earnings realizations. We additionally provide a detailed econometric "cookbook" to make these methods accessible to those who simply want to apply and modify our codes.

A benefit of our Kalman filter and EM algorithm approach is that it provides a natural way of incorporating observations with very low or zero earnings.[2] Motivated by economic theories of human capital depreciation during unemployment (e.g., Ljungqvist and Sargent (1998), among others), we posit a law of motion for persistent earnings when individuals have zero earnings (with a slight abuse of convention, we use 'zero earnings' and 'unemployment' interchangeably). During unemployment spells, individuals receive shocks to persistent earnings; these shocks have a different mean and variance than those received during periods of employment. We provide an explicit microfoundation for such a process in a search model with risky human capital, learning by doing, and skill depreciation in unemployment. Despite individuals' lack of earnings information during unemployment spells, the law of motion for persistent earnings is identified via earnings upon re-entry to work.

We estimate our filter on a linked sample of SSA-CPS earnings records from 1981 to 2019.[3] We begin by pooling the data and estimating a stationary income process. The estimated parameters reveal that the unemployed (i.e., those with very low or zero earnings) face substantial earnings risk.[4] We estimate that the variance of shocks to persistent earnings of the unemployed is nearly four-times that of the employed, and we find that the unemployed face persistent earnings losses of nearly 19% per year of unemployment (compared with a 0.05% gain for the employed).

Our second contribution is to examine how earnings risk has varied over time. In addition to employment status, we extend our filter to allow for age- and time-dependent variances of persistent and temporary earnings shocks as well as for age- and time-dependent means of persistent earnings shocks. We document an upward trend in persistent earnings risk since the 1980s. Among the employed, the variance of persistent earnings shocks rose by over 20%.

[2] Additionally, our approach is robust to non-Gaussian innovations, as we discuss further in Appendix A.4. Through a series of Monte Carlo exercises in Appendix B.6, we show that our estimates of individual-level persistent earnings exhibit bias of less than 0.1% when we apply our algorithms on simulated data in which innovations are highly non-normal and drawn from Guvenen et al. (2021)'s mixture distribution.

[3] In Appendix C.2, we establish representativeness of our link SSA-CPS sample by showing that trends in median earnings and earnings inequality closely align with Guvenen et al. (2022), who utilize the universe of SSA records, across time and demographic groups.

[4] In our baseline estimation, we define an individual to be unemployed if their annual earnings are below the average Social Security Administration (SSA) cutoff for receiving a full-year of credits toward SSA retirement benefits, which is $3,350 (in 2005 PCE dollars) between 1978 and 2019.

Conversely, there has been an offsetting downward trend in temporary earnings risk over the same time period. Through the combination of these two factors, overall earnings risk among the employed has a downward trend, indicating that examining only overall earnings risk can mask heterogeneous trends in the underlying temporary and persistent components.

Leveraging one of the strengths of our approach, we also study the changing dynamics of persistent earnings levels and risk during periods of non-employment. We find that persistent earnings losses have accelerated for unemployed individuals. Entering full-year unemployment results in a -10% decline in persistent earnings in 1985, but by 2015, this rate of loss more than doubles, reaching -22% per annum. On top of this more rapid negative drift in persistent earnings, the *variance* of persistent earnings shocks during unemployment has also nearly doubled between 1985 and 2015. Using our estimates of persistent earnings risk for employed and unemployed workers, we create a *combined* measure of persistent income risk that explicitly takes into account the risk workers face from entering into unemployment and how the likelihood of entering into unemployment has changed over time. We find that since the mid-1980s, combined persistent income risk has increased by nearly 20%.

A central component of the measure of combined persistent risk is the likelihood of entering into unemployment. Consistent with Fujita (2018), we find that the likelihood of entering into unemployment has been declining in the US since the mid-1980s. This decline in the likelihood of entering unemployment has been a mitigating factor in the rise of persistent income risk. In a counterfactual exercise, we show that if the rate of entry into unemployment had remained at its 1985 value, combined persistent income risk would have increased by nearly 30% (instead of 20%).

Our third contribution is to examine why persistent earnings risk rose. By linking our administrative earnings database to survey responses in the CPS, we test a number of explanations of rising persistent earnings risk. Using the occupation and geographic information from the CPS we do not find any empirical link between the trends in risk we document and empirical proxies for exposure to several other secular changes occurring during our period: declining manufacturing employment and union coverage at the state level, as well as declining wages and employment in routine occupations.

Instead, we find that rising persistent earnings risk is concentrated among high-skill workers and empirically linked with adoption of skill-biased technologies. We begin by documenting rising persistent earnings risk based on three definitions of "high-skill" work. First, we estimate changes in persistent income risk by education level. We find that the largest increase in persistent income risk has occurred among individuals with a college degree or more than a college degree. Second, we examine how persistent income risk has evolved over time by

ONET job zones, which measure the degree of preparation required to enter into an occupation. We show that the increase in persistent income risk was driven by the highest ONET job zone, which requires "extensive skill, knowledge, and experience." Third, we measure changes in persistent earnings risk over time within detailed occupation codes (e.g., Autor and Dorn (2013)). We show that occupations with a greater degree of non-routine cognitive task content, as measured in Acemoglu and Autor (2011), have seen a larger increase in persistent earnings risk.[5]

One potential mechanism for why high skill workers are facing greater persistent risk is that they face greater exposure to the introduction and diffusion of new, skill-biased technologies (e.g., Krueger (1993) and Deming and Noray (2020)). While existing work (e.g., Krusell et al. (2000)) suggests that these new technologies are complementary to skilled labor, they also create a risk of skill displacement. For instance, workers may be unable to easily acquire the skills required to adapt to the new technology, or they may find that previously valuable/scarce expertise is no longer required to produce with the new technology. Hence, new vintages of technology can create winners and losers, where the biggest losses and gains occur among skilled workers.[6] Since it is costly to acquire new skills and changes in the technological frontier are permanent, such a phenomenon naturally generates substantial and persistent variation in earnings across workers.

We provide evidence consistent with this technology-adoption mechanism by linking Burning Glass vacancy data to our SSA-CPS data. The Burning Glass data allows us to measure the extent to which occupations introduced intensive computer and software use in the workplace as a proxy for the introduction of new skill-biased technologies.[7] We find that occupations with high computer use in the 2010s on average saw considerably larger increases in persistent earnings risk, larger acceleration of earnings losses during unemployment, and larger increases in earnings variability among the unemployed.

Lastly, we integrate our income process into a Bewley-Huggett-Aiyagari model, and we compute the welfare implications of rising persistent earnings risk. One benefit of our filtering approach is that the statistical assumptions required for our estimator to maximize the conditional likelihood of earnings (given the observed conditioning variables such as employment

[5]We also show that we obtain similar results using other proxies for the degree to which an occupation is high skill (e.g., average years of completed education and average earnings).

[6]Quantitative papers with this mechanism include Chari and Hopenhayn (1991), and Violante (2002). For direct empirical evidence and related theory for how high skilled workers see larger increases in risk following innovations, see Braxton and Taska (2023), Kogan et al. (2020), and Kogan et al. (2023). See also Goldin and Katz (2010), Akerman et al. (2015), Atack et al. (2019), and Feigenbaum and Gross (2020).

[7]Burning Glass collects detailed information on the skills listed in vacancies. We follow recent work by Hershbein and Kahn (2018) and Atalay et al. (2020), which argues that the skill requirements in vacancy postings are informative on the technology of the firm posting the vacancy.

status) nest a large variety of job loss processes. We find empirically that prior earnings are negatively related to the probability of being unemployed, a feature which we show is straightforward to incorporate into quantitative models. In our quantitative model, rising persistent earnings risk reduces welfare by 5% of lifetime consumption, whereas the decline in temporary earnings risk only has a minimal impact on welfare.

Related literature. This paper contributes to the large literature on income process estimation (e.g., Meghir and Pistaferri (2004), Storesletten et al. (2004), Blundell et al. (2008), Altonji et al. (2013), Guvenen et al. (2014), Chatterjee et al. (2021), Guvenen et al. (2021) among others[8]), and on the evolution of income risk (e.g., Gottschalk and Moffitt (1994), Sabelhaus and Song (2010), Bloom et al. (2017), Ziliak et al. (2020), Moffitt and Zhang (2018), Moffitt (2020)). Influential work by Bloom et al. (2017) showed that dispersion in earnings growth rates at 1-year and 5-year horizons has been flat or declining in the US since the 1980s. Moffitt (2020) summarizes recent work on the topic.

Relative to this literature, we make three contributions. First, we extend the canonical persistent-temporary income process to include spells of unemployment and shocks to persistent earnings during spells of unemployment. Our model of job loss parsimoniously generates non-Gaussian income risk (e.g. Geweke and Keane (2000), Altonji et al. (2013), Arellano and Bonhomme (2016), Arellano et al. (2017), De Nardi et al. (2020), Guvenen et al. (2021) and Halvorsen et al. (2023)) and yields a novel measure of income risk, which we refer to as *combined persistent income risk*, which accounts for changes in risk within an employment spell as well as changes in the likelihood of transitioning across employment states. Second, we show that despite flat or declining dispersion in earnings growth rates at short and long run horizons, persistent earnings risk has risen while temporary earnings risk has fallen. Third, unlike much of the literature that employs generalized method of moments estimators, our Kalman filter-EM algorithm allows us to examine heterogeneity in the evolution of earnings risk over time and shows how persistent earnings risk has evolved over time by age, education level, coarse occupation groups, and detailed occupations, etc. We find that persistent earnings risk rose in "high-skill" jobs most exposed to technology adoption.

Methodologically, we apply the EM approach to maximizing the likelihood of linear state space models (see, e.g., Shumway and Stoffer, 1982) to a panel setting, then extend the model to allow for mean and variance parameters of the state space system to be affine and exponentially-affine in a set of observed variables, respectively. While the first extension for conditional means is straightforward, to our knowledge, demonstrating the tractability the EM approach given our assumption on conditional variances appears to be new to the literature. While several papers

[8] See Meghir and Pistaferri (2011) and Guvenen et al. (2021) for detailed discussions of the literature.

apply the Kalman filter to income data, including Nakata and Tonetti (2015), Borella et al. (2019) and Chatterjee et al. (2021), our paper makes several contributions: (1) we allow for unemployment and thus incorporate tail risk into earnings risk measures, (2) we validate our measures of persistent and temporary risk using job transitions and other survey metrics, (3) our EM estimation algorithm allows us to tractably incorporate rich heterogeneity by conditioning on a potentially large number of observables,[9] and finally, (4) we use our estimates to understand the drivers of persistent and temporary earnings risk in both the cross-section and over time.

Our paper contributes to two additional literatures. Influential work by Kopczuk et al. (2010) and DeBacker et al. (2013) argues that the rise in income inequality is due to a rise in persistent income inequality. Our work complements these findings by highlighting an increase in persistent income risk as a reason for an increase in persistent income inequality, as opposed to increased initial dispersion in persistent earnings or an increase in persistence of the income process. Finally, our paper contributes to the large literature about how technological change has impacted the labor market (e.g., Autor et al. (1998), Krusell et al. (2000), Autor et al. (2003), Autor and Dorn (2013), Hershbein and Kahn (2018), Atalay et al. (2018), Deming and Noray (2020), Kogan et al. (2023), Heathcote et al. (2020), and Braxton and Taska (2023)). Our contribution is to demonstrate that exposure to technological change is linked with increases in persistent earnings risk.

1 Empirical framework

In this section, we describe our econometric framework for modeling income. We start in Section 1.1 by describing a simple income process that extends the canonical persistent-temporary structure to include spells of unemployment. In Section 1.2, we show how the parameters of our income process are identified. We then discuss in Section 1.3 how we use the Kalman filter along with an EM algorithm to estimate the parameters of our income process and obtain estimates of persistent and temporary earnings for every individual in every period. In Section 1.4, we discuss how to use our method to estimate much richer models of income dynamics by allowing the mean and variance parameters of the low dimensional model from Section 1.1 – including those governing income dynamics while unemployed – to be functions of other observables (e.g., time, education, detailed occupation code, etc.). Finally, we conclude with a

[9] There are several other approaches to modeling rich income heterogeneity including parametric and non-parametric approaches, e.g. Browning et al. (2010) and Jensen and Shore (2011), respectively. While our approach is more restrictive in terms of restrictions on unobservables relative to these articles, our approach is significantly more tractable and allows for flexibly studying the role played by various observable variables in generating heterogeneity in income dynamics.

step-by-step guide on how to apply our estimation routine in Section 1.5.

1.1 An income process with unemployment risk

We begin with a panel dataset of income, $Y_{i,t}$, where $i \in \{1, \ldots, N\}$ indexes individuals and $t \in \{1, \ldots, T\}$ indexes years.[10] We are interested in understanding the evolution of earnings, net of predictable lifecycle components. Let $\widehat{y}_{i,t}$ characterize how observable lifecycle components influence log earnings. Then define residual log earnings, denoted $y_{i,t}$, as $y_{i,t} = \log(Y_{i,t}) - \widehat{y}_{i,t}$. In the remainder of the paper, we focus on the factors that influence changes in residual log earnings $y_{i,t}$, henceforth *income*.

The income process we define below depends upon whether an individual is employed in a given year. Let $l_{i,t} = [l_{E,i,t} \ l_{U,i,t}]'$ be a vector that identifies an individual's labor market status. Element $l_{E,i,t}$ is an indicator variable that equals one when individual i is employed in year t and zero otherwise. Likewise, $l_{U,i,t}$ equals one when individual i is unemployed, and zero otherwise. We define an individual to be employed when they have labor income above a minimum earnings criterion $\bar{y}$ (i.e., $Y_{i,t} > \bar{y}$) and unemployed otherwise. In our econometric derivations, we assume $\bar{y} = 0$ to avoid any issues with censoring or misclassification of employment status. However, in practice, we assume small positive values of $\bar{y} > 0$ to capture our economic intuition that extremely low values of earnings are associated with non-employment.[11] We first discuss the law of motion for earnings conditional on $l_{i,t}$, then will discuss the data generating process for $l_{i,t}$ below.

For employed individuals, we model the process for income $y_{i,t}$ as the sum of persistent and temporary earnings. There is a missing data problem in the sense that the persistent and temporary components of income $y_{i,t}$ are not separately observable. Let $z_{i,t}$ denote the unobserved persistent component of income (whose evolution we discuss next), and let $\omega_{i,t}$ denote the temporary shock. When an individual is unemployed, their observed income ($y_{i,t}$) is not observable

[10] For ease of notation, we assume here that the panel is balanced, but the extension to an unbalanced panel setting is immediate.

[11] For the small positive values of $\bar{y}$ considered in this paper, we use Monte Carlos to establish that employment misclassifications are extremely rare and have a minuscule impact upon our estimates of the shocks to temporary and persistent earnings (see Appendix B.7). An alternative approach is to have entry into unemployment governed via a Tobit style selection mechanism. Conceptually, incorporating a Tobit structure is easy but it makes implementation considerably more complicated and slower as the filtering problem becomes non-linear requiring the use of a particle filter. We view the simulations discussed above as suggesting that the results are unlikely to change with this extra machinery.

to the econometrician. We assume that, conditional on $l_{i,t}$, observed earnings satisfy,

$$\begin{aligned} y_{i,t} &= \begin{cases} z_{i,t} + \omega_{i,t} & \text{if } l_{E,i,t} = 1 \\ \cdot & \text{if } l_{E,i,t} = 0 \end{cases} \\ \mathbb{V}(\omega_{i,t} \mid l_{i,t}) &= R(l_{i,t}), \end{aligned} \tag{1}$$

where $\omega_{i,t}$ is independent of $z_{i,t}$ conditional on $l_{i,t}$. Temporary shocks to an individual's earnings ($\omega_{i,t}$) are drawn from a normal distribution with mean zero and variance $R(l_{i,t})$, where the variance depends upon the individual's labor market status.[12]

We next discuss the law of motion for persistent earnings. We model the process for persistent earnings $z_{i,t}$ as an autoregressive process subject to innovations with means and variances that depend on $l_{i,t}$. For each individual, persistent earnings evolves according to,

$$\begin{aligned} z_{i,t+1} &= F\, z_{i,t} + B(l_{i,t+1}) + \nu_{i,t+1} \\ \mathbb{V}(\nu_{i,t+1} \mid l_{i,t+1}) &= Q(l_{i,t+1}) \\ z_{i,0} &\sim N(0, Q_0). \end{aligned} \tag{2}$$

where the parameter F denotes the persistence of $z_{i,t}$. $B(l_{i,t})$ captures the drift of an individual's persistent income in period t. Observe that the drift (mean) of the shock to persistent earnings varies by an individual's employment status $l_{i,t}$. $\nu_{i,t}$ denotes the shock to persistent income for individual i in period t, which – conditional on employment status $l_{i,t}$ – is independently drawn from all other shocks from a normal distribution with mean zero and variance $Q(l_{i,t})$. Finally, $z_{i,0}$ denotes the initial draw of persistent earnings for an individual i, which is assumed to be drawn from a normal distribution with mean zero and variance Q_0.

This income process extends the persistent-temporary structure that is common in the literature (e.g., Storesletten et al. (2004), Blundell et al. (2008)) to incorporate spells of unemployment. Although unobserved by the econometrician, persistent income continues to evolve during spells of unemployment and labor market status $l_{i,t}$ is informative about these dynamics. Since income is set to missing during unemployment, contemporaneous observations of income during unemployment contain no incremental information about $z_{i,t}$ conditional on $l_{i,t}$. However, as we discuss in more detail in Section 1.2, the mean and standard deviation of income at re-employment will inform the parameters governing the law of motion for persistent earnings $z_{i,t}$ during unemployment.

[12]In particular, individuals do not receive temporary shocks when unemployed.

Economic interpretation of income risk among the unemployed. Our income process allows individuals to receive persistent income shocks during unemployment. One natural interpretation of these persistent income shocks is human capital obsolescence. The economics are similar to Ljungqvist and Sargent (1998) and Alvarez, Borovičková, and Shimer (2016) in which human capital is subject to shocks and a downward drift while unemployed. This type of latent risk to the unemployed is real in the sense that it is a shock to perhaps the most important asset of these individuals: human capital. In Appendix A.3, we show that the income process specified in equations (1) and (2) is consistent with a labor search model with on-the-job human capital accumulation and skill depreciation during unemployment.

Timing and sequential exogeneity. While our income process is largely agnostic about the process for employment – it is simply observed and conditioned on – in order for our estimator to maximize the conditional likelihood of observed income $\{y_\tau\}_{\tau \leq T}$ given $\{l_\tau\}_{\tau \leq T}$, we impose an assumption of *sequential exogeneity*. More formally, we impose that $z_{i,t-1} \perp l_{i,t} | \{l_\tau, y_\tau\}_{\tau \leq t-1}$, which does not restrict the dependence between $l_{i,t}$ and both lagged employment status $\{l_\tau,\}_{\tau \leq t-1}$ as well as lagged income $\{y_\tau\}_{\tau \leq t-1}$.[13] As we discuss in Section 1.4, employment realizations can also be functions of lagged observables (e.g., age, tenure, etc.). In addition, our data generating process does not allow employment realizations to correlate with contemporaneous or future realizations of permanent or temporary shocks $(\nu_{i,t}, \omega_{i,t})$.

The dynamics of the system evolve as follows. The individual first draws an observation of latent initial persistent income $(z_{i,0})$ from a normal distribution with mean zero and variance Q_0. Given our sampling restrictions, all individuals start employed–i.e., $l_{E,i,1} = 1$. Moving forward, we draw independent persistent and temporary innovations ($\nu_{i,1}$ and $\omega_{i,1}$) from distributions that depend on $l_{i,1}$. We then draw $l_{i,t}$ conditional on observables (e.g., lagged income $y_{i,t-1}$), and this process repeats. This timing assumption yields *sequential exogeneity*, and it is naturally satisfied in the way we write many labor search models and Bewley models (see Appendix A.3 and Appendix D, respectively).

Job transition process. As econometricians, we observe complete histories of job transitions in our data, $\{l_{i,t}\}_{t=1}^T$. Job transitions are a central component of overall income risk, as we make clear in the combined earnings risk metrics of Section 3. A variety of job destruction and job finding processes are consistent with our econometric model (e.g., job loss and job finding rates that depend on lagged income, demographics, and rich histories of experience and job tenure), but we are agnostic about the process until Section 5.1. In Section 5.1, we posit a law of motion for employment that can be used in quantitative models, and we explore its

[13] Under this assumption, we can factor each term in the joint likelihood of $l_t, y_t | \{l_\tau, y_\tau\}_{\tau \leq t-1}$ into two pieces: one which captures the evolution of $l_t | \{l_\tau, y_\tau\}_{\tau = \leq t-1}$ and another which captures $y_t | l_t, \{l_\tau, y_\tau\}_{\tau \leq t-1}$.

properties. As one would expect, individuals with higher prior income are significantly less likely to lose their job. Lower prior income individuals are on a more slippery ladder, which they repeatedly fall off. We also show that our econometric model is consistent with labor search models in Appendix A.3. Lastly, we show in Appendix C.7 that our econometric model is tractable enough to incorporate job-to-job transitions, which are observed in our data. Our econometric model is consistent with these job-to-job transitions depending on rich histories of income, demographics, experience, and tenure.

1.2 Identification

Before discussing our method for estimating the income process outlined in equations (1) and (2), we discuss the identification of the parameters. To ease the presentation of identification, we assume that $F = 1$.[14] Let B_E (B_U) denote the drift of persistent earnings for the employed (unemployed), and Q_E (Q_U) denote the variance of persistent earnings for the employed (unemployed).[15]

We begin by discussing the identification of parameters that govern income risk for the employed. Consider the set of individuals who are employed in periods t through $t-k$. Given the income process specified in equations (1) and (2), and assuming $F = 1$, we can write the variance of income changes over these k periods ($y_t - y_{t-k}$) as,

$$var(y_{i,t} - y_{i,t-k} | l_{E,i,t} = \cdots = l_{E,i,t-k} = 1) = kQ_E + 2R. \tag{3}$$

By considering equation (3) for two different values of k, we arrive at a system of two-equations and two-unknowns which allow us to obtain Q_E and R. The intuition from this expression is that we identify the variance of temporary and persistent earnings among the employed by looking at the variance of income changes over different horizons. Finally, we can obtain an estimate of the drift in persistent earnings among the employed, by taking the mean of earnings changes over a 1-year horizon for individuals employed in both periods, i.e., $E(y_{i,t} - y_{i,t-1} | l_{E,i,t+1} = l_{E,i,t} = 1) = B_E$.

We next discuss the identification of parameters for the unemployed. Assume that an individual i is employed in period $t-1$, unemployed in period t and then employed in period $t+1$. Let $\nu^U_{i,t}$ denote the shock to persistent earnings while unemployed in period t, and let $\nu^E_{i,t+1}$ denote the shock to persistent earnings while employed in period $t+1$. The change in

[14] In Appendix A.1, we show identification when $F < 1$ and show how the persistence parameter (F) as well as variance of initial persistent earnings are identified.

[15] Formally, $B(l_{i,t}) = B_E$ if $l_{E,i,t} = 1$, $B(l_{i,t}) = B_U$ if $l_{U,i,t} = 1$, $Q(l_{i,t}) = Q_E$ if $l_{E,i,t} = 1$, $Q(l_{i,t}) = Q_U$ if $l_{U,i,t} = 1$. $R(l_{i,t}) = R$ if $l_{E,i,t} = 1$, $R(l_{i,t}) = \cdot$ if $l_{U,i,t} = 1$ (note individuals do not receive temporary shocks when unemployed).

income for individual i between periods $t-1$ and $t+1$ is then given by,

$$y_{i,t+1} - y_{i,t-1} = B_t^U + \nu_{i,t}^U + B_{t+1}^E + \nu_{i,t+1}^E + \omega_{i,t-1} + \omega_{i,t+1}. \tag{4}$$

Taking the mean and variance of equation (4), we are able to identify the parameters Q_U and B_U. In particular, letting $\Delta y_{t-1,t+1} = y_{i,t+1} - y_{i,t-1}$ denote the change in earnings around an unemployment spell in t, we have that $var(\Delta y_{t-1,t+1} | l_{E,i,t+1} = l_{E,i,t-1} = 1, l_{E,i,t} = 0) = Q_U + Q_E + 2R$. Since Q_E and R are already identified from the earlier moments for the employed, we are able to identify Q_U. Finally, we have that $E(\Delta y_{t-1,t+1} | l_{E,i,t+1} = l_{E,i,t-1} = 1, l_{E,i,t} = 0) = B_U + B_E$, and since B_E is known from the employed estimates we are able to identify B_U. Thus, despite earnings not being observed in period t, we are able to identify the shocks to persistent earnings during the unemployment spell in period t by examining earnings upon re-entry. We next discuss our method for estimating the income process specified in equations (1) and (2).

1.3 Estimation

In this section, we discuss how we estimate the income process presented in equations (1) and (2). The estimation proceeds by iterating between two steps, which we discuss in more detail below. The first step uses the Kalman filter to recover estimates of the time series of temporary and persistent earnings for each person. The second step recovers the parameters that govern the income process.[16]

1.3.1 Estimating time series of temporary and persistent income shocks.

In this section, we discuss how we use the Kalman filter to recover the time series of temporary and persistent income shocks for each individual. For now we assume that the parameters that govern the income process are known and we will discuss how these parameters are estimated in Section 1.3.2.

The first step in estimating the time series of temporary and persistent income shocks is recasting the income process as a state-space system. In practice, we make the state variable the current realization of persistent earnings as well as its lag. Let $\zeta_{i,t}$ denote an individual i's

[16]One way wonder, why not simply use the moments discussed in Section 1.2 to estimate unknown parameters of the simple model we discuss here? In addition to potential efficiency gains that come from using a full information method, the EM algorithm extends easily to the richer model we will present in Section 1.4 in which parameters of the model can depend flexibly on observable variables.

unobserved state in period t:[17]

$$\zeta_{i,t} = \begin{bmatrix} z_{i,t} \\ z_{i,t-1} \end{bmatrix},$$

where $z_{i,t}$ is persistent earnings of individual i in period t. From equation (2), the state vector $\zeta_{i,t}$ evolves according to the following law of motion – with autocorrelation matrix $\hat{F}$ and mean $\hat{B}$ – which we hereafter refer to as the *state equation*,

$$\zeta_{i,t} = \begin{bmatrix} z_{i,t} \\ z_{i,t-1} \end{bmatrix} = \underbrace{\begin{bmatrix} B(l_{i,t}) \\ 0 \end{bmatrix}}_{\hat{B}(l_{i,t})} + \underbrace{\begin{bmatrix} F & 0 \\ 1 & 0 \end{bmatrix}}_{\hat{F}} \zeta_{i,t-1} + \begin{bmatrix} \nu_{i,t} \\ 0 \end{bmatrix}. \tag{5}$$

Using the definition of the state vector $\zeta_{i,t}$ and the income process specified in equation (1), labor income evolves according to the following equation while employed, which we hereafter refer to as the *measurement equation*,

$$y_{i,t} = H(l_{i,t})\zeta_{i,t} + l_{E,i,t}\,\omega_{i,t}, \tag{6}$$

where $H(l_{i,t}) = \begin{bmatrix} l_{E,i,t} & 0 \end{bmatrix}$ governs the relationship between the state vector $(\zeta_{i,t})$ and earnings $y_{i,t}$ among individuals who are employed $(l_{E,i,t} = 1)$.[18]

Equations (5) and (6) recast our income process as a state-space system where persistent earnings (and its lag) are the unobserved state variable. As discussed in Hamilton (1994a), the Kalman filter provides a method for estimating the unobserved state variable with the minimum mean squared error. Hence, we recover the time series of persistent and temporary earnings for each individual using the Kalman filter. We next outline our Kalman filtering algorithm.

As noted above, in the Kalman filtering step of the estimation, we assume the parameters of the income process are known. Starting with estimates $\hat{\zeta}_{i,1|0}$ and $M_{i,1|0}$, which we will define below, we obtain an estimate of the time series of persistent earnings (and its lag) from the Kalman filter, which we denote $\hat{\zeta}_{i,t}$, as follows:

1. Estimate the Kalman Gain :

$$K_{i,t} = M_{i,t|t-1}H^{'}(l_{i,t})\left[H(l_{i,t})M_{i,t|t-1}H^{'}(l_{i,t}) + R(l_{i,t})\right]^{-1}.$$

[17] It will be convenient to make the state vector persistent earnings as well as its lag when we derive the expressions for updating the parameters of the income process in Section 1.3.2.

[18] When an agent is unemployed ($l_{E,i,t} = 0$), the value of the observation $y_{i,t}$ provides no additional signal about latent earnings other than what can be inferred from other observables, so the Kalman filter will not directly use $y_{i,t}$ to update its guess about $z_{i,t}$.

2. Update the state vector:

$$\hat{\zeta}_{i,t|t} = \hat{\zeta}_{i,t|t-1} + K_{i,t}\left(y_{i,t} - H(l_{i,t})\hat{\zeta}_{i,t|t-1}\right)$$
$$\hat{\zeta}_{i,t+1|t} = \hat{F}\hat{\zeta}_{i,t|t} + \hat{B}(l_{i,t}).$$

3. Update the MSE matrix:

$$\begin{aligned} M_{i,t|t} &= M_{i,t|t-1} - K_{i,t}H(l_{i,t})M_{i,t|t-1} \\ M_{i,t+1|t} &= \hat{F}M_{i,t|t}\hat{F}' + Q(l_{i,t+1})e_1^2 e_1^{2'}. \end{aligned}$$

where $e_1^2 = [1,0]'$. Repeat steps 1-3 for $t = 2,...T$, and for each individual $i \in \{1,..,N\}$. In the algorithm above the MSE matrix $M_{i,t+k|t}$ represents the uncertainty about the estimate of the state vector at time $t+k$ using information up to to time t (for $k = 0,1$). Finally, to run the Kalman filter we need an initial estimate of the mean of the state-vector ($\hat{\zeta}_{i,1|0}$) and the MSE matrix ($M_{i,1|0}$). We set $\hat{\zeta}_{i,1|0} = \hat{B}(l_{i,1})$, and we initialize the variance of the state vector as,

$$M_{i,1|0} = \begin{bmatrix} F^2 Q_0 + Q(l_1) & FQ_0 \\ FQ_0 & Q_0 \end{bmatrix}.$$

Smoothing. In order to apply the EM algorithm below – which relies on the expected full-information log likelihood – we must apply the Kalman smoother. Moreover, Hamilton (1994b) comments that when the value of the state vector is of interest in its own right, as in our application, we can improve the inference about the historical values of the state vector in the middle of the sample by using the Kalman smoother. The Kalman smoother is used after running the Kalman filter and incorporates the full time series of data to generate mean squared error minimizing estimates of the unobserved state variable. We use the Kalman smoother and denote the updated paths for the mean and variance covariance matrices for persistent earnings and its lag by $\{\{\hat{\zeta}_{i,t|T}\}_{t=1}^T\}_{i=1}^N$ and $\{\{M_{i,t|T}\}_{t=1}^T\}_{i=1}^N$, respectively. For ease of exposition we present the algorithm for the Kalman smoother in Appendix A.2.

1.3.2 Estimating parameters of the income process.

The final step in the estimation is to recover the parameters of the income process. The parameters that govern the income process can be found by maximizing the joint likelihood across all individuals i and time periods t. The likelihood for individual i in period t is given by,

$$\begin{aligned} LL_{i,t}(y_{i,t}|l_{i,t},\{y_{i,j},l_{i,j}\}_{j=1}^{t-1}) = &(2\pi)^{-1/2}|H(l_{i,t})'M_{i,t|t-1}H(l_{i,t})+R(l_{i,t})|^{-1/2} \\ &\times \exp\{-\frac{1}{2}(y_t - H(l_{i,t})'\hat{\zeta}_{t|t-1})'(H(l_{i,t})'M_{i,t|t-1}H(l_{i,t})+R(l_{i,t}))^{-1}\} \\ &\times (y_{i,t}-H(l_{i,t})'\hat{\zeta}_{i,t|t-1}). \end{aligned} \tag{7}$$

While there are many approaches to maximizing the likelihood, our preferred approach is to use an EM algorithm (e.g., Dempster et al. (1977)). The EM algorithm starts by treating the unobserved state variable ($z_{i,t}$) as data and writing down the full-information log likelihood, under the assumption that the shocks are normally distributed. By taking expectations, the likelihood becomes a function of our estimate of persistent earnings from the Kalman smoother ($\hat{z}_{i,t|T}$) and the data. The only remaining unknowns in the expected full-information log likelihood are the parameters of the income process. By taking first order conditions and rearranging, we are able to obtain closed form expressions to update the parameters of the income process. These expressions resemble GLS regression equations and can be very tractably implemented.

For example, in Appendix B we show that the persistence parameter (F), as well as the drift to persistent earnings while employed (B_E) and unemployed (B_U) can be estimated via the following GLS style regression,

$$[F\ \ B_E\ \ B_U]' = \left[X_C'Q^{-1}X_C + g_1\right]^{-1}\left[X_C'Q^{-1}Y_C + g_2\right], \tag{8}$$

where g_1 and g_2 are known functions of the covariance of persistent earnings with its lag and the variance of persistent earnings[19], respectively, and

$$X_C \equiv \begin{bmatrix} \hat{z}_{1,0|T} & l_{E,1,1} & l_{U,1,1} \\ \hat{z}_{1,1|T} & l_{E,1,2} & l_{U,1,2} \\ \vdots & \vdots & \vdots \\ \hat{z}_{1,T-1|T} & l_{E,1,T} & l_{U,1,T} \\ \hat{z}_{2,0|T} & l_{E,2,1} & l_{U,2,1} \\ \vdots & \vdots & \vdots \\ \hat{z}_{N,T-1|T} & l_{E,N,T} & l_{U,N,T} \end{bmatrix}_{NT\times 3} \qquad Y_C \equiv \begin{bmatrix} \hat{z}_{1,1|T} \\ \vdots \\ \hat{z}_{1,T|T} \\ \vdots \\ \hat{z}_{N,T|T} \end{bmatrix}_{NT\times 1} \qquad Q^{-1} \equiv \text{diag}\left(\frac{1}{Q(l_{i,t})}\right)_{NT\times NT}. \tag{9}$$

Equation (8) shows that the persistence parameter (F) is updated by regressing lagged persistent earnings (the first column of X_C) onto current persistent earnings (Y_C), and is then adjusted

[19]These functions capture the impact of filtering uncertainty about both current and lagged persistent earnings.

to take into account the covariance of persistent earnings with its lag as well as the variance of lagged persistent earnings. The drift of persistent earnings when employed (B_E) is updated by regressing a dummy variable for being employed (the second column of X_C) onto current persistent earnings (Y_C), while controlling for lagged persistent earnings. Similarly, the drift of persistent earnings when unemployed (B_U) is updated by regressing a dummy variable for being unemployed (the third column of X_C) onto current persistent earnings (Y_C), while controlling for lagged persistent earnings. The GLS regression formula in equation (8) shows that these parameters are identified by running regressions that are informative about the evolution of persistent earnings over time, as well as during employment and unemployment spells.

Similarly, we can obtain simple expressions to update the variance parameters, $Q(\cdot)$ and $R(\cdot)$. For example, in Appendix B we show that these updates for the variance of shocks to persistent earnings among the employed and unemployed are given by

$$Q_E = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} l_{E,i,t} \left[\left(\hat{z}_{i,t|T} - F\hat{z}_{i,t-1|T} - B_E \right)^2 + [1, -F] M_{i,t|T} [1, -F]' \right]}{\sum_{i=1}^{N} \sum_{t=1}^{T} l_{E,i,t}} \tag{10}$$

$$Q_U = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} l_{U,i,t} \left[\left(\hat{z}_{i,t|T} - F\hat{z}_{i,t-1|T} - B_U \right)^2 + [1, -F] M_{i,t|T} [1, -F]' \right]}{\sum_{i=1}^{N} \sum_{t=1}^{T} l_{U,i,t}}, \tag{11}$$

where the first and second terms in the numerator capture the posterior mean and filtering uncertainty of each shock, respectively. Analogously, the updating formula for the variance of shocks to temporary earnings is given by

$$R = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T} l_{E,i,t} \left[\left(y_{i,t} - \hat{z}_{i,t|T} \right)^2 + [1, 0] M_{i,t|T} [1, 0]' \right]}{\sum_{i=1}^{N} \sum_{t=1}^{T} l_{E,i,t}}, \tag{12}$$

and an analogous expression holds for the variance of the initial draw of permanent earnings.

The Role of Distributional Assumptions. In the presentation of the income process and discussion of the EM algorithm we assumed that income shocks were normally distributed. In Appendix A.4 we discuss how neither the Kalman filter nor the EM algorithm require the use of normally distributed shocks; absent this distributional assumption, the Kalman Filter remains unbiased and has an interpretation as a linear minimum variance estimator of the latent state variable. We further provide Monte-Carlo analysis when shocks are non-normal. Finally, we additionally discuss how incorporating spells of unemployment and making shocks functions of observables ($x_{i,t}$) – the subject of our next section – results in earnings that exhibit non-

Gaussian features after integrating out observables even if shocks are Gaussian conditional on observables.

1.4 Adding flexibility: Linking income process parameters to observables

Given the tractability of the EM approach, we discuss in this section how our method easily extends to estimate a rich income process where the parameters depend upon observables (e.g., age, calendar time, education occupation, etc.).

In Section 1.3.2, we showed that via the EM algorithm we are able to sequentially update different blocks of parameters of our income process using closed form updated equations. We show in Appendix B.3.3, that the GLS type formula in equation (8) extends with only minimal modifications to making the drift parameters ($B(\cdot)$) a linear function of observables, denoted $x_{i,t}$. As an example of this tractable implementation, the drift parameters (B_E and B_U) can be allowed to vary over time by adding a series of indicator variables for year and employment status into the X_C matrix from equation 8.

Similarly, making the variance parameters ($Q(\cdot)$ and $R(\cdot)$) log-linear functions of observables results in tractable estimation with observables. It is easy to prove that this functional form for variances yields a concave objective function when we maximize the expected log-likelihood to recover unknown variance parameters. While extensions to missing data and observables in mean equations are well-established in the literature, to our knowledge our way of modeling variances is new to the literature, motivate our description of our Kalman filtering approach as "generalized," mimicking the distinction between ordinary and generalized least squares.

Given this tractability, we can easily extend the EM approach to handle a generalization of equation 1, in which we replace $R(l_{i,t})$ with $R(l_{i,t}, x_{i,t})$, where

$$\mathbb{V}(\omega_{i,t} \mid l_{i,t}; x_{i,t}) = R(l_{i,t}; x_{i,t}) = \exp\left[f_R(l_{i,t}; x_{i,t})\right], \tag{13}$$

and $f_R(l_{i,t}; x_{i,t})$ is linear-in-parameters and $\omega_{i,t}$ is independent of $z_{i,t}$ conditional on $l_{i,t}$ and $x_{i,t}$.[20] For example, consider temporary shocks that only depend on a quadratic in age $a_{i,t}$:

$$R(l_{i,t}; x_{i,t}) = \exp\left[\lambda_1 a_{i,t} + \lambda_2 a_{i,t}^2\right], \;\Rightarrow\; f_R(l_{i,t}; x_{i,t}) = \lambda_1 a_{i,t} + \lambda_2 a_{i,t}^2.$$

The parameters λ_1 and λ_2 are unknown and estimated in the EM algorithm. Likewise, in equa-

[20] "Linear-in-parameters" means that f_R can be written $f_R(l_{i,t}; x_{i,t}) = g(l_{i,t}; x_{i,t})'\Lambda$ for some unknown vector of coefficients Λ and some known linear function of the data $g(l_{i,t}; x_{i,t})'$. We assume that $\mathbb{E}[g(l_{i,t}; x_{i,t})g(l_{i,t}; x_{i,t})']$ is full rank.

tion 2, we can replace $B(l_{i,t})$ and $Q(l_{i,t})$ with

$$B(l_{i,t}; x_{i,t}) = f_B(l_{i,t}; x_{i,t}) \tag{14}$$

$$\mathbb{V}(\nu_{i,t} \mid l_{i,t}; x_{i,t}) = Q(l_{i,t}; x_{i,t}) = \exp\left[f_Q(l_{i,t}; x_{i,t})\right], \tag{15}$$

where $f_B(l_{i,t}; x_{i,t})$ and $f_Q(l_{i,t}; x_{i,t})$ are linear-in-parameters and $\nu_{i,t}$ is drawn independently across individuals and time. In this general case the timing assumptions on observables ($x_{i,t}$) follow in a symmetric manner to the assumptions on the employment process ($l_{i,t}$).[21]

The tractability that we obtain in this general setup allows for estimating rich models of income dynamics. Section 2.2 discusses the set of income models we estimate in this paper, beginning with our baseline from Section 1.1 and extending to models with income risk profiles which are heterogeneous by age, calendar time, and occupation or education. These rich specifications are easily handled via our econometric approach and help to shed additional light on drivers of earnings dynamics.

1.5 Putting it together: the econometric "cookbook"

We conclude this section by summarizing the three steps of the estimation algorithm:

(1.) **Guess the parameters of the income process:** $\{F^{(0)}, B(\cdot)^{(0)}, Q(\cdot)^{(0)}, R(\cdot)^{(0)}\}$

(2.) **Apply the Kalman filter and smoother:** This yields estimates of the path of persistent earnings ($\hat{z}_{i,t|T}$) for each person in every period (Section 1.3.1).

(3.) **Update parameters:** Use $\hat{z}_{i,t|T}$ to evaluate the expected full-information log likelihood and update parameters. First order conditions of the expected full-information log likelihood yield the closed form expressions from the EM algorithm (Section 1.3.2). These imply $\{F^{(1)}, B(\cdot)^{(1)}, Q(\cdot)^{(1)}, R(\cdot)^{(1)}\}$. Iterate until convergence.

The primary gain from step (3.) is fast, tractable updates of parameters in high-dimensional models. An alternative to our estimation procedure is to simply evaluate the conditional likelihood (equation (7)) on grids of parameters, and then select parameters that yield highest value of the conditional likelihood. Such grid search methods, or other global optimization methods, may be feasible in some settings. But the high-dimensional parameter space – especially when

[21] The sequential exogeneity assumption becomes $z_{i,t-1} \perp x_{i,t}, l_{i,t} | \{l_\tau, x_\tau, y_\tau\}_{\tau \le t-1}$. The joint likelihood of $l_t, x_t, y_t | \{l_\tau, x_\tau, y_\tau\}_{\tau \le t-1}$ into two pieces: one which captures the evolution of $l_t, x_t | \{l_\tau, x_\tau, y_\tau\}_{\tau = \le t-1}$ and another which captures $y_t | l_t, x_t, \{l_\tau, x_\tau, y_\tau\}_{\tau \le t-1}$. Finally, the employment process $l_{i,t}$ can also be a function of lagged observables $x_{i,t}$ (e.g., age, tenure, etc.).

parameters are interacted with time trends and demographics – necessitates our use of the EM algorithm.[22] In Appendix B.6 we present the results of a Monte Carlo exercise which shows that the estimation procedure is able to accurately recover unbiased estimates of persistent earnings as well as the parameters of the income process.

2 Data and Estimation

In this section, we introduce the data we will use for our estimation procedure and discuss the income processes we will estimate. We then briefly discuss the results from estimating the income process presented in Section 1.1.

2.1 Data

To estimate the parameters governing the income process and the time series of temporary and persistent earnings, we use annual labor earnings from administrative earnings records that have been linked to survey information. Our source of administrative earnings records is the Social Security Administration's Detailed Earnings Records (DER). The DER is a database of job-level W-2 earnings from 1978 to 2019. We supplement the DER with survey responses from the Annual Social and Economic Supplement (ASEC) to the Current Population Survey (CPS). The ASEC asks questions about labor income, job characteristics (e.g., occupation, industry), labor market events (e.g., layoffs, weeks unemployed) as well as detailed demographic information. Using scrambled Social Security numbers, called Protected Identification Keys (PIKs), the Census Bureau links individuals from the CPS ASEC to their earnings information in the DER.[23]

Our sample includes individuals who were in the ASEC in the years 1973, 1979, 1981-1991, 1994, and 1996-2020. Earnings records from the DER are included in all years in which the individual is observed and not just the years for which an individual is in the ASEC. As in Song et al. (2018) and Guvenen et al. (2020), we use earnings from the DER from 1981 through 2019, owing to concerns about data quality before 1981.[24] For the majority of individuals in our sample, the

[22] Dempster et al. (1977) prove that the EM algorithm is guaranteed to increase the likelihood function at each iteration for general MLE problems. While poor choices of starting values could in principle lead to convergence to a local maximum, we have found our results to be generally quite insensitive to these choices in our applications and convergence to be quite rapid.

[23] See Wagner and Layne (2014) for more information on the assignment of PIKs to survey and administrative data. Note also that going forward we interchangeably use "the CPS", "March CPS", and "CPS ASEC" to denote the CPS ASEC.

[24] See Song et al. (2018) and Guvenen et al. (2020) for additional details. Our results are not sensitive to starting in 1981 rather than 1978.

ASEC provides 2 years of detailed information on demographics, income (labor and non-labor income), labor market information (e.g., weeks worked, occupation, etc.), and a full time series of an individual's labor income over their career from the DER.[25] The survey responses on employment transitions allow us to validate our earnings-based filter and the information on education and occupation allows us to create finely partitioned groups to examine for whom earnings risk has changed over time.

To study earnings dynamics, we focus on a sample of individuals with a minimum degree of labor force attachment. To be included in our estimation sample, an individual must: (1) satisfy a minimum earnings requirement in at least 5 (non-consecutive) years, (2) satisfy the minimum earnings criterion in at least 50% of years (inclusive) between the first and last year that they satisfy the minimum earnings criterion, (3) be between the ages of 25 and 60, and (4) enter the sample by 2013. For conditions (1) and (2), we impose a minimum earnings criterion equal to the average Social Security Administration (SSA) cutoff for receiving a full-year of credits toward SSA retirement benefits, which is $3,350 (in 2005 PCE dollars) between 1978 and 2019.

These criteria allow for long spells of zero earnings, potentially equal to half of the individual's panel of earnings. Condition (4) is included so that entrants to the sample in the final year are not selected towards individuals with the strongest labor force attachment (i.e., individuals with earnings above the minimum threshold for 5 consecutive years). Finally, to focus on labor market risk for workers, we additionally remove from the sample individuals who have self-employment income that exceeds 50% of their total income (labor income plus self-employment income) in at least 5 years. These sampling criteria result in a sample of over 1.7 million individuals.[26] We explore alternate sampling criteria in Appendix C.9.

We use earnings information from the DER to study income risk. Our measure of income is the sum of Box 1 (total wages, tips, and bonuses) and Box 12 (earnings deferred to a 401(k) type account) earnings across all jobs the individual held during the year. We report earnings in 2005 dollars, where earnings are deflated by the PCE price index. To remove the impact of outliers, we winsorize real earnings at the 99.9th percentile in each year. Table 1 provides summary statistics for the individuals in our sample.[27]

Representativeness. In Appendix C.2, we assess sample representativeness. We show that the time series of standard deviations of earnings by gender closely mirror those reported in Guvenen et al. (2022), hereafter referred to as GKSW, who use the full sample of SSA earnings

[25]In our estimation, we use an individual's sampling weight from the ASEC.

[26]Owing to Census Bureau disclosure rules, the number of individuals is rounded to the nearest thousand.

[27]Note that we do not use an individual's reported education if they are less than 25 when in the ASEC. We do so to avoid mis-classifying individuals who have not yet completed their education. In Table 1 these individuals are classified as *Education Not Reported*.

Table 1: Summary statistics

Variable	Mean
Real Annual Earnings	\$41,240
Age	41.1
Share Unemployed	6.6%
Share Less than College Degree	59.6%
Share College Degree Plus	28.1%
Share Education Not Reported	12.2%
Share Male	51.2%
Observations	37,540,000
Individuals	1,736,000

Note: See Section 2.1 for sample selection criteria. Real annual earnings are measured in 2005 dollars. The variable "share unemployed" is the share of individuals whose average earnings in a given year do not satisfy the minimum earnings criterion. We classify an individual's education level as not reported if they are less than 25 when in the ASEC.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

records and impose very similar earnings criteria for sample inclusion.

2.2 Income models

In this section, we present the three models of income dynamics that we will estimate as part of this paper. Our first model is based on Section 1.1. Our second model includes means and variances that depend on time and age. Our third model allows all parameters from the second model to vary by education or occupation.

Model 1. This is our benchmark model outlined in Section 1.1.

$$\begin{aligned}
y_{i,t} &= \begin{cases} z_{i,t} + \omega_{i,t}, & \omega_{i,t} \sim N(0, R) \quad \text{if } l_{E,i,t} = 1 \\ \cdot & \quad\quad\quad\quad\quad\quad\quad\;\; \text{if } l_{E,i,t} = 0 \end{cases} \\
z_{i,t+1} &= \begin{cases} F\, z_{i,t} + B_E + \nu_{i,t+1}, & \nu_{i,t+1} \sim N(0, Q_E) \quad \text{if } l_{E,i,t} = 1 \\ F\, z_{i,t} + B_U + \nu_{i,t+1}, & \nu_{i,t+1} \sim N(0, Q_U) \quad \text{if } l_{E,i,t} = 0 \end{cases} \\
z_{i,0} &\sim N(0, Q_0)
\end{aligned} \tag{16}$$

Model 2. We modify Model 1 to allow means and variances to depend on time and age.[28] We also split unemployment spells into the first year of unemployment and all future years of unemployment. In this model, E denotes employment, U denote the first year of unemployment, and N denote future years of unemployment. We use this model in Section 3 to examine the trends in earnings risk over time. Let $\mathbf{1}_t$ denote time fixed effects, let a denote age, and let $f_j(a) = \lambda_1^j a + \lambda_2^j a^2$ denote a quadratic in age for a given labor market status and/or parameter j. As we discuss in Section 1.4 to tractably estimate the variance parameters ($Q(\cdot)$ and $R(\cdot)$) as functions of age and time, we model log variances as being linear in parameters. In the specification below, we use lower case letters to denote these "logarithmic" parameters. When we report results, we exponentiate the lower case variables so that they correspond to $Q(\cdot)$ and $R(\cdot)$.

$$
\begin{aligned}
y_{i,t} &= \begin{cases} z_{i,t} + \omega_{i,t} & \omega_{i,t} \sim N\left(0, e^{\sum_t r_t \mathbf{1}_t + f_R(a)}\right) \quad \text{if } l_{E,i,t} = 1 \\ \cdot & \text{if } l_{E,i,t} = 0 \end{cases} \\
z_{i,t+1} &= \begin{cases} F\, z_{i,t} + \sum_t B_{E,t}\mathbf{1}_t + f_{B,E}(a) + \nu_{i,t+1}, & \nu_{i,t+1} \sim N\left(0, e^{\sum_t q_{E,t}\mathbf{1}_t + f_{Q,E}(a)}\right) \quad \text{if } l_{E,i,t} = 1 \\ F\, z_{i,t} + \sum_t B_{U,t}\mathbf{1}_t + f_{B,U}(a) + \nu_{i,t+1}, & \nu_{i,t+1} \sim N\left(0, e^{\sum_t q_{U,t}\mathbf{1}_t + f_{Q,U}(a)}\right) \quad \text{if } l_{E,i,t} = 0,\ l_{E,i,t-1} = 1 \\ F\, z_{i,t} + \sum_t B_{N,t}\mathbf{1}_t + f_{B,N}(a) + \nu_{i,t+1}, & \nu_{i,t+1} \sim N\left(0, e^{\sum_t q_{N,t}\mathbf{1}_t + f_{Q,N}(a)}\right) \quad \text{if } l_{E,i,t} = 0,\ l_{E,i,t-1} = 0 \end{cases} \\
z_{i,0} &\sim N\left(0, e^{\sum_t q_{0,t}\mathbf{1}_t + f_{Q,0}(a) + f_{Q,0,t\leq 1984}(a)\mathbf{1}_{t\leq 1984}}\right)
\end{aligned}
\tag{17}
$$

Model 3. We modify Model 2 to allow all parameters to vary by an individual's education (Section 4.1), coarse occupation groups ((Section 4.2)), and detailed occupation codes ((Section 4.3)). In particular, this estimation allows the persistence parameter F to differ by group (e.g., by education, or occupation).

In using these models, a first step in the estimation is removing the predictable life-cycle component of log earnings (i.e., residualizing). We perform this step as in Guvenen et al. (2014) (see Appendix C.1 for details). When removing the predictable life-cycle component of log earnings, we only use earnings that satisfy our minimum earnings requirement. When estimating Model 3, where we allow parameters to vary by education (occupation), we remove the predictable component of earnings separately for each group.[29]

[28]This estimation combines ideas from Blundell et al. (2008), who estimate time-varying temporary and persistent risk, with Karahan and Ozkan (2013), who estimate how temporary and persistent risk vary over the life-cycle. Additionally, we allow for spells of unemployment.

[29]Similarly in Appendix C.8, when we allow parameters to vary by gender, we remove the predictable component of earnings separately for each gender.

Results of estimating Model 1. Table 2 presents the parameter estimates from estimating Model 1.[30] The autoregressive parameter shows that the process is persistent ($F = 0.93$).[31] Additionally, the estimation reveals that compared with employed individuals, unemployed individuals experience persistent earnings shocks with both a different mean (captured by the drift) and a different variance. In particular, when an individual is unemployed, the variance of persistent earnings shocks is nearly four times larger than employed individuals ($Q_U = 0.275$, $Q_E = 0.069$). Additionally, an unemployed individual's persistent earnings decline on average by nearly 19% per year, while an employed individual's persistent earnings increase by 0.05% per year. Hence unemployed individuals draw persistent earnings shocks from a distribution with a significantly lower mean and greater dispersion relative to that of employed individuals.

Model validation. In addition to recovering income process parameters, a benefit of our approach is that we obtain estimates of persistent earnings for each individual in each year, which we hereafter refer to as "filtered estimates." In Appendix C.3, we compare these filtered estimates with observable labor market events (e.g., layoffs, job switching) in order to validate the filter. We show that job loss is associated with large declines in persistent and temporary earnings, but being recalled to a previous employer is linked with more muted declines in temporary and persistent earnings. Additionally, we find that job switchers experience greater dispersion in temporary and persistent shocks relative to stayers, and that moving to a higher (lower) paying firm is associated with more positive (negative) shocks to persistent earnings.

3 The changing nature of earnings risk

In this section, we estimate Model 2 to measure time trends in persistent and temporary earnings risk.[32] We show that since the 1980s, persistent earnings risk has risen, while temporary

[30] We compute standard errors on our parameters using a block bootstrap procedure. See Appendix A.5 for details.

[31] Existing estimates of persistence of persistent earnings in mixture models range from 0.953 to 0.999 (see, e.g., Guvenen et al. (2014)). However, we make two departures from the literature, which make comparison difficult. First, the mean (drift) of the shock to persistent earnings depends upon an individual's employment status. Second, the prior literature drops observations with zero earnings (i.e., earnings below a minimum earnings criteria). Since our approach produces an estimate of persistent earnings even when an individual has zero labor earnings, these observations are taken into account when estimating F.

[32] Model 2 splits unemployment spells into the first year of unemployment and all future years of unemployment. As shown in Section 1.1, we require four years worth of data to identify persistent earnings for future years of unemployment. Therefore, we bin together the first four years (1981-1984) and last four years (2016-2019) into single fixed effects. Since they do not affect our main results, we omit them from our graphs for ease of exposition. Finally, Model 2 includes age quadratics to control for compositional changes over the sample period. When presenting time trend results, we present the results for a 25-year old and relegate more detailed life-cycle analysis to Appendix C.4.

Table 2: Parameter estimates for Model 1

Description	Parameter	Value
Persistence	F	0.9316
		(0.0002)
Variance of Shocks to Peristent Earnings (Emp.)	Q_E	0.06894
		(0.0001)
Variance of Shocks to Peristent Earnings (Unemp.)	Q_U	0.2751
		(0.0007)
Drift of Persistent Earnings (Emp.)	B_E	0.0005
		(0.0001)
Drift of Persistent Earnings (Unemp.)	B_U	-0.1877
		(0.0006)
Variance of Initial Draw of Persistent Earnings	Q_0	0.6733
		(0.0014)

Note: Table presents parameter estimates from estimating the income process in Section 1.1 (Model 1). Bootstrapped standard errors in parenthesis.

Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

earnings risk has declined. We further show that the decline in persistent earnings during spells of unemployment has worsened. In the following section, we exploit the rich demographic, geographic, and occupation information in the linked SSA-CPS data to characterize for whom earnings risk has changed.

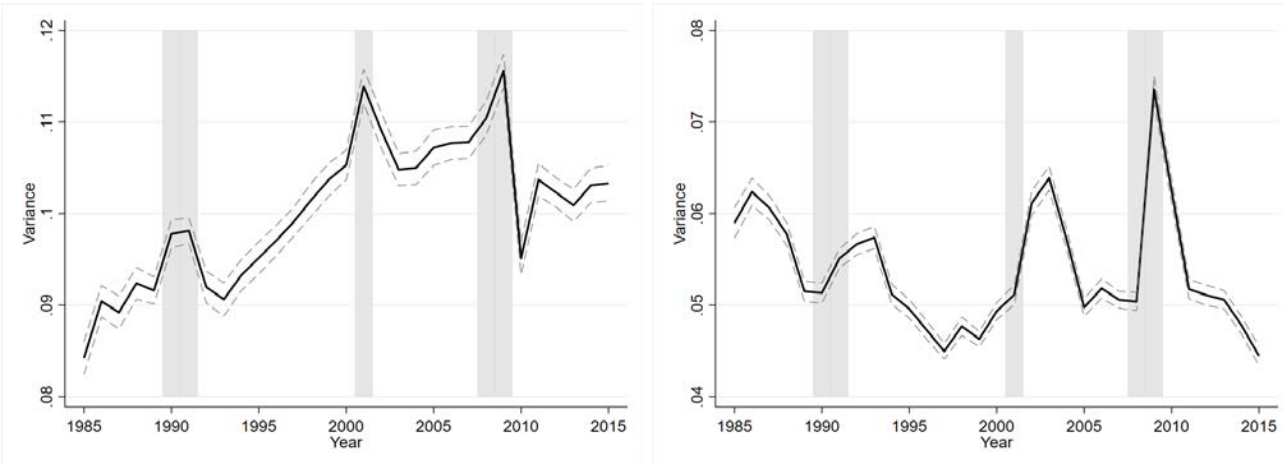
Employed Risk. We first present how earnings risk among the employed has evolved over time. Panel (a) of Figure 1 presents the evolution of persistent earnings risk among the employed between 1985 and 2015. The figure shows that there was a steady increase in the variance of persistent earnings shocks between 1985 and the start of the Great Recession in 2007. Over this time period, the variance of persistent earnings shocks increased by more than 35 percent. In the aftermath of the Great Recession, there was a decline in persistent earnings risk. In Appendix C.6, we show that the decline in persistent earnings risk following the Great Recession occurred among individuals over the age of 45. Despite the decline in persistent risk after the Great Recession, persistent risk is still 20 percent higher in 2015 than 1985.[33]

At the same time that persistent earnings risk has been increasing, temporary earnings risk has declined. Panel (b) presents the variance of shocks to temporary earnings between 1985

[33]Estimating Model 2 also produces an estimate of the mean of persistent shocks to the employed over time. We present this time series in Appendix C.4. The time series does not show any long-run trends, but insteady is highly cyclical with expansions characterized as periods of positive mean persistent shocks and recessions as periods with negative persistent shocks.

Figure 1: Persistent and temporary earnings risk of the employed

(a) Variance of Persistent Earnings Shocks (b) Variance of Temporary Earnings Shocks



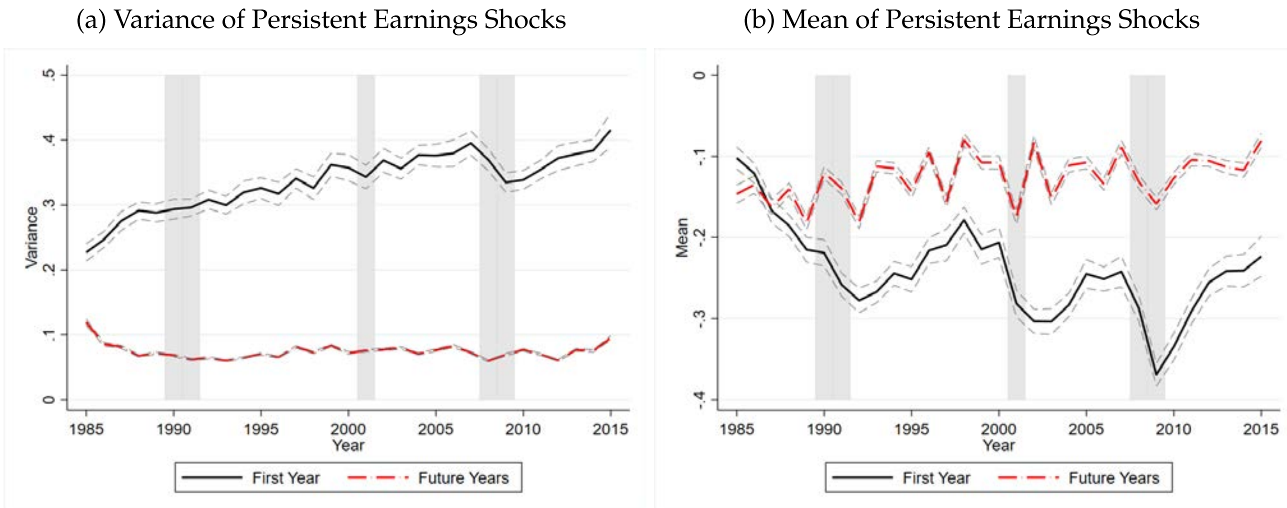
Note: Figure presents the results of estimating Model 2. Panel (a) presents the variance of persistent earnings shocks among the employed. Panel (b) presents the variance of temporary earnings shocks. Gray dashed lines represent a 95% confidence interval, and gray bars denote NBER recession dates.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

and 2015. The figure shows that over this time period, temporary earnings risk has declined by over 20 percent (from 0.059 in 1985 to 0.045 in 2015). The time series also shows that there is substantial cyclical variation in temporary earnings risk. In and around each recession, temporary earnings risk spikes as larger shares of individuals pass through job loss spells that do not necessarily result in a full-year unemployment.

While earnings risk among the employed is the focus of much of the existing literature on income risk, it provides an incomplete characterization of overall earnings risk. A benefit of our approach is that it allows us to incorporate spells of unemployment and estimate the risk that unemployed workers face. We next examine how persistent risk among the unemployed has evolved over time.

Unemployed Risk. Figure 2 presents how the shocks to persistent earnings among the unemployed have evolved over time. In estimating Model 2, we split the unemployed into individuals who are in their first year of unemployment and individuals who are in their second or later years of unemployment, which we hereafter refer to as "future years" of unemployment. In panel (a) of Figure 2, we present the variance of persistent earnings shocks to the unemployed over time. First, we find that the variance of shocks to persistent earnings in an individual's first year of unemployment (black, solid line) is larger than in their future years of unemployment

Figure 2: Persistent earnings risk of the unemployed



Note: Figure presents the results of estimating Model 2. Panel (a) presents the variance of persistent earnings shocks among the unemployed. Panel (b) presents the mean of persistent earnings shocks among the unemployed. The black, solid line presents estimates for individuals who are in their first year of unemployment. The red, dashed line represents individuals who are in their second or later years of unemployment (future years). Gray dashed lines represent a 95% confidence interval, and gray bars denote NBER recession dates.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

(red, dashed line). Second, we find that the variance of shocks to persistent earnings among the unemployed in their first year of unemployment increased by 80 percent over our sample period. Conversely, the variance of shocks to persistent earnings among unemployed individuals in their future years of unemployment declines slightly at the start of our sample and then remains largely stable for the rest of our time period.

Panel (b) of Figure 2 presents the mean persistent earnings shock among the unemployed over time. The figure shows that for most of the sample period, the decline in persistent earnings during the first year of unemployment is larger than during their future years of unemployment. Additionally, the figure shows that the decline in persistent earnings during the first year of unemployment has become more severe over time. In 1985, entering unemployment was associated with a decline in persistent earnings of 10 percent. By 2010, entering unemployment was associated with a decline of persistent earnings of nearly 30 percent. There is also substantial cyclical variation in the mean persistent earnings shock for individuals entering unemployment. We find that the decline in persistent earnings from entering unemployment is larger in recessions, consistent with the work of Davis and von Wachter (2011) who show that the costs of job loss are larger in recessions.

Combined Risk. In Figures 1 and 2, we showed how risk has evolved separately for the employed and unemployed. These estimates are informative about risk within an employment state, but they ignore transitions across employment states (e.g., entering into unemployment). In this section, we introduce a measure of *combined* persistent income risk, which explicitly takes into account the likelihood that an individual transitions across employment states and how the likelihood of these transitions has evolved over time. We compute our measure of combined persistent income risk, denoted by Q_t, using the law of total variance. The parameter estimates from Model 2 and the share of individuals in each employment state are sufficient statistics for combined income risk.[34]

The black, solid line in panel (a) of Figure 3 presents our estimate of how combined persistent income risk has evolved since 1985. The figure shows that combined persistent income risk increases from the mid-1980s up to the Great Recession. In the aftermath of the Great Recession, combined persistent income risk falls but remains nearly 20 percent above its value in 1985.

A benefit of our measure of combined persistent income risk is that it allows us to examine how changes in the likelihood of transitioning across employment states impacts the risk that individuals face in the labor market. In panel (b) of Figure 3, we show that the share of individuals in their first year of unemployment – which can be interpreted as the annual entry rate into unemployment – has been steadily declining over our sample period. Recent work by Fujita (2018) shows that the likelihood of making an employment-to-unemployment (EU) transition has been declining in the US over this time period in the CPS. We plot Fujita (2018)'s monthly EU rate on the right axis of Figure 3. Despite the conceptual differences between the two series (i.e., we use an earnings threshold while Fujita (2018) uses self-reported employment status), our entry rate into unemployment is highly correlated with Fujita (2018)'s EU rate.

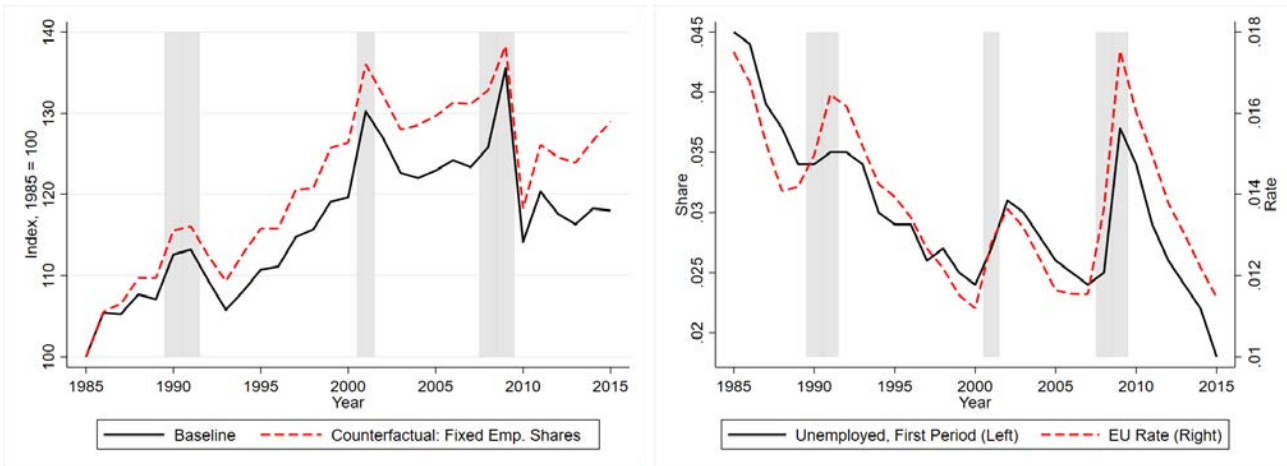
Given that the variance of shocks to persistent earnings is higher among individuals entering into unemployment relative to the employed, the decline in the share of individuals entering unemployment puts downward pressure on combined persistent earnings risk. To quantify how changes in employment shares over time have influenced the evolution of combined persistent risk, we compute a counterfactual measure that freezes the shares of employment and

[34] Let Q_t denoted combined persistent risk in period t. Let $Q_{E,t}$ denotes the variance of persistent shocks to the employed in period t, $Q_{U,t}$ ($Q_{N,t}$) denotes the variance of persistent income shocks to the unemployed in their first (future) period of unemployment. Define $B_{k,t}$ as the mean of shocks for the employed $k = E$, unemployed in their first period ($k = U$) and future periods of unemployment ($k = N$). Finally, Let $\hat{u}_t$ denote the share of agents who are in their first period of unemployment, $\hat{n}_t$ denote the share of individuals who are in future periods of unemployment, and $\hat{e}_t = 1 - \hat{u}_t - \hat{n}_t$ as the share of agents who are employed. Using the law of total variance we can compute combined persistent risk in period t via,

$$Q_t = Q_{E,t}\hat{e}_t + Q_{U,t}\hat{u}_t + Q_{N,t}\hat{n}_t + (B_{E,t})^2(\hat{e}_t)(1-\hat{e}_t) + (B_{U,t})^2(\hat{u}_t)(1-\hat{u}_t) + (B_{N,t})^2(\hat{n}_t)(1-\hat{n}_t) - 2\left[B_{U,t}\hat{u}_tB_{E,t}\hat{e}_t + B_{N,t}\hat{n}_tB_{E,t}\hat{e}_t + B_{N,t}\hat{n}_tB_{U,t}\hat{u}_t\right]$$

Figure 3: Combined persistent risk over time

(a) Combined Persistent Risk (b) Share Entering into Unemployment



Note: Panel (a) plots combined persistent income risk over time (black, solid line) and a counterfactual measure of combined persistent income risk, which holds employment and unemployment shares fixed at their 1985 values (red, dashed line). Panel (b) plots the share of individuals who are in their first period of unemployment. The black, solid line presents the share of individuals in our sample in their first year of unemployment (left, axis). The red, dashed line is the monthly employment-to-unemployment (EU) transition rate as estimated in the CPS and produced by Fujita (2018) (right, axis). Gray bars denote NBER recessions.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

unemployment at their 1985 values. The red, dashed line in Panel (a) of Figure 3 presents this counterfactual path of combined persistent risk and shows that had employment shares remain fixed at their 1985 values, combined persistent risk would have increased by 30 percent over the sample period (instead of 20 percent). Thus, declining inflows into unemployment have been a mitigating factor in the increase in persistent risk that individuals face in the labor market.

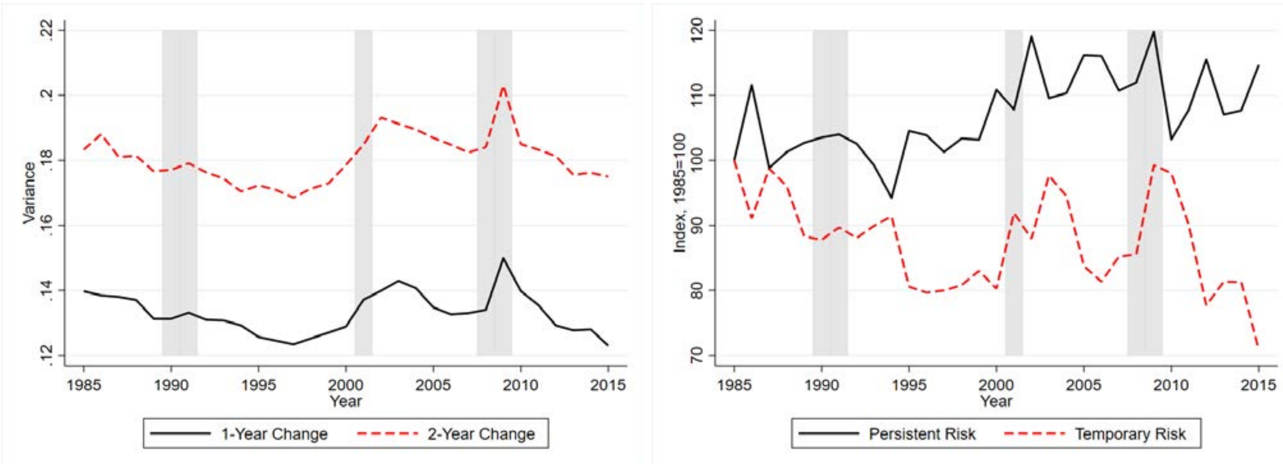
3.1 Identifying trends in income risk

In this section, we examine the data moments that allow us to identify how income risk has evolved over time. The arguments are similar to Section 1.2, except that we relax the assumption of a unit root. Intuitively, just as in the unit root case, comparing quasi-differences ($:= x_t - F^k x_{t-k}$) at different horizons k is sufficient to identify persistent and temporary income risk. Appendix A.1 formally proves this.

Panel (a) of Figure 4 plots the variance of the quasi-difference of changes in log earnings over a 1-year horizon (black, solid line) and a 2-year horizon (red, dashed line) among individ-

Figure 4: Identifying changes in risk over time among employed

(a) Variance of Earnings Changes At Different Horizons (b) Implied Path of Temporary and Persistent Risk



Note: Panel (a) plots of the variance of the quasi-difference in log earnings over a one year horizon (black, solid line) and a two year horizon (red, dashed line), where the individual was employed in the middle year. Panel (b) plots the implied path of persistent income risk (black, solid line) and temporary income risk (red, dashed line) using the moments from Panel (a) and the identification argument in Appendix A.1. Gray bars denote NBER recession dates.

Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

uals who are employed in the middle year.[35] The figure shows that the variance of earnings changes over a 1-year horizon has declined between 1985 and 2015, while the variance of earnings changes over a two-year horizon have been largely stable.

In panel (b) of Figure 4, we plot the path of persistent income risk (black, solid line) and temporary earnings risk (red, dashed line) implied by the identification argument in Appendix A.1. The figure shows that persistent income risk rises between 1985 and 2015, while temporary earnings risk decreases. The figure also shows that we obtain similar cyclical patterns using these "simple moments" as in our full estimation from Section 3. In particular, the implied path of persistent earnings risk experiences a notable decline after the Great Recession, while temporary earnings risk spikes around the 2001 and 2008-09 recessions.

These results highlight that using 1-year and k-year log earnings changes cannot be treated as separate proxies for temporary and persistent risk. In fact, the variance of 2-year earnings changes can have no trend or can even decline, while persistent earnings risk can rise. Only by positing a model structure can the joint evolution of the variance of log earnings changes

[35] For the quasi-differences, we use the measure of persistence (F) as implied by equation (20), which measures F using the ratio of covariances of earnings over different horizons. This produces an estimate of $F = 0.9392$.

be used to identify persistent and temporary risk. A benefit of our filtering exercise is that it examines the full path of an individual's earnings (and changes in earnings) to inform the estimates of temporary and persistent earnings risk.

3.2 Heterogeneity and robustness

We conclude this section by briefly describing a series of additional results.

Heterogeneity by Age. In Appendix C.6, we examine how income risk has changed over time across the age distribution by estimating a model where the means and variances of shocks to persistent and temporary earnings are allowed to vary by age and decade. In panel (a) of Figure 5, we show how combined persistent income risk has changed between the 1980s and the 2010s. The figure shows that the increase in combined persistent income risk has occurred for workers of all ages but is most pronounced among young workers. In panel (b) of Figure 5, we show how the decline in persistent persistent earnings during the first year of unemployment has evolved between the 1980s and the 2010s by age. The figure shows that there are larger declines in persistent earnings during unemployment spells in the 2010s relative to the 1980s for workers of all ages, but that this acceleration has been most pronounced among older workers. Finally, in Appendix C.6 we show that the decline in temporary risk has been largest among younger workers.

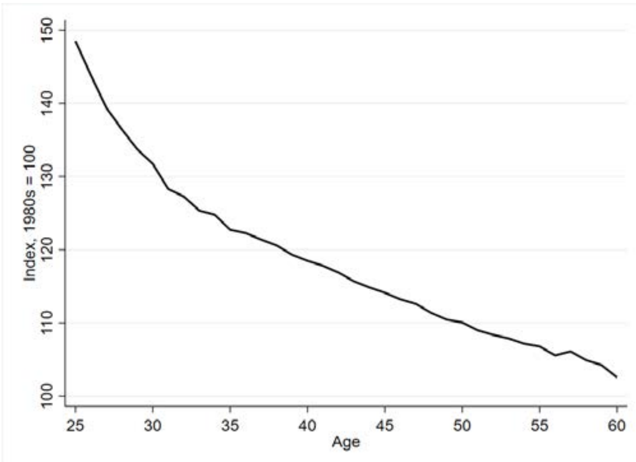
Role of Job Switching / Staying. In Appendix C.7, we examine how income risk has evolved over time for job switchers and job stayers. We find that persistent earnings risk among the employed has increased for both job switchers and job stayers. Similarly, we find that the decline in temporary earnings risk has occurred among both job switchers and job stayers.

Heterogeneity by Gender. In Appendix C.8, we present the results for the evolution of persistent and temporary income risk of men and women over time. We find that persistent income risk has risen for both men and women, while temporary risk has declined for both genders. Despite the similar secular trends, we find that there is more cyclical variation in the income risk for men, reflecting that men tend to work in jobs that are more exposed to the business cycle (e.g., Doepke and Tertilt (2016)).

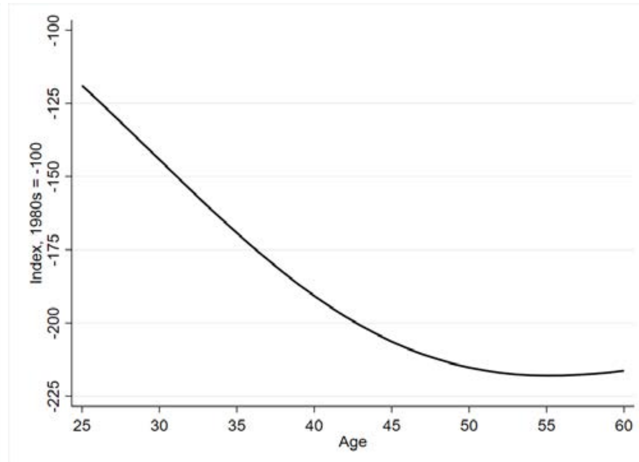
Robustness. Finally, we briefly discuss the results of several robustness exercises. In Appendix C.9, we show that our headline results are robust to alternate minimum earnings cutoffs. In Appendix C.10, we show that we obtain nearly identical estimates of the parameters of our income process over time if we start the estimation 4-years later or end the estimation 4-years earlier.

Figure 5: Changes in persistent risk by age from 1980s to 2010s

(a) Combined Persistent Risk, ratio 2010s to 1980s



(b) Drift First Period Unemployed, ratio 2010s to 1980s



Note: The figure presents the results of estimating Model 4 as presented in Appendix C.6. Panel (a) presents combined persistent earnings in the 2010s relative to the 1980s as function of age, where 1980 is normalized to 100. Panel (b) presents the drift to persistent earnings during the first period of unemployment in the 2010s relative to the 1980s as a function of age where the 1980s is normalized to -100 (values less than -100 imply more negative drift in 2010s).
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

4 For whom is earnings risk changing?

We have shown that persistent earnings risk has increased, and the decline in persistent earnings during spells of unemployment has worsened. In this section, we exploit the demographic, geographic, and occupation information in the linked SSA-CPS data to characterize for whom these changes are occurring. We frame this discussion by testing several hypotheses for rising persistent earnings risk. We do not find evidence supporting hypotheses related to regional explanations of persistent earnings losses, including the decline of the Rust Belt. We additionally do not find evidence to support explanations based on declining routine employment (e.g., Acemoglu and Autor (2011)). Instead, we show that the rise in persistent earnings risk is a high skill worker phenomenon and provide supportive evidence that the rise in persistent risk among high skill workers is due to exposure to the introduction of new technologies. We organize this section around tests of three hypotheses:

Hypothesis 1: Rising persistent earnings risk is due to the introduction of new, skill-biased technologies and concentrated among high-skill workers.

Hypothesis 2: Rising persistent earnings risk is due to the decline of manufacturing and union protection and concentrated geographically in the Rust Belt.

Hypothesis 3: Rising persistent earnings risk is due to the decline of routine skill intensive occupations.

To rule out concerns regarding parameter restrictions among these disparate groups, we utilize Model 3 in this section, which extends Model 2 to make all parameters vary by group (e.g., education or occupation) and we separately residualize earnings by group. For ease of presentation, we only present results for Hypothesis 1 in the main text, and we provide evidence against Hypotheses 2 and 3 in Appendices C.11 and C.12.1, respectively. In the subsections that follow, we provide three pieces of evidence in favor of the high skill hypothesis by examining how risk has evolved by education level (Section 4.1), coarse occupation groups called 'job zones' (Section 4.2) as well as by more detailed occupation classifications (Section 4.3). Finally, we show that rising persistent earnings risk is associated with greater exposure to new, skill-biased technologies in Section 4.3.

4.1 Increasing risk among the high skill: Education

Our first piece of evidence for the high-skill worker hypothesis is an analysis of income risk among workers of different education levels. To do so, we estimate Model 3 in which the parameters of the income process differ by an individual's recorded education level in the CPS. We consider five educational groups: (1) less than high school, (2) high school graduate, (3) some college, (4) college graduate, and (5) more than a college degree.[36]

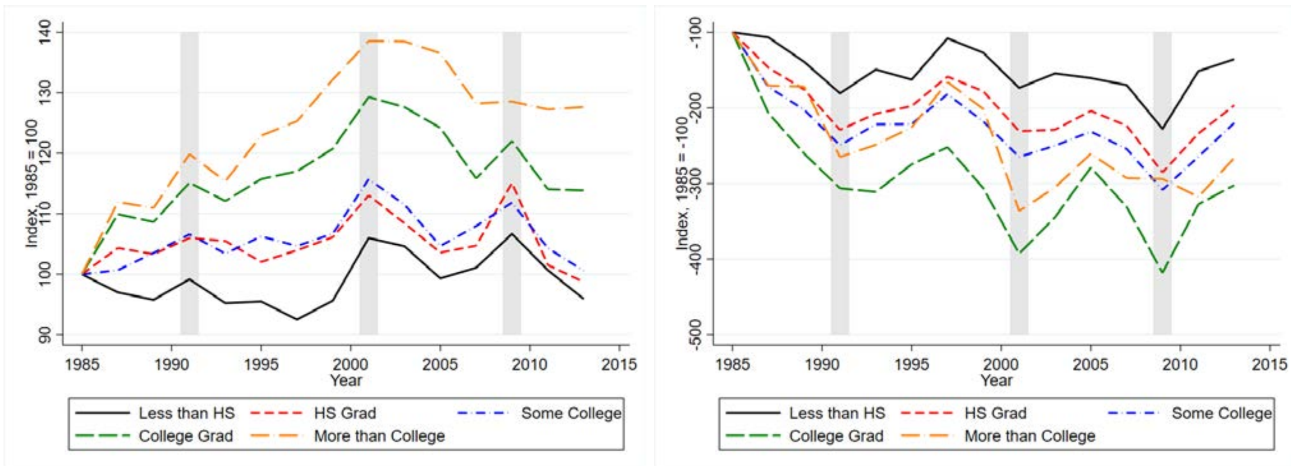
Panel (a) of Figure 6 presents our estimates of combined persistent income risk by education group. The figure shows that the increase in combined persistent income risk has been largest among the most highly educated individuals in the economy. For individuals with more than a college degree (orange, long dashed-dotted line), combined persistent risk has increased by nearly 30 percent from the mid-1980s to the mid-2010s. Among college graduates (green, long dashed line), the increase in persistent income risk has been approximately 15 percent. Conversely, by the end of the sample period, there has been effectively no change in combined persistent income risk for individuals with less than a college degree.

We next examine how the mean shock to persistent earnings has evolved during unemployment spells by education level in panel (b) of Figure 6. Similarly, the figure shows that the greater scarring effect of unemployment over time has been most pronounced among highly

[36]To avoid misclassification of education levels, we only use an individual's reported education if they are 30 or older in their CPS observation. Given the larger number of groups, we use 2-year binned fixed effects in the estimation.

Figure 6: Changes in persistent risk over time by education

(a) Combined Persistent Risk (b) Drift First Period of Unemployment



Note: The figure presents the results of estimating Model 3, where parameters vary by education group. Panel (a) presents combined persistent earnings risk over time by education group. Panel (b) presents the drift to persistent earnings during the first period of unemployment by education group. The black, solid line corresponds to individuals with less than a high school degree. The red, dashed line corresponds to individuals with a high school degree. The blue, dash-dotted line corresponds to individuals with some college. The green, long dashed line corresponds to individuals with a college degree. The orange, long dashed-dotted line corresponds to individuals with more than a college degree. Gray bars denote NBER recession dates.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

educated workers. By the mid-2010s, the decline in persistent earnings from entering unemployment worsened by a factor of three for individuals with a college degree as well as for individuals with more than a college degree. Conversely, for individuals with less than a high school degree (black, solid line) the decrease in the mean shock to persistent earnings is approximately 35 percent larger by the end of the sample.

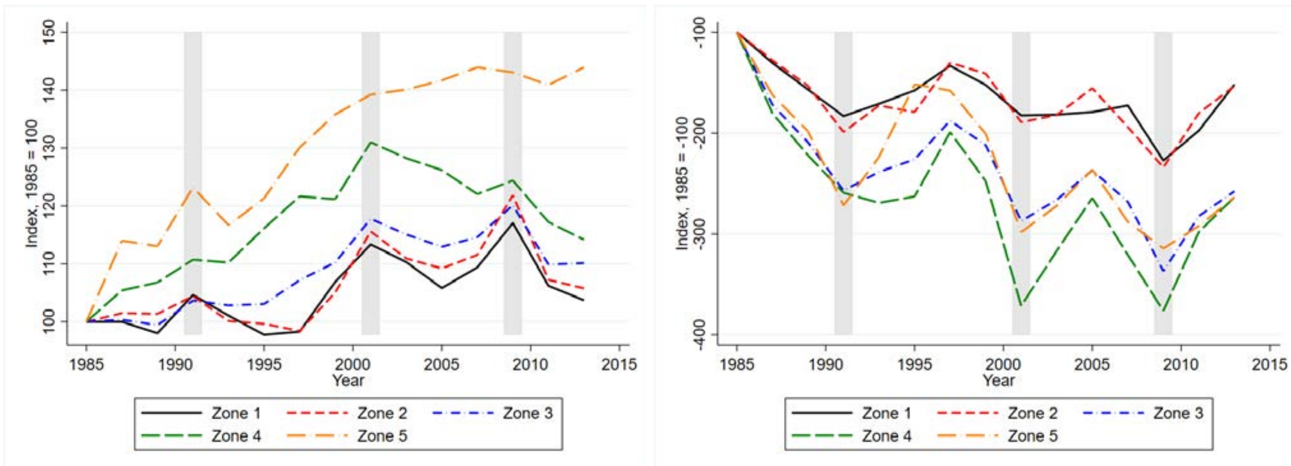
4.2 Increasing risk among the high skill: ONET job zones

Our second piece of evidence in favor of the high skill hypothesis examines how income risk has evolved over time using a coarse grouping of occupations based upon ONET job zones. ONET places an occupation into one of 5 “job zones,” where occupations in zone (group) 1 include jobs that that require "little or no preparation", while jobs in zone 5 require "extensive preparation."[37] Thus, these job zones give us another way to sort occupations by their skill level

[37] Jobs in zone 2 are described as requiring "some preparation," jobs in zone 3 require "medium preparation," and jobs in zone 4 require "considerable preparation." See https://www.onetonline.org/help/online/zones for

Figure 7: Changes in persistent risk over time by ONET job zone

(a) Combined Persistent Risk (b) Drift First Period of Unemployment



Note: The figure presents the results of estimating Model 3, where parameters vary by ONET Job zone. Panel (a) presents combined persistent earnings risk over time by ONET job zone. Panel (b) presents the drift to persistent earnings during the first period of unemployment by ONET job zone. The black, solid line corresponds to individuals in job zone 1. The red, dashed line corresponds to individuals in job zone 2. The blue, dash-dotted line corresponds to individuals in job zone 3. The green, long dashed line corresponds to individuals in job zone 4. The orange, long dashed-dotted line corresponds to individuals in job zone 5. Gray bars denote NBER recession dates. Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

with zones 1 through 5 representing an ordering of skill levels with zone 5 being the highest.[38] In Appendix C.14, we present more details on ONET job zones, and include examples of which occupations are included in each job zone.

Panel (a) of Figure 7 plots combined persistent income risk by ONET job zone. The figure shows that the increase in combined persistent income risk has been largest among individuals who work in the highest ONET job zone (orange, long dashed-dotted line). In particular, for individuals in these occupations, which require extensive preparation, combined persistent income risk has increased by over 40 percent between the mid-1980s and the mid-2010s. At the other end of the spectrum, for individuals employed in job zone 1 (black, solid line), which require little to no preparation, combined persistent income risk increased by approximately 5 percent over the same period.

Panel (b) of Figure 7 presents the decline in persistent earnings associated with entering

more information on ONET job zones. Link last accessed on October 9, 2023.

[38]We use ONET job zones from ONET Vintage 15.1. In classifying an individual's occupation, we use an individual's first reported occupation in the CPS, and only include individuals who are 30 or older in their first CPS observation. Given the larger number of groups, we use 2-year binned fixed effects in the estimation.

into unemployment by ONET job zone. The figure shows that the acceleration of the decline in persistent earnings from entering unemployment has occurred among the highest ONET job zones. In these occupations which range from requiring "medium preparation" (zone 3) to "extensive preparation" (zone 5), we see that the decline in persistent earnings from entering unemployment has increased by a factor of three over the sample period. Conversely, among the lowest ONET job zones, we have seen a much more muted acceleration of the decline in persistent earnings associated with entering unemployment.

4.3 Increasing risk among the high skill: Detailed occupations

Finally, we present a third piece of evidence in favor of the high skill hypotheses by exploiting changes in risk within detailed occupations and relating these changes in risk to the skill content of an occupation.[39] Using an individual's earliest reported occupation in the CPS, we classify individuals into one of 334 time-consistent occupation codes developed by Autor and Dorn (2013).[40] We then estimate Model 3 where the parameters of our income process are allowed to differ by each occupation. Given the large number of occupations, we use 5-year windows for the time fixed effects.

We test the high skill workers hypotheses by using three measures of an occupation's skill intensity. We first split occupations by their degree of "Non-Routine Cognitive Analysis" skills (henceforth, *non-routine cognitive skills*) as measured by Acemoglu and Autor (2011) using O*NET data.[41] We additionally measure the degree to which an occupation is high-skill by using its mean years of completed education and mean earnings.[42] To ease the comparison across measures, we normalize each measure to have mean zero and unit standard deviation. Let X_o be a measure of the skill-content of occupation o (e.g. non-routine cognitive task content, etc.). Let $\Delta Y_o = Y_{o,(2010-2015)} - Y_{o,(1985-1989)}$ denote the change in parameter Y (e.g., the standard deviations of shocks to persistent earnings among employed etc.) for occupation o between time period $2010-2015$ and $1985-1989$. The specification we use is of the form,

$$\Delta Y_o = \alpha + \eta X_o + \epsilon_o. \tag{18}$$

[39] In Appendix C.13, we repeat this analysis using an individuals industry as recorded in the LBD during their first CPS year. We find results supporting the high skill hypotheses using industry variation.

[40] We thank Bryan Seegmiller for creating the mapping from CPS occupation codes to the Autor and Dorn (2013) occupation codes.

[41] This measure is created from the O*NET tasks measures on the importance of: (1) analyzing data/information, (2) thinking creatively, and (3) interpreting information for others. The index is constructed to be mean zero and have unit variance.

[42] We measure average years of completed education and average earnings in the years 1985-1989.

Table 3: Non-routine cognitve skills and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Non-Routine Cognitive Skills	0.00538***	0.00583***	0.0646***	-0.0289***
	(0.00127)	(0.00123)	(0.00762)	(0.00844)
Round N (Occupations)	300	300	300	300
R-squared	0.079	0.114	0.179	0.054

Note: Table presents parameter results of estimating equation (18)*, where the independent variable is the degree of non-routine cognitive skills in an occupation as measured by Acemoglu and Autor (2011). Non-routine cognitive skills are normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses.* ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

The parameter of interest is η which measures whether occupations with a greater task requirement X experienced a larger increase in parameter Y over the sample period. Hence, if $\eta > 0$ then an occupation with a greater amount of skill content X experienced a larger change in earnings risk over the sample period.

Table 3 presents the results of estimating equation 18, where the measure of skill intensity (X_o) is the degree of non-routine cognitive skills. In column (1) of Table 3, the dependent variable is the change in combined persistent income risk. The positive coefficient on non-routine cognitive skills in column (1) indicates that occupations with a larger degree of non-routine cognitive skill content have seen a larger increase in combined persistent income risk over our sample period. The coefficient indicates that an occupation one standard deviation (SD) above the mean level of non-routine cognitive skill has seen combined persistent income risk increase by over 1 percentage point more relative to an occupation one SD below the mean. In columns (2) and (3) of Table 3, we show that we obtain similar results for the change the variance of shocks to persistent earnings among the employed (column (2)) and the unemployed (column (3)).

In column (4) of Table 3, we present the results of estimating equation 18 where the dependent variable is the change in the drift to persistent earnings during the first year of unemployment. The negative coefficient in column (4) indicates that occupations with a greater degree of non-routine cognitive skill content have seen larger declines in persistent earnings from entering unemployment over the sample period. In particular, an occupation 1-SD above the mean has seen the decline in persistent earnings from entering unemployment become nearly 6 percentage points larger over time relative to an occupation 1-SD below the mean.

In Appendix C.12.2, we show that we obtain similar results using mean years of education and mean earnings as our measure of the high skill nature of an occupation. We also find similar results using mean earnings and education in an individual's industry (Appendix C.13).

Mechanism: Exposure to new technologies. Finally, we examine a potential mechanism for why high skill workers experience a larger increase in persistent earnings risk. Previous work has shown that high skill workers are more exposed to the introduction of new, skill-biased technologies (e.g., Krueger (1993) and Deming and Noray (2020)).[43] New technologies allow workers to increase their output, but they also require workers to have new skills to perform their job. Hence, for workers with the sufficient skill to use the new technology their output increases, which increases their wages. For workers who do not have the skills to use the new technology, the demand for their services declines in their original occupation (or the worker has to move to another occupation where their skills are still employable, but wages are often lower), lowering their wages.[44]

We test this mechanism by exploiting variation in the introduction of new technologies across occupations. In particular, we measure changes in persistent earnings risk among occupations that adopted greater computer and software skill requirements by the 2010s. Since computers were not prevalent in the workplace during the 1980s, individuals in these occupations faced a greater degree of new technology introduction.[45] Using the estimates from Braxton and Taska (2023), who leverage the detailed skill requirements from online vacancies in the Burning Glass (Lightcast) database, we measure the degree of computer use in an occupation by measuring the share of vacancies that list a computer or software requirement in the years 2007 to 2017.[46] To facilitate comparison to our other measures of occupation skill content,

[43]Krueger (1993) provides evidence that the computer revolution of the workplace was more pronounced for high skill workers. Recently, Deming and Noray (2020) showed that there are greater changes in the skill requirements (a proxy for technologies used by firms) of jobs over time for workers with technology intensive college majors, e.g. science, technology, and business.

[44]Consistent with this mechanism Braxton and Taska (2023) shows that workers displaced from occupations that have experienced a greater increase in computer and software requirements suffer larger earnings losses. Additionally, Kogan et al. (2020) find that within industry increases in the rate of innovation are associated with substantial increases in earnings risk, particularly for high income workers.

[45]From Card and DiNardo (2002), "... many observers date the beginning of the *computer revolution* to the introduction of the IBM-PC in 1981..."

[46]Burning Glass (Lightcast) Technologies is a data analytics firm which uses web scraping algorithms to identify newly posted vacancies and collect all of the information included in the vacancy. A central feature of their database is that they collect the detailed skill requirements included in each vacancy along with occupation. Recent papers using the Burning Glass database include Hershbein and Kahn (2018), Deming and Noray (2020), Hazell and Taska (2020), Schubert et al. (2022), Braxton and Taska (2023), among others. We use a measure of computer requirements by occupation based upon the Burning Glass database from Braxton and Taska (2023), who measure the share of vacancies that list a computer and software requirements by occupation and year for the years 2007 and 2010-2017.

Table 4: Computer skills and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Computer Skills	0.00294***	0.00324***	0.0254***	-0.0232***
	(0.00106)	(0.00107)	(0.00892)	(0.00657)
Round N (Occupations)	300	300	300	300
R-squared	0.029	0.043	0.034	0.042

*Note: Table presents parameter results of estimating equation (18), where the independent variable is the exposure to computer and software requirements as measured by Braxton and Taska (2023). Exposure to computer and software requirements is normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$*
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

we normalize the measure of computer skills to have mean zero and standard deviation equal to one.

Table 4 presents the results of estimating equation (18) where the independent variable is an occupation's exposure to computer skills. We find that computer use is a strong predictor of the increase in persistent earnings risk. In particular, Table 4 shows that occupations with a greater amount of computer skill requirements have experienced (i) a larger increase in combined persistent earnings risk (column (1)), as well as a larger increase in persistent income risk among the employed (column (2)) and unemployed (column (3)), and (ii) larger persistent earnings losses while unemployed (column (4)). The magnitudes are sizable. Consider persistent risk among the employed (column (2)). The differential increase in persistent earnings risk between an occupation one SD above versus one SD below mean computer skills is 0.00648 ($= 2 \times 0.00324$). This corresponds to roughly 33% ($= 100 \times 0.00648/0.0191$) of the rise in persistent earnings risk over the time period in Panel (a) of Figure 1.[47]

For whom has persistent risk risen? The evidence presented in Section 4 suggests that increasing labor income risk is concentrated among individuals who (1) have a college degree (or more), (2) are most exposed to computer requirements, and (3) work in occupations that require "extensive skill, knowledge, and experience." These findings support our conclusion that rising persistent income risk a "high-skill" phenomenon, and in the next section, we show that rising persistent earnings risk has important implications for inequality, worker welfare and the macroeconomy.

[47] $Q_{E,1980} = 0.08422$, $Q_{E,2015} = 0.1033$ and so $Q_{E,2015} - Q_{E,1980} = 0.0191$.

5 Welfare

In the preceding sections, we showed that employed and combined persistent income risk rose 20% since the mid-1980s, and persistent earnings dispersion and losses among newly unemployed individuals doubled. In this section and corresponding appendix, we integrate our income process into a Bewley-Huggett-Aiyagari model to demonstrate that our findings have important welfare implications. We find that despite falling temporary risk, the rise in persistent earnings risk is large enough to make workers worse off.

Two steps are necessary to integrate our income process into a Bewley-Huggett-Aiyagari model:

1. Posit and estimate a law of motion for employment status that satisfies sequential exogeneity (e.g, Section 5.1).

2. Discretize the income process. Appendix D.4 provides details on how to discretize our income process using Tauchen's method.

5.1 Employment status law of motion

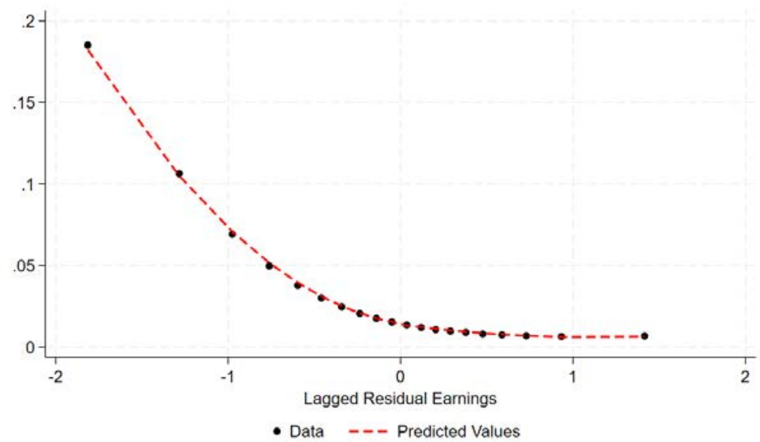
The first step in incorporating our income process into a quantitative model is specifying a law of motion for employment. In this section, we show empirically that the likelihood of becoming unemployed is a function of prior earnings. Let $U_{i,t}$ be an indicator for an individual i being in their first period of unemployment in period t. We model the likelihood that an individual enters into unemployment using the following functional form:

$$U_{i,t} = \mathbb{I}\{y_{t-1} \geq 0\} \left[\sum_{k=0}^{2} \alpha^{+}_{k,E} y^{k}_{t-1} \right] + \mathbb{I}\{y_{t-1} < 0\} \left[\sum_{k=0}^{2} \alpha^{-}_{k,E} y^{k}_{t-1} \right] \tag{19}$$

The functional form in equation (19) allows for probability of entering unemployment to be a quadratic function of prior (residual) earnings (y_{t-1}) estimated separately for positive (residual) earnings or negative prior (residual) earnings.

To gauge the plausibility of the functional form, Figure 8 compares the observed share of unemployed individuals by ventile of prior earnings (black, solid line) to the predicted value based on equation (19) (red, dashed line). The fit is excellent. Finally, we must also define the unemployment probability for the individuals who are currently unemployed. We use a constant unemployment probability for these individuals. This employment process is consistent with the sequential exogeneity assumptions in 1.1.

Figure 8: Probability of entering unemployment



Note: This figure shows the predicted probability of unemployment as estimated from equation (19) *plotted against the actual observed probability of unemployment.*
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

5.2 Welfare implications of rising persistent risk

Finally, we examine the welfare implications of rising persistent income risk. For this exercise, we compare the 2010s and the 1980s using the estimates of our income process presented in Appendix C.6 and the law of motion for employment status outlined above in Section 5.1.[48] Due to space constraints, we present the model environment and estimation details in Appendix D.

Key welfare findings. The increase in persistent earnings risk decreased welfare by 5.1% of lifetime consumption. The decline in temporary risk immaterially offset these losses. On the other hand, the declining rate at which workers enter unemployment offset these losses by 40%. If the inflow into unemployment remained at its 1980s levels, the welfare losses from rising persistent risk would have been 8.5%. We leave it to future research to explore the implications of these findings for optimal tax and transfer systems and other normative questions.

6 Conclusion

For whom has earnings risk changed, and why? By answering these questions our paper makes several contributions. First, we write down a simple persistent-temporary income process that allows for spells of unemployment and show how it can be estimated via the Kalman filter

[48] As we discuss in Section 3.2, this income process allows the means and variances of shocks to persistent and temporary earnings to vary by age and decade.

and EM algorithm. The EM algorithm makes estimating the parameters of the income process highly tractable even in settings where parameters depend upon a large number of observables (e.g., age, education, detailed occupation codes, etc.).

Second, we estimate our income process on administrative earnings records that have been linked to survey responses from the CPS ASEC to examine how persistent and temporary earnings risk have changed over time. From our estimates of persistent earnings risk among the employed and unemployed, we create a novel measure of "combined persistent risk," which takes unemployment risk into account. We showed that employed and combined persistent income risk rose 20% since the mid-1980s, and persistent earnings dispersion and losses among newly unemployed individuals doubled. These increases in risk imply sizable welfare losses in Bewley-Huggett-Aiyagari models of incomplete markets.

Third, we examine why persistent earnings risk has increased. We show that the increase in persistent earnings risk has been largest among: (1) individuals with a college degree or higher, (2) individuals who work in occupations that require "extensive skill, knowledge, and experience," and (3) occupations with a larger degree of non-routine cognitive task content as measured by Acemoglu and Autor (2011). We argue that the increase in persistent earnings risk among the high skill is due to their greater exposure to new, skill-biased technologies. We show that workers employed in occupations which have introduced computer and software requirements the most intensively have seen the largest increase in persistent earnings risk.

We view this paper as part of a broader research agenda that aims to open the black box of earnings dynamics. By recovering shocks to persistent and temporary earnings for *every* person in *every* period, the method presented in this paper can be used to further understand the factors that shape earnings at the individual level as well as how individuals respond to temporary and persistent shocks.

References

Acemoglu, D. and D. Autor (2011). Skills, tasks and technologies: Implications for employment and earnings. In *Handbook of labor economics*, Volume 4, pp. 1043–1171. Elsevier.

Akerman, A., I. Gaarder, and M. Mogstad (2015). The skill complementarity of broadband internet. *The Quarterly Journal of Economics 130*(4), 1781–1824.

Altonji, J. G., A. A. Smith Jr, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica 81*(4), 1395–1454.

Alvarez, F. E., K. Borovičková, and R. Shimer (2016). Decomposing duration dependence in a stopping time model. Technical report, National Bureau of Economic Research.

Arellano, M., R. Blundell, and S. Bonhomme (2017). Earnings and consumption dynamics: a nonlinear panel data framework. *Econometrica 85*(3), 693–734.

Arellano, M. and S. Bonhomme (2016). Nonlinear panel data estimation via quantile regressions.

Atack, J., R. A. Margo, and P. W. Rhode (2019). " automation" of manufacturing in the late nineteenth century: The hand and machine labor study. *Journal of Economic Perspectives 33*(2), 51–70.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2018). New technologies and the labor market. *Journal of Monetary Economics 97*, 48–67.

Atalay, E., P. Phongthiengtham, S. Sotelo, and D. Tannenbaum (2020). The evolution of work in the united states. *American Economic Journal: Applied Economics 12*(2), 1–34.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the us labor market. *American economic review 103*(5), 1553–1597.

Autor, D. H., L. F. Katz, and M. S. Kearney (2008). Trends in us wage inequality: Revising the revisionists. *The Review of economics and statistics 90*(2), 300–323.

Autor, D. H., L. F. Katz, and A. B. Krueger (1998). Computing inequality: have computers changed the labor market? *The Quarterly journal of economics 113*(4), 1169–1213.

Autor, D. H., F. Levy, and R. J. Murnane (2003). The skill content of recent technological change: An empirical exploration. *The Quarterly journal of economics 118*(4), 1279–1333.

Barbe, P. and P. Bertail (2012). *The weighted bootstrap*, Volume 98. Springer Science & Business Media.

Bloom, N., F. Guvenen, L. Pistaferri, J. Sabelhaus, S. Salgado, and J. Song (2017). The great micro moderation.

Blundell, R., L. Pistaferri, and I. Preston (2008). Consumption inequality and partial insurance. *The American Economic Review*, 1887–1921.

Borella, M., M. De Nardi, and F. Yang (2019). Are marriage-related taxes and social security benefits holding back female labor supply? Technical report, National Bureau of Economic Research.

Braxton, J. C., K. F. Herkenhoff, J. L. Rothbaum, and L. Schmidt (2021). Changing income risk across the us skill distribution: Evidence from a generalized kalman filter. Technical report, National Bureau of Economic Research.

Braxton, J. C. and B. Taska (2023). Technological change and the consequences of job loss. *American Economic Review 113*(2), 279–316.

Browning, M., M. Ejrnaes, and J. Alvarez (2010). Modelling income processes with lots of heterogeneity. *The Review of Economic Studies 77*(4), 1353–1381.

Card, D. and J. E. DiNardo (2002). Skill-biased technological change and rising wage inequality: Some problems and puzzles. *Journal of labor economics 20*(4), 733–783.

Chari, V. V. and H. Hopenhayn (1991). Vintage human capital, growth, and the diffusion of new technology. *Journal of political Economy 99*(6), 1142–1165.

Chatterjee, A., J. Morley, and A. Singh (2021). Estimating household consumption insurance. *Journal of Applied Econometrics* (forthcoming).

Davis, S. J. and T. von Wachter (2011). Recessions and the costs of job loss. *Brookings Papers on Economic Activity 2011*(2), 1–72.

De Nardi, M., G. Fella, and G. Paz-Pardo (2020). Nonlinear household earnings dynamics, self-insurance, and welfare. *Journal of the European Economic Association 18*(2), 890–926.

DeBacker, J., B. Heim, V. Panousi, S. Ramnath, and I. Vidangos (2013). Rising inequality: transitory or persistent? new evidence from a panel of us tax returns. *Brookings Papers on Economic Activity 2013*(1), 67–142.

Deming, D. J. and K. Noray (2020). Earnings dynamics, changing job skills, and stem careers. *The Quarterly Journal of Economics 135*(4), 1965–2005.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society: Series B (Methodological) 39*(1), 1–22.

Doepke, M. and M. Tertilt (2016). Families in macroeconomics. In *Handbook of macroeconomics*, Volume 2, pp. 1789–1891. Elsevier.

Feigenbaum, J. and D. P. Gross (2020). Automation and the fate of young workers: Evidence from telephone operation in the early 20th century. Technical report, National Bureau of Economic Research.

Fort, T. C. and S. D. Klimek (2016). The effects of industry classification changes on us employment composition.

Fujita, S. (2018). Declining labor turnover and turbulence. *Journal of Monetary Economics 99*, 1–19.

Geweke, J. and M. Keane (2000). An empirical analysis of earnings dynamics among men in the psid: 1968–1989. *Journal of econometrics 96*(2), 293–356.

Goldin, C. and L. F. Katz (2010). *The race between education and technology*. harvard university press.

Gottschalk, P. and R. Moffitt (1994). The growth of earnings instability in the us labor market. *Brookings Papers on Economic Activity 25*(2), 217–272.

Guvenen, F., G. Kaplan, and J. Song (2020). The glass ceiling and the paper floor: Gender differences among top earners, 1981–2012. In *NBER Macroeconomics Annual 2020, volume 35*. University of Chicago Press.

Guvenen, F., G. Kaplan, J. Song, and J. Weidner (2022). Lifetime earnings in the united states over six decades. *American Economic Journal: Applied Economics*.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021). What do data on millions of us workers reveal about lifecycle earnings dynamics? *Econometrica 89*(5), 2303–2339.

Guvenen, F., S. Ozkan, and J. Song (2014). The nature of countercyclical income risk. *Journal of Political Economy 122*(3), 621–660.

Halvorsen, E., H. A. Holter, S. Ozkan, and K. Storesletten (2023). Dissecting idiosyncratic earnings risk. *Journal of the European Economic Association*, jvad047.

Hamilton, J. (1994a). *Time series analysis*. Princeton University Press, Princeton, NJ.

Hamilton, J. D. (1994b). State-space models. *Handbook of econometrics 4*, 3039–3080.

Hazell, J. and B. Taska (2020). Downward rigidity in the wage for new hires. *Available at SSRN 3728939*.

Heathcote, J., F. Perri, and G. L. Violante (2020). The rise of us earnings inequality: Does the cycle drive the trend? *Review of economic dynamics 37*, S181–S204.

Heathcote, J., K. Storesletten, and G. L. Violante (2017). Optimal tax progressivity: An analytical framework. *The Quarterly Journal of Economics 132*(4), 1693–1754.

Hershbein, B. and L. B. Kahn (2018). Do recessions accelerate routine-biased technological change? evidence from vacancy postings. *American Economic Review 108*(7), 1737–72.

Jarmin, R. and J. Miranda (2002). The longitudinal business database. Technical report, US Census Bureau, Center for Economic Studies.

Jensen, S. T. and S. H. Shore (2011). Semiparametric bayesian modeling of income volatility heterogeneity. *Journal of the American Statistical Association 106*(496), 1280–1290.

Juhn, C., K. M. Murphy, and B. Pierce (1993). Wage inequality and the rise in returns to skill. *Journal of political Economy 101*(3), 410–442.

Kaplan, G. and G. Menzio (2016). Shopping externalities and self-fulfilling unemployment fluctuations. *Journal of Political Economy 124*(3), 771–825.

Kaplan, G. and G. L. Violante (2010). How much consumption insurance beyond self-insurance? *American Economic Journal: Macroeconomics 2*(4), 53–87.

Karahan, F. and S. Ozkan (2013). On the persistence of income shocks over the life cycle: Evidence, theory, and implications. *Review of Economic Dynamics 16*(3), 452–476.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2023). Technology and labor displacement: Evidence from linking patents with worker-level data.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and J. Song (2020). Technological innovation and labor income risk. Technical report, National Bureau of Economic Research.

Kopczuk, W., E. Saez, and J. Song (2010). Earnings inequality and mobility in the united states: evidence from social security data since 1937. *The Quarterly Journal of Economics 125*(1), 91–128.

Krueger, A. B. (1993). How computers have changed the wage structure: evidence from microdata, 1984–1989. *The Quarterly Journal of Economics 108*(1), 33–60.

Krusell, P., L. E. Ohanian, J.-V. Ríos-Rull, and G. L. Violante (2000). Capital-skill complementarity and inequality: A macroeconomic analysis. *Econometrica 68*(5), 1029–1053.

Ljungqvist, L. and T. Sargent (1998). The european unemployment dilemma. *Journal of Political Economy 106*(3), 514–550.

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*(1), 1–32.

Meghir, C. and L. Pistaferri (2011). Earnings, consumption and life cycle choices. In *Handbook of labor economics*, Volume 4, pp. 773–854. Elsevier.

Moffitt, R. and S. Zhang (2018). Income volatility and the psid: Past research and new results. In *AEA Papers and Proceedings*, Volume 108, pp. 277–80.

Moffitt, R. A. (2020). Reconciling trends in us male earnings volatility: Results from a four data set project. Technical report, National Bureau of Economic Research.

Nakata, T. and C. Tonetti (2015). Small sample properties of bayesian estimators of labor income processes. *Journal of Applied Economics 18*(1), 121–148.

Sabelhaus, J. and J. Song (2010). The great moderation in micro labor earnings. *Journal of Monetary Economics 57*(4), 391–403.

Schubert, G., A. Stansbury, and B. Taska (2022). Employer concentration and outside options. *Available at SSRN 3599454*.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *American economic review*, 25–49.

Shumway, R. H. and D. S. Stoffer (1982). An approach to time series smoothing and forecasting using the em algorithm. *Journal of time series analysis 3*(4), 253–264.

Song, J., D. J. Price, F. Guvenen, N. Bloom, and T. Von Wachter (2018). Firming up inequality. *The Quarterly Journal of Economics 134*(1), 1–50.

Storesletten, K., C. Telmer, and A. Yaron (2004). Cyclical dynamics in idiosyncratic labor market risk. *Journal of Political Economy 112*(3), 695–717.

Violante, G. L. (2002). Technological acceleration, skill transferability, and the rise in residual inequality. *The Quarterly Journal of Economics 117*(1), 297–338.

Wagner, D. and M. Layne (2014). The Person Identification Validation System (PVS): Applying the Center for Administrative Records and Research and Applications' record linkage software. *U.S. Census Bureau CARRA Report Series #2014-01*.

Ziliak, J. P., C. Hokayem, and C. R. Bollinger (2020). *Trends in Earnings Volatility using Linked Administrative and Survey Data*. US Census Bureau, Center for Economic Studies.

A Additional details: estimation

In this appendix, we provide additional details on our estimation procedure. In Appendix A.1 we discuss how our income process is identified when the persistence parameters (F) is less than one. In Appendix A.2 we present the "Kalman smoother." In Appendix A.3, we show how the income process presented in Section 1.1 can be microfounded from a labor search model. In Appendix A.4 we discuss the role of normally distributed shocks in our estimation procedure. In Appendix A.5, we discuss how we compute standard errors.

A.1 Identification with persistent shocks

In this appendix, we show how the parameters of the income process specified in Section 1.1 can be identified with persistent shocks, i.e., $F < 1$.

First, we discuss how the persistence parameter F can be identified. Observe that $cov(y_t, y_{t-1}) = Fvar(z_{i,t-1})$ and that $cov(y_{t+1}, y_{t-1}) = F^2 var(z_{i,t-1})$. Thus, we can recover F via,

$$F = \frac{cov(y_{t+1}, y_{t-1})}{cov(y_t, y_{t-1})} = \frac{F^2 var(z_{i,t-1})}{Fvar(z_{i,t-1})} \tag{20}$$

With F identified, we can define $\tilde{\Delta} y_{i,t} = y_{i,t} - F y_{i,t-1}$ to be the "quasi-difference" in earnings for an individual i in year t. Using the income process specified in equations 1 and 2, we then have that,

$$\tilde{\Delta} y_{i,t} = B_{i,t}(l_{i,t}) + \nu_{i,t}(l_{i,t}) + \omega_t - F\omega_{t-1}$$

Assuming this individual was employed in both periods (i.e., in $t-1$ and t), then we have,

$$\tilde{\Delta} y_{i,t} = B_E + \nu^E_{i,t} + \omega_t - F\omega_{t-1} \tag{21}$$

where $\nu^E_{i,t}$ denotes the shock to persistent earnings drawn from employed distribution of persistent shocks. Taking the mean and variance of the expression in equation 21 returns,

$$\text{mean}(\tilde{\Delta} y_{i,t}) = B_E \tag{22}$$

$$\text{var}(\tilde{\Delta} y_{i,t}) = Q_E + (1 + F^2)R. \tag{23}$$

Next, consider the second "quasi-difference",

$$\begin{aligned}\tilde{\Delta}^2 y_{i,t} &= y_{i,t} - F^2 y_{i,t-2} \\ &= FB_{i,t-1}(l_{i,t-1}) + F\nu_{i,t-1}(l_{i,t-1}) + B_{i,t}(l_{i,t}) + \nu_{i,t}(l_{i,t}) + \omega_{i,t} - F^2\omega_{i,t-2}.\end{aligned}$$

The individual i is employed in periods t and $t-2$ since we observe their income. If they were employed in $t-1$ then their change in earnings can be written as,

$$\tilde{\Delta}^{2,E} y_{i,t} = FB_E + F\nu^E_{i,t-1} + B_E + \nu^E_{i,t} + \omega_{i,t} - F^2\omega_{i,t-2}$$

Taking the variance of $\tilde{\Delta}^{2,E} y_{i,t}$ returns

$$\text{var}(\tilde{\Delta}^{2,E} y_{i,t}) = (1+F^2)Q_E + (1+F^4)R, \tag{24}$$

Alternatively, if the individual i is unemployed in period $t-1$, then we can write their change in earnings as,

$$\tilde{\Delta}^{2,U} y_{i,t} = FB_U + F\nu^U_{i,t-1} + B_E + \nu^E_{i,t} + \omega_{i,t} - F^2\omega_{i,t-2} \tag{25}$$

where $\nu^U_{i,t-1}$ denotes the shock to persistent earnings drawn from the unemployed distribution of persistent shocks. Taking the mean and the variance of equation 25 returns,

$$\text{mean}(\tilde{\Delta}^{2,U} y_{i,t}) = FB_U + B_E, \tag{26}$$

$$\text{var}(\tilde{\Delta}^{2,U} y_{i,t}) = F^2 Q_U + Q_E + (1+F^4)R. \tag{27}$$

Finally, we have observed initial income observations:

$$y_{i,1} = z_{i,1} + \omega_{i,1} = Fz_{i,0} + B_E + \nu^E_{i,1} + \omega_{i,1}, \tag{28}$$

Taking the variance of equation 28, we have:

$$\text{var}(y_{i,1}) = F^2 Q_0 + Q_E + R. \tag{29}$$

The income process specified in Section 1.1 contains 7 parameters $(Q_E, Q_U, Q_0, R, B_E, B_U, F)$. The structure of the income process and these parameters then make predictions about the mean and variance of (quasi) earnings changes at different horizons and for different employment statuses, which are summarized by the following 7 equations: (20), (22), (23), (24), (26), (27), and (29).

A.2 Kalman Smoother

As discussed in Section 1.3.1 after running the Kalman filter, we run the Kalman smoother to update our estimates of the unobserved state variable (i.e., persistent earnings and its lag.) In this appendix, we present the algorithm for running the Kalman smoother.

The steps for the smoothed Kalman filter are:

1. Run the Kalman filter as presented in Section 1.3.1 storing the sequences $\{M_{i,t|t-1}\}_{t=1}^{T}$ and $\{M_{i,t|t}\}_{t=1}^{T}$ as well as $\{\hat{\zeta}_{i,t|t-1}\}_{t=1}^{T}$ and $\{\hat{\zeta}_{i,t|t}\}_{t=1}^{T}$.

2. Store the element $\hat{\zeta}_{i,T|T}$ from $\{\hat{\zeta}_{i,t|t}\}_{t=1}^{T}$.

3. Calculate the sequence of smoothed estimations $\{\hat{\zeta}_{i,t|T}\}_{t=1}^{T-1}$ in reverse order by iterating on:

$$\hat{\zeta}_{i,t|T} = \hat{\zeta}_{i,t|t} + J_{i,t}(\hat{\zeta}_{i,t+1|T} - \hat{\zeta}_{i,t+1|t})$$

   for $t = T-1, T-2, ..., 1$, where $J_{i,t} = M_{i,t|t}\hat{F}'M_{i,t+1|t}^{-1}$.

4. Update the sequence of MSE by iterating on:

$$M_{i,t|T} = M_{i,t|t} + J_{i,t}(M_{i,t+1|T} - M_{i,t+1|t})J_{i,t}'.$$

A.3 Labor search model

In this appendix, we show how the income process presented in Section 1.1 can be obtained from a random search model with generalized Nash Bargaining, as in Kaplan and Menzio (2016). There is a unit measure of individuals who live indefinitely. Time is discrete and runs forever. Let $e \in \{W, U\}$ denote employment status. We assume that persistent and temporary human capital evolve as follows:

$$\ln h^{p'} = \rho \ln h^p + h_e + \epsilon, \qquad \epsilon \sim N(0, \sigma^p)$$

$$\ln h^t = u, \qquad u \sim N(0, \sigma^t)$$

where $h_e > 0$ for $e = W$ and $h_u < 0$ for $e = U$ capture on-the-job learning while employed and human capital depreciation while unemployed, respectively.

We assume the production function is multiplicative in human capital, i.e., $f(h^p, h^t) = h^p h^t$. We assume the outside option is a γ replacement rate and is thus $\gamma h^p h^t$. Let ω denote the bargaining weight of the workers. We assume generalized Nash Bargaining in which the outside options of the worker and firms are to not produce but remain matched and renegotiate next period (see Kaplan and Menzio (2016)).[49] Worker income is therefore given by,

$$w = h^p h^t (\omega + \gamma(1-\omega)).$$

Thus log income maps exactly into our framework,

$$\ln w = \ln h^p + \ln h^t + \ln(\omega + \gamma(1-\omega)).$$

The timing of events from the start to end of a period is: (1) workers produce and consume, (2) unemployed workers search for a job/employed workers are laid off at rate δ (which can depend on the income w), and lastly (3) at the end of the period, human capital transitions are realized; thus, contemporaneous shocks to temporary and persistent human capital do not determine employment status. In terms of our econometric model, this timing assumption is equivalent to *sequential exogeneity* discussed in Section 1.1.

Formally, an unemployed worker continuation value is given by

$$U(h^p, h^t) = \gamma h^p h^t + \beta\{p(\theta) E_{h^{p'}, h^{t'}|h^p, W} W(h^{p'}, h^{t'}) + (1 - p(\theta)) E_{h^{p'}, h^{t'}|h^p, U} U(h^{p'}, h^{t'})\}.$$

An employed worker's continuation value is similarly defined, where we assume there is an job-loss probability $\delta(w)$ each period,

$$W(h^p, h^t) = w + \beta\{(1 - \delta(w)) E_{h^{p'}, h^{t'}|h^p, W} W(h^{p'}, h^{t'}) + \delta(w) E_{h^{p'}, h^{t'}|h^p, U} U(h^{p'}, h^{t'})\},$$

$$w = h^p h^t (\omega + \gamma(1-\omega)).$$

Income is given by $\ln w = \ln h^p + \ln h^t + \ln(\omega + \gamma(1-\omega))$. Since our residualization removes

[49] Kaplan and Menzio (2016) assume that in the event of non-agreement, the outside options of the workers and firm are to remain matched and renegotiate next period. The generalized Nash Bargaining objective is

$$\max_w \Big\{ w + \beta(1-\delta(w)) EW(h^{p'}, h^{t'}) + \beta\delta(w) EU(h^{p'}, h^{t'}) - \underbrace{\gamma h^p h^t - \beta(1-\delta(w)) EW(h^{p'}, h^{t'}) - \beta\delta(w) EU(h^{p'}, h^{t'})}_{\text{Keep match, collect UI}} \Big\}^{\omega} \ldots$$

$$\ldots \times \Big\{ h^p h^t - w + \beta(1-\delta(w)) EJ(h^{p'}, h^{t'}) - \underbrace{0 - \beta(1-\delta(w)) EJ(h^{p'}, h^{t'})}_{\text{Keep match, dont produce}} \Big\}^{1-\omega}$$

This yields $w = h^p h^t (\omega + \gamma(1-\omega))$

predictable components of w, this maps exactly to our income process in the text. Also note that the layoff rate can depend on income w while maintaining *sequential exogeneity*. Hence, this fairly standard labor search environment yields an identical income process to the one presented in Section 1.1.

A.4 Role of Normally Distributed Shocks

In this appendix, we discuss the role of distributional assumptions for our analysis: (1) we argue that our estimates do not rely upon normality of the shocks by relying on theoretic results, (2) we verify these theoretic results through Monte Carlo simulations, and (3) we show that our benchmark income process exhibits non-normal distributions of persistent and temporary innovations.

Theoretic justification. In deriving the likelihood function used to estimate parameters and compute posteriors, we assume that the shocks to temporary and persistent earnings (for both the employed and unemployed) are normally distributed; however, each step of our estimation procedure does not require the shocks to be normally distributed.

The key step of the Kalman filter infers the unobserved state variable (persistent earnings) using the linear projection updating formula (detailed in Chapter 4.5 and Chapter 13.2 of Hamilton (1994a)). The linear projection updating formula infers the unobserved state variable by minimizing the mean squared error of the forecast. As long as the underlying state-space model is linear, the linear projection updating formula yields unbiased estimates of the unobserved state under non-normality.[50]

Additionally, the EM algorithm produces consistent estimates of the income process parameters even in cases when the shocks are not normally distributed.[51] The intuition for the EM result is that the formulas to update the parameters of the income process resemble GLS-style regression formulas. Hence, the Gauss-Markov theorem applies, and we obtain the best linearly unbiased estimator (BLUE) for the parameters.

Monte Carlo evidence. In Appendix B.6, we verify these theoretic results in small samples via Monte Carlo simulation exercises. We simulate non-normal innovations to equations (1)

[50]See the longer discussion in the handbook chapter Hamilton (1994b). In particular, in Section 2.5 Hamilton (1994b) writes, "Thus, while the Kalman filter forecasts need no longer be optimal for systems that are not normal, no other forecast based on a linear function of [z_t] will have a smaller mean squared error [see Anderson and Moore (1979, pp. 92-98) or Hamilton (1994, Section 13.2)]. These results parallel the Gauss-Markov theorem for ordinary least squares regression."

[51]See Chapter 13 of Hamilton (1994a). In particular, on p.389 of Hamilton (1994a) entitled "Quasi-Maximum Likelihood Estimation," he writes, that even with non-normal shocks, the likelihood function can be interpreted as a quasi-maximum likelihood function and estimation "will still yield consistent estimates of the elements of F, Q, A, H, and R".

and (2) for $N = 2500$ individuals and $T = 30$ years (comparable to samples from the PSID). We use Guvenen et al. (2021)'s estimated mixture distribution of innovations to both persistent and temporary earnings (parameters presented in their Table 4, column (3)). We then apply our filtering methods to the simulated data. We regress the true persistent earnings (z_{it}) on the recovered estimate of persistent earnings ($\hat{z}_{it}$) and we report the mean and standard deviation of that coefficient in order to assess the algorithm's performance. Our method produces an extremely good fit of the true latent states, with a bias of less than 0.1%. We vary time horizons and find similar results.

Non-normal shocks in estimated process. In Appendix D.5, we show that our estimated income process yields higher order moments that are consistent with Guvenen et al. (2021). In our estimates and Guvenen et al. (2021)'s, the standard deviation and skewness of earnings are negatively correlated with an individual's lagged ranking in the earnings distribution. Moreover, kurtosis is positively correlated with an individual's lagged ranking in the earnings distribution. Thus our simple income process yields non-degenerate higher-order moments that mirror the data, and so we contribute a tractable income process that allows researchers to incorporate rich earnings dynamics into theoretic frameworks.

Non-normality of recovered shocks. Recent work has emphasized that log income changes are non-Gaussian, and exhibit negative skewness as well as excess kurtosis (e.g., Guvenen et al. (2021)). While the shocks to temporary and persistent earnings in our income process are drawn from normal distributions, our income process produces skewness and kurtosis in log earnings changes by incorporating unemployment spells as well as making the shocks functions of other observables. By conditioning on these observables, we naturally estimate mixture distributions; therefore, integrating out these observables yields non-Gaussian shock distributions even if shocks were Gaussian conditional on $l_{i,t}$ and $x_{i,t}$. In an earlier version of this paper (Braxton et al. (2021), Figure 1), we showed that our estimates of persistent and temporary earnings shocks exhibit negative skewness as well as excess kurtosis relative to a normal distribution. These deviations became especially stark when we incorporate observables such as job switching into the estimation.

A.5 Computing standard errors

In this appendix, we discuss how we compute standard errors of our model parameters. We obtain standard errors on our parameter estimates using a block bootstrap procedure. In the bootstrap procedure, we draw a 5% random sample and run the filtering algorithm on this sub-sample of the data. To obtain greater variation, we also randomly draw weights from an exponential distribution and rescale the survey weights by multiplying the survey weights by these draws (e.g., Barbe and Bertail (2012)). We repeat this exercise 100 times. We obtain standard errors by taking the standard deviation of the parameter estimates across the 100 replications and multiply by the square root of the sampling probability (5%).

B EM Algorithm

In this appendix, we outline the EM algorithm we use to estimate the parameters of the income process presented in Section 1. In Appendix B.1, we give an overview of the EM algorithm. In Appendix B.2, we present the full-information log likelihood. In Appendix B.3, we derive the expressions for updating the mean (drift) parameters (B) and the persistence parameter F. In Appendix B.4, we drive the expression for updating the variance parameters (Q and R). In Appendix B.5, we write out the full EM algorithm. In Appendix B.6 we present a series of Monte-Carlo exercises to validate that our estimation procedure is able to accurately recover the path of persistent earnings at the individual level as well as the parameters of the income process. Finally, in Appendix B.7 we show that the potential for misclassifying an individual's employment status is rare and has minimal impact on our results in simulations.

B.1 Overview of EM Algorithm

The EM algorithm is an iterative algorithm to update the parameters that govern the income process. To start the algorithm we make an initial guess of the parameters of the income process, and using these parameters create an estimate of the state vector using the Kalman filter presented in Section 1.3.1. The next step in the EM algorithm is to use the estimates of the state vector along with the data to update the estimates of the parameters. The parameters are updated using a series of equations that we drive below. The algorithm then repeats by using the new parameters to update the estimate of the state vector, and then using the estimated state vector and data to update the parameters. This process continues until the log likelihood has been maximized. In the subsections below we derive the equations that will allow for closed form updating of the parameters. Finally, we provide a detailed description of the EM algorithm.

B.2 Log Likelihood

The EM algorithm uses a set of closed form updating equations to uncover parameters which allow the log likelihood function to be maximized. To derive these formulas we start with the full-information conditional log likelihood, which is the likelihood function *if* the state-variables are observed. For an individual i, the full information log likelihood appears as:[52]

$$
\begin{aligned}
LL_i(\{y_{i,t}\}_{t=0}^{T}, \{z_{i,t}\}_{t=0}^{T} \mid \{l_{i,t}, x_{i,t}\}_{t=1}^{T}, \theta_0) = &-\frac{T+1}{2}\log(2\pi) \\
&-\frac{1}{2}\log(Q_0) - \frac{1}{2}\frac{(z_{i0})^2}{Q_0} \\
&-\frac{1}{2}\sum_{t=1}^{T}\log(Q(l_{i,t})) - \frac{1}{2}\sum_{t=1}^{T}\frac{(z_{i,t} - Fz_{i,t-1} - B(l_{i,t}))^2}{Q(l_{i,t})} \\
&-\frac{1}{2}\sum_{t=1}^{T}\log(R(l_{i,t})) - \frac{1}{2}\sum_{t=1}^{T}\frac{(l_{E,i,t})(y_{i,t} - z_{i,t})^2}{R(l_{i,t})}
\end{aligned}
$$

B.3 Updating Means

In this appendix, we derive the expressions that are used to update the mean parameters of the income process (e.g. the persistence of persistent earnings, and drifts of persistent earnings when employed/unemployed). Before deriving the formulas we present a series of useful expressions that will ease the derivations of the updating equations. Additionally, note the following notation. Define $E_T\left[z_{it}|\{x_{it}, y_{it}, l_{it}\}\right] = \hat{z}_{it|T}$, that is the expected value of individual i's persistent earnings in period t (given the data) is denoted by $\hat{z}_{it|T}$, which corresponds to the output of the smoothed Kalman filter. Define $\Sigma_{i0|T}(1,1)$ to be the estimated the variance of initial persistent earnings. Define $\Sigma_{it|T}(1,2)$ to be the estimated covariance between $\hat{z}_{it|T}$ and $\hat{z}_{it-1|T}$.[53]

For simplicity and ease of notation, we will first discuss in detail how to update parameters for the simple income process outlined in section 1.1. Then, we discuss how things extend to the more general case in which F and B are both assumed to be linear in a set of unknown parameters in Section B.3.3.

B.3.1 Useful Expressions

In this section, we derive a series of useful expressions that will aid in the derivation of the updating equations in the following subsections.

[52] Note the full information log likelihood is used to derive the equations which update the parameters of the income process via the EM algorithm. The likelihood that is maximized as part of the estimation is given by (7).

[53] Note this covariance term is the $(1,2)$ element of the matrix $M_{i,t|T}$.

First, we show that $E_T\left[z_{i0}^2|\{x_{it}, y_{it}, l_{it}\}\right] = \Sigma_{i0|T}(1,1) + \hat{z}_{i0|T}^2$,

$$\begin{aligned}
E_T\left[z_{i0}^2|\{y_{it}, x_{it}, l_{it}\}\right] &= E_T\left[\left(z_{i0} - \hat{z}_{i0|T} + \hat{z}_{i0|T}\right)^2 |\{y_{it}, x_{it}, l_{it}\}\right] \\
&= E_T\left[\left(z_{i0} - \hat{z}_{i0|T}\right)^2 + \hat{z}_{i0|T}^2 + 2\left(z_{i0} - \hat{z}_{i0|T}\right)\hat{z}_{i0|T}|\{y_{it}, x_{it}, l_{it}\}\right] \\
&= E_T\left[\left(z_{i0} - \hat{z}_{i0|T}\right)^2 + \hat{z}_{i0|T}^2|\{y_{it}, x_{it}, l_{it}\}\right] \\
&= \Sigma_{i0|T}(1,1) + \hat{z}_{i0|T}^2 \qquad (30)
\end{aligned}$$

where in the third equality we used the fact that $E_T\left[z_{i0} - \hat{z}_{i0|T}|\{y_{it}, x_{it}, l_{it}\}\right] = 0$.

Next, we show that $E_T\left[z_{it}z_{i,t-1}|\{y_{it}, x_{it}, l_{it}\}\right] = \Sigma_{it|T}(1,2) + \hat{z}_{it|T}\hat{z}_{it-1|T}$,

$$\begin{aligned}
E_T\left[z_{it}z_{i,t-1}|\{y_{it}, x_{it}, l_{it}\}\right] &= E_T\left[\left(z_{it} - \hat{z}_{it|T} + \hat{z}_{it|T}\right)\left(z_{i,t-1} - \hat{z}_{it-1|T} + \hat{z}_{it-1|T}\right)|\{y_{it}, x_{it}, l_{it}\}\right] \\
&= E_T\left[\left(z_{it} - \hat{z}_{it|T}\right)\left(z_{i,t-1} - \hat{z}_{it-1|T}\right) + \hat{z}_{it|T}\hat{z}_{it-1|T}|\{y_{it}, x_{it}, l_{it}\}\right] \\
&= \Sigma_{it|T}(1,2) + \hat{z}_{it|T}\hat{z}_{it-1|T} \qquad (31)
\end{aligned}$$

where in the second equality we have used the fact that $E_T\left[\left(z_{it} - \hat{z}_{it|T}\right)|\{y_{it}, x_{it}, l_{it}\}\right] = 0$ and $E_T\left[\left(z_{it-1} - \hat{z}_{it-1|T}\right)|\{y_{it}, x_{it}, l_{it}\}\right] = 0$.

B.3.2 Updating F, B_E, B_U

In this section, we derive the expression we will use to update the parameters $\{F, B_E, B_U\}$. The relevant part of the log likelihood for updating the parameters $\{F, B_E, B_U\}$ is given by:

$$\frac{1}{Q(l_{i,t})}\sum_{t=1}^{T}\left(z_{i,t} - Fz_{i,t-1} - B(l_{i,t})\right)^2$$

The expected value can be written as:

$$\frac{1}{Q(l_{i,t})}E_T\left[\left(z_{i,t} - Fz_{i,t-1} - B(l_{i,t})\right)^2|\{y_{it}, x_{it}, l_{it}\}\right]$$

Completing the square we obtain the following expression:

$$\begin{aligned}\frac{1}{Q(l_{i,t})} E_T \Big[z_{i,t}^2 - z_{i,t} F z_{i,t-1} - z_{i,t} B(l_{i,t}) \\ + F^2 z_{i,t-1}^2 - F z_{i,t-1} z_{i,t} + F z_{i,t-1} B(l_{i,t}) \\ + B(l_{i,t})^2 - B(l_{i,t}) z_{i,t} + F B(l_{i,t}) z_{i,t-1} | \{y_{it}, x_{it}, l_{it}\} \Big]\end{aligned}$$

Combining terms we have:

$$\begin{aligned}\frac{1}{Q(l_{i,t})} E_T \Big[z_{i,t}^2 - 2F z_{i,t} z_{i,t-1} - 2 z_{i,t} B(l_{i,t}) \\ + F^2 z_{i,t-1}^2 + 2F z_{i,t-1} B(l_{i,t}) \\ + B(l_{i,t})^2 | \{y_{it}, x_{it}, l_{it}\} \Big]\end{aligned} \tag{32}$$

We will next use expressions from Section B.3.1 to simplify equation (32). First using equation (30) (adjusted for period t, and period $t+1$), we have:

$$\begin{aligned}\frac{1}{Q(l_{i,t})} \left(\Sigma_{it|T}(1,1) + \hat{z}_{it|T}^2 + F^2 \left[\Sigma_{it-1|T}(1,1) + \hat{z}_{it-1|T}^2 \right] \right) \\ + \frac{1}{Q(l_{i,t})} E_T \big[-2F z_{i,t} z_{i,t-1} - 2 z_{i,t} B(l_{i,t}) + 2F z_{i,t-1} B(l_{i,t}) \\ + B(l_{i,t})^2 | \{y_{it}, x_{it}, l_{it}\} \big]\end{aligned}$$

Next using equation (31), we have:

$$\begin{aligned}\frac{1}{Q(l_{i,t})} \left(\Sigma_{it|T}(1,1) + \hat{z}_{it|T}^2 + F^2 \left[\Sigma_{it-1|T}(1,1) + \hat{z}_{it-1|T}^2 \right] \right) \\ + \frac{1}{Q(l_{i,t})} \left(-2F \left[\Sigma_{it|T}(1,2) + \hat{z}_{it|T} \hat{z}_{it-1|T} \right] \right) \\ + \frac{1}{Q(l_{i,t})} E_T \left[-2 z_{i,t} B(l_{i,t}) + 2F z_{i,t-1} B(l_{i,t}) + B(l_{i,t})^2 | \{y_{it}, x_{it}, l_{it}\} \right]\end{aligned}$$

Then taking the expectation over the remaining terms we have:

$$\frac{1}{Q(l_{i,t})}\left(\Sigma_{it|T}(1,1)+\hat{z}^2_{it|T}+F^2\left[\Sigma_{it-1|T}(1,1)+\hat{z}^2_{it-1|T}\right]\right) \tag{33}$$
$$\frac{1}{Q(l_{i,t})}\left(-2F\left[\Sigma_{it|T}(1,2)+\hat{z}_{it|T}\hat{z}_{it-1|T}\right]\right)$$
$$+\frac{1}{Q(l_{i,t})}\left[\left(-2\hat{z}_{it|T}B(l_{i,t})+2F\hat{z}_{it-1|T}B(l_{i,t})+B(l_{i,t})^2\right)\right]$$

We want to optimize equation (33) with respect to F, B_E and B_U. For ease of exposition, we drop the terms in equation (33) that do not include F, B_E and B_U, which returns:

$$\frac{1}{Q(l_{i,t})}\left(F^2\left[\Sigma_{it-1|T}(1,1)+\hat{z}^2_{it-1|T}\right]\right) \tag{34}$$
$$\frac{1}{Q(l_{i,t})}\left(-2F\left[\Sigma_{it|T}(1,2)+\hat{z}_{it|T}\hat{z}_{it-1|T}\right]\right)$$
$$+\frac{1}{Q(l_{i,t})}\left[\left(-2\hat{z}_{it|T}B(l_{i,t})+2F\hat{z}_{it-1|T}B(l_{i,t})+B(l_{i,t})^2\right)\right]$$

The expression in (34) gives the expected contribution to the likelihood for individual i in period t. We want to maximize the likelihood across all individuals and time periods. To perform this optimization it will be convenient to define the following vectors and matrices. Define:

$$X_C \equiv \begin{bmatrix} \hat{z}_{1,0|T} & l_{E,1,1} & l_{U,1,1} \\ \hat{z}_{1,1|T} & l_{E,1,2} & l_{U,1,2} \\ \vdots & \vdots & \vdots \\ \hat{z}_{1,T-1|T} & l_{E,1,T} & l_{U,1,T} \\ \hat{z}_{2,0|T} & l_{E,2,1} & l_{U,2,1} \\ \vdots & \vdots & \vdots \\ \hat{z}_{N,T-1|T} & l_{E,N,T} & l_{U,N,T} \end{bmatrix}_{NT\times 3} \qquad C \equiv \begin{bmatrix} F \\ B_E \\ B_U \end{bmatrix}_{3\times 1} \quad Y_C \equiv \begin{bmatrix} \hat{z}_{1,1|T} \\ \vdots \\ \hat{z}_{1,T|T} \\ \vdots \\ \hat{z}_{N,T|T} \end{bmatrix}_{NT\times 1} \tag{35}$$

We can rewrite terms in matrix notation as follows:

$$
\begin{aligned}
C^{'}X_{C}^{'}X_{C}C &= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(F\hat{z}_{it-1|T} + B_{E}l_{Eit} + B_{U}l_{Uit}\right)^{2} \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(F^{2}\hat{z}^{2}_{it-1|T} + 2FB_{E}\hat{z}_{it-1|T}l_{Eit} + 2FB_{U}\hat{z}_{it-1|T}l_{Uit} + (B_{E}l_{Eit})^{2} + (B_{U}l_{Uit})^{2}\right) \\
&= \sum_{i=1}^{N}\sum_{t=1}^{T}\left(F^{2}\hat{z}^{2}_{it-1|T} + 2F\hat{z}_{it-1|T}B(l_{i,t}) + B(l_{i,t})^{2}\right) \qquad \text{using } B(l_{it}) \text{ def. from above.}
\end{aligned}
$$

We also have,

$$
\begin{aligned}
Y_{C}^{'}X_{C}C &= F\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{it|T}\hat{z}_{it-1|T} + B_{E}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{it|T}l_{Eit} + B_{U}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{it|T}l_{Uit} \\
&= F\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{it|T}\hat{z}_{it-1|T} + \sum_{i=1}^{N}\sum_{t=1}^{T}\hat{z}_{it|T}B(l_{i,t}) \qquad \text{using } B(l_{it}) \text{ def. from above.}
\end{aligned}
$$

To complete writing the sum of the log likelihood across individuals it will be convenient to define the following vectors:

$$
\vec{\sigma}_{t-1}(1,1) \equiv \begin{bmatrix} \Sigma_{1,0|T}(1,1) \\ \Sigma_{1,1|T}(1,1) \\ \vdots \\ \\ \Sigma_{N,T-1|T}(1,1) \end{bmatrix}_{NT\times 1} \qquad \vec{\sigma}_{t}(1,2) \equiv \begin{bmatrix} \Sigma_{1,2|T}(1,1) \\ \Sigma_{1,2|T}(1,2) \\ \vdots \\ \\ \Sigma_{N,T|T}(1,2) \end{bmatrix}_{NT\times 1} \qquad e_{1}^{3} \equiv \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \qquad e^{NT} \equiv \underbrace{\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}}_{(NT\times 1)}
$$

We can make further progress with matrix notation by noting,

$$
\sum_{i=1}^{N}\sum_{t=1}^{T}\left(F^{2}\Sigma_{it-1|T}(1,1)\right) = C^{'}e_{1}^{3}e_{1}^{3'}C\vec{\sigma}_{t-1}^{'}(1,1)e^{NT} \tag{36}
$$

and

$$
\sum_{i=1}^{N}\sum_{t=1}^{T}\left(F\Sigma_{it|T}(1,2)\right) = e^{NT'}\vec{\sigma}_{t}(1,2)e_{1}^{3'}C. \tag{37}
$$

Using (34) and the definitions above we have the following:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} E_T\left[(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2\,|\{y_{it},x_{it},l_{it}\}\right] = C^{'}X_C^{'}X_C C - 2Y_C^{'}X_C C$$
$$- 2e^{NT^{'}}\vec{\sigma}_t(1,2)e_1^{3'}C + C^{'}e_1^3 e_1^{3^{'}} C\vec{\sigma}_{t-1}^{'}(1,1)e^{NT}$$

Finally, define:

$$Q^{-1} \equiv \begin{bmatrix} \frac{1}{Q_1(l_{1,1})} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & \frac{1}{Q_2(l_{1,2})} & 0 & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & \vdots & \vdots \\ \vdots & \vdots & 0 & \frac{1}{Q_T(l_{1,T})} & 0 & \vdots & \vdots \\ \vdots & \vdots & \vdots & 0 & \frac{1}{Q_1(l_{2,1})} & 0 & \vdots \\ \vdots & \vdots & \vdots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & \cdots & \cdots & \cdots & 0 & \frac{1}{Q_T(l_{N,T})} \end{bmatrix}$$

Using (34) and the definitions above we have the following:

$$\sum_{i=1}^{N}\sum_{t=1}^{T} \frac{1}{Q(l_{i,t})} E_T\left[(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2\,|\{y_{it},x_{it},l_{it}\}\right] = C^{'}X_C^{'}Q^{-1}X_C C - 2Y_C^{'}Q^{-1}X_C C$$
$$- 2e^{NT^{'}}Q^{-1}\vec{\sigma}_t(1,2)e_1^{3'}C$$
$$+ C^{'}e_1^3 e_1^{3^{'}} C\vec{\sigma}_{t-1}^{'}(1,1)Q^{-1}e^{NT}$$

Taking the FOC with respect to C returns:

$$0 = 2C^{'}X_C^{'}Q^{-1}X_C - 2Y_C^{'}Q^{-1}X_C - 2e^{NT^{'}}Q^{-1}\vec{\sigma}_t(1,2)e_1^{3'} + 2C^{'}e_1^3 e_1^{3^{'}}\vec{\sigma}_{t-1}^{'}(1,1)Q^{-1}e^{NT} \tag{38}$$

Rearranging equation (38) returns:

$$C^{'}\left[X_C^{'}Q^{-1}X_C + e_1^3 e_1^{3^{'}}\vec{\sigma}_{t-1}^{'}(1,1)Q^{-1}e^{NT}\right] = Y_C^{'}Q^{-1}X_C + e^{NT^{'}}Q^{-1}\vec{\sigma}_t(1,2)e_1^{3'} \tag{39}$$

Taking the transpose of both sides of equation (39) returns:

$$\left[X_C^{'}Q^{-1}X_C + e_1^3 e_1^{3^{'}}\vec{\sigma}_{t-1}^{'}(1,1)Q^{-1}e^{NT}\right]C = X_C^{'}Q^{-1}Y_C + e_1^3\vec{\sigma}_t^{'}(1,2)Q^{-1}e^{NT} \tag{40}$$

where we have exploited the fact that the matrices on the LHS of equation (39) are symmetric.

Equation (40) gives us a closed form equation for updating the parameters $\{F, B_E, B_U\}$.

B.3.3 Extension to the general case

Above, we assumed that the income process was quite simple. In particular, one could write $F_{it} = [0, 0, 1]C = e_1^{3'}C = F$ and $B_{it} = [0, l_{E,i,t}, l_{U,i,t}]C$. It turns out that it is straightforward to extend to a much more flexible setting in which

$$F(l_{it}; X_{it}) \equiv f_F(l_{i,t}; x_{i,t}) = g_F(l_{i,t}; x_{i,t})'\Lambda_{B,F} \tag{41}$$

$$B(l_{i,t}; x_{i,t}) \equiv f_B(l_{i,t}; x_{i,t}) = g_B(l_{i,t}; x_{i,t})'\Lambda_{B,F}, \tag{42}$$

where $g_F(l_{i,t}; x_{i,t})$ and $g_B(l_{i,t}; x_{i,t})$ are *known* functions of l_{it} and x_{it}. $\Lambda_{B,F}$ is the set of unknown parameters which captures information which is relevant for the conditional mean in the state equation, which involves both the AR(1) coefficient on lagged persistent income as well as the drift in the state equation. Here, we allow for considerably more flexibility, but simply require that both $F(l_{it}; X_{it})$ and $B(l_{i,t}; x_{i,t})$ are linear in these parameters. In the vast majority of applications, one would tend to expect that things are partitioned so that the j^{th} element of $g_F(l_{i,t}; x_{i,t})$ is always zero if the j^{th} element of $g_B(l_{i,t}; x_{i,t})$ is nonzero with positive probability, but we don't need to require this per se.[54]

Let us define X_F as the design matrix constructed by concatenating the $[g_F(l_{i,t}; x_{i,t})]'$ vectors vertically, and X_B be the analogous object constructed by concatenating the $[g_B(l_{i,t}; x_{i,t})]'$ vectors vertically. Then, let us redefine

$$X_C \equiv \text{diag}(\vec{z}_{t-1})X_F + X_B,$$

where $\vec{z}_{t-1}$ is the first column of the definition of X_C in equation ((35))–i.e., the vector of lagged posterior means. In the special case in which F is constant, $g_F(l_{i,t}; x_{i,t})$ has a 1 in its first element and a zero otherwise, and $g_B(l_{i,t}; x_{i,t})$ has a zero in its first element, we can use this extended X_C matrix in place of the one defined above, and the updating formulas defined in ((40)) apply without modification.

If F_{it} is not constant, we also need to make a minor modification to the additional terms which appear in the likelihood function which involve filtering uncertainty about current and

[54]For example, in our base case above, F was assumed to be the first element of C and the remaining two parameters captured the unknown parameters which governed $B(l_{it})$.

lagged z_{it}. In the more general case, the expressions in equations ((36)-(37)) simplify to

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(F_{it}^2\Sigma_{it-1|T}(1,1)\right)=\sum_{i=1}^{N}\sum_{t=1}^{T}\left(C'g_F(l_{i,t};x_{i,t})\Sigma_{it-1|T}(1,1)[g_F(l_{i,t};x_{i,t})]'C\right)=C'X_F'\text{diag}(\vec{\sigma}_{t-1}(1,1))X_FC$$

and

$$\sum_{i=1}^{N}\sum_{t=1}^{T}\left(F_{it}\Sigma_{it|T}(1,2)\right)=\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\Sigma_{it|T}(1,2)[g_F(l_{i,t};x_{i,t})]'C\right)=\vec{\sigma}_t(1,2)'X_FC.$$

If we use the alternative formulation which allows for F to vary as a linear function of $g_F(l_{i,t};x_{i,t})$, we obtain the closely related expression to equation ((40)):

$$\left[X_C'Q^{-1}X_C+X_F'Q^{-1}\text{diag}(\vec{\sigma}_{t-1}(1,1))X_F\right]C=X_C'Q^{-1}Y_C+X_F'Q^{-1}\vec{\sigma}_t(1,2), \tag{43}$$

which still resembles a GLS regression equation. Clearly, this will not work for completely arbitrary X_F and X_B; we will need to be able to impose restrictions which ensure that the matrix $\left[X_C'Q^{-1}X_C+X_F'Q^{-1}\text{diag}(\vec{\sigma}_{t-1}(1,1))X_F\right]$ is invertible.

B.4 Updating Variances

In this appendix, we derive the expressions that will be used to update the variance parameters. As above, we will economize on notation by restricting attention to the notation of the model in Section 1.1. However, the extension to the general case is immediate. Notice that, below, we already assume that log variances are linear in unknown sets of parameters. As such, allowing for a more flexible linear-in-parameters structure simply requires reinterpreting l_{it} as a broader set of observables than just employment/unemployment dummies.[55]

B.4.1 Shocks to Persistent Earnings When Employed and Unemployed (Q_E and Q_U)

In this section we discuss how we update the variance of persistent earnings for the employed and unemployed. We can write the variance of persistent earnings as:

$$Q(l_{it}) = \exp(l_{it}'\Lambda_Q)$$

where $\Lambda_Q=[\Lambda_{Q,E},\Lambda_{Q,U}]$. The relevant part of the negative log likelihood which depends on Λ_Q is:

[55] Also, in expressions below, we would need to replace F with F_{it} and $B(l_{it})$ with B_{it} where appropriate.

$$\Theta(\Lambda_Q;\beta,\omega) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T}\log(Q(l_{it})) + \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2}{Q(l_{it})} \tag{44}$$

To arrive at an updating formula for the variance of persistent earnings, we will take the conditional expectation using the posterior distribution of the latent states given all of the missing data, and then take FOC with respect to Λ_Q. Taking the conditional expectation using the posterior distribution of latent states (given all of the missing data) returns:

$$\Theta(\Lambda_Q;\beta,F) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T}\log(Q(l_{it})) + \sum_{i=1}^{N}\sum_{t=1}^{T}\frac{E_T\left[(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2|x_{it},y_{it},l_{it}\right]}{Q(l_{it})} \tag{45}$$

Observe that this function is convex in Λ_Q. Therefore, if we take a second order approximation of the objective, we obtain the following:

$$\Theta(\Lambda_Q;\beta,F)-\Theta(\Lambda_{Q,0};\beta,\omega) \equiv (\Lambda_Q-\Lambda_{Q,0})'\nabla\Theta + \tfrac{1}{2}(\Lambda_Q-\Lambda_{Q,0})'\,\nabla^2\Theta(\Lambda_Q-\Lambda_{Q,0}),$$

where the Jacobian matrix is defined as

$$\nabla\Theta(\Lambda_{Q,0};\beta,F) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T}\left[1-\frac{E\left[(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2|x_{it},y_{it},l_{it}\right]}{Q(l_{it})}\right] l_{i,t}$$

and the Hessian matrix $\nabla^2\Theta$ is defined as

$$\nabla^2\Theta(\Lambda_{Q,0};\beta,F) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\frac{E\left[(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2|x_{it},y_{it},l_{it}\right]}{Q(l_{it})}\right] l_{i,t}l_{i,t}'$$

Taking first order conditions, we obtain the familiar expressions for Newton's method:

$$\nabla^2\Theta\Lambda_Q = \nabla^2\Theta\Lambda_{Q,0}-\nabla\Theta \qquad \Longleftrightarrow \qquad \Lambda_Q = \Lambda_{Q,0}-\left[\nabla^2\Theta\right]^{-1}\nabla\Theta, \tag{46}$$

which gives us a simple way of updating Λ_Q.

Implementation Note that we can write the conditional expectation term as:

$$\begin{aligned} E\left[(z_{i,t}-Fz_{i,t-1}-B(l_{i,t}))^2|x_{it},y_{it},l_{it}\right] &= E\left[z_{i,t}-Fz_{i,t-1}-B(l_{i,t})|x_{it},y_{it},l_{it}\right]^2 \\ &\quad + var(z_{i,t}-Fz_{i,t-1}-B(l_{i,t})|x_{it},y_{it},l_{it}) \end{aligned} \tag{47}$$

Let $A_{it} = z_{it} - Fz_{i,t-1}$, then we can write the conditional variance expression as follows:

$$\begin{aligned}
var(z_{i,t} - Fz_{i,t-1} - B(l_{i,t})|x_{it}, y_{it}, l_{it}) &= var(A_{it} - B(l_{i,t})|x_{it}, y_{it}, l_{it}) \\
&= var(A_{it}|x_{it}, y_{it}, l_{it}) + var(B(l_{i,t})|x_{it}, y_{it}, l_{it}) \\
&\quad - 2cov(A_{it}, B(l_{i,t})|x_{it}, y_{it}, l_{it}) \\
&= var(A_{it}|x_{it}, y_{it}, l_{it})
\end{aligned}$$

where in the final equality we are using the fact that we are conditioning on l_{it}. Then using the definition of A_{it}, we have:

$$\begin{aligned}
var(z_{it} - Fz_{it-1} - B(l_{i,t})|x_{it}, y_{it}, l_{it}) &= var(z_{it}|x_{it}, y_{it}, l_{it}) + F^2 var(z_{i,t-1}|x_{it}, y_{it}, l_{it}) \\
&\quad - 2Fcov(z_{it}, z_{i,t-1}|x_{it}, y_{it}, l_{it})
\end{aligned} \tag{48}$$

Combining equations (47) and (48), we have the following expression for the conditional expectations terms.

$$\begin{aligned}
E\left[(z_{it} - Fz_{it-1} - B(l_{i,t}))^2|x_{it}, y_{it}, l_{it}\right] &= E\left[(z_{it} - Fz_{it-1} - B(l_{i,t}))|x_{it}, y_{it}, l_{it}\right]^2 \\
&\quad + var(z_{it}|x_{it}, y_{it}, l_{it}) + F^2 var(z_{i,t-1}|x_{it}, y_{it}, l_{it}) \\
&\quad - 2Fcov(z_{it}, z_{i,t-1}|x_{it}, y_{it}, l_{it})
\end{aligned} \tag{49}$$

Closed form expression for Q. Using equation 49 when $Q(\cdot)$ depends only on labor market status, we can obtain a closed form expression for the variance of persistent shocks to the employed and unemployed as a function of the individual level estimates of persistent earnings ($z_{i,t|T}$), its lag ($z_{i,t-1|T}$), and the variance-covariance matrix ($M_{i,t|T}$).

$$Q_E = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} l_{E,i,t}\left[\left(\hat{z}_{i,t|T} - F\hat{z}_{i,t-1|T} - B_E\right)^2 + [1, -F]M_{i,t|T}[1, -F]'\right]}{\sum_{i=1}^{N}\sum_{t=1}^{T} l_{E,i,t}} \tag{50}$$

$$Q_U = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} l_{U,i,t}\left[\left(\hat{z}_{i,t|T} - F\hat{z}_{i,t-1|T} - B_U\right)^2 + [1, -F]M_{i,t|T}[1, -F]'\right]}{\sum_{i=1}^{N}\sum_{t=1}^{T} l_{U,i,t}}, \tag{51}$$

B.4.2 Updating Variance of Temporary Earnings (R)

In this section we discuss how we update the variance of temporary earnings. We can write the variance of persistent earnings as:

$$R(l_{E,i,t}) = \exp(l_{E,i,t}\Lambda_R)$$

where $l_{E,i,t}$ is a dummy variable denoting whether an individual i is employed in period t.

The relevant part of the negative log likelihood which depends on Λ_R is:

$$\Theta(\Lambda_R) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T} \log(R(l_{E,i,t})) + \sum_{i=1}^{N}\sum_{t=1}^{T} \frac{H(l_{i,t})(y_{i,t}-z_{i,t})^2}{R_{i,t}(l_{E,i,t})} \tag{52}$$

Taking the conditional expectation using the posterior distribution of latent states (given all of the missing data) returns:

$$\Theta(\Lambda_R) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T} \log(R(l_{E,i,t})) + \sum_{i=1}^{N}\sum_{t=1}^{T} H(l_{i,t}) \frac{E\left[(y_{i,t}-z_{it})^2|x_{it},y_{it},l_{it}\right]}{R_{i,t}(l_{E,i,t})} \tag{53}$$

Similar to above, observe that this function is convex in Λ_R. Therefore, if we take a second order approximation of the objective, we obtain the following:

$$\Theta(\Lambda_R) - \Theta(\Lambda_{R,0}) \equiv (\Lambda_R - \Lambda_{R,0})'\nabla\Theta + \tfrac{1}{2}(\Lambda_R - \Lambda_{R,0})'\,\nabla^2\Theta(\Lambda_R - \Lambda_{R,0}),$$

where the Jacobian matrix is defined as

$$\nabla\Theta(\Lambda_{R,0}) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T} \left[1 - \frac{E\left[(y_{i,t}-z_{it})^2|x_{it},y_{it},l_{it}\right]}{R(l_{E,i,t})}\right] l_{E,i,t}$$

and the Hessian matrix $\nabla^2\Theta$ is defined as

$$\nabla^2\Theta(\Lambda_{R,0}) \equiv \sum_{i=1}^{N}\sum_{t=1}^{T} \left[\frac{E\left[(y_{i,t}-z_{it})^2|x_{it},y_{it},l_{it}\right]}{R(l_{E,i,t})}\right] l_{E,i,t}l_{E,i,t}'$$

Taking first order conditions, we obtain the familiar expressions for Newton's method:

$$\nabla^2\Theta\Lambda_R = \nabla^2\Theta\Lambda_{R,0} - \nabla\Theta \qquad \Longleftrightarrow \qquad \Lambda_R = \Lambda_{R,0} - \left[\nabla^2\Theta\right]^{-1}\nabla\Theta, \tag{54}$$

Implementation Note that we can write the conditional expectations term as:

$$E\left[(y_{i,t}-z_{it})^2|x_{it},y_{it},l_{it}\right] = E\left[y_{i,t}-z_{it}|x_{it},y_{it},l_{it}\right]^2 + var(y_{i,t}-z_{it}|x_{it},y_{it},l_{it})$$

Since we condition on $y_{i,t}$, the conditional variance term can be written as:

$$var(y_{i,t} - z_{it}|x_{it}, y_{it}, l_{it}) = var(z_{it}|x_{it}, y_{it}, l_{it})$$

Then we have that:

$$E\left[(y_{i,t} - z_{it})^2|x_{it}, y_{it}, l_{it}\right] = E\left[y_{i,t} - z_{it}|x_{it}, y_{it}, l_{it}\right]^2 + var(z_{it}) \tag{55}$$

Closed form expression for R**.** Using equation 55 when R depends only on labor market status, we can obtain a closed form expression for the variance of temporary shocks as a function of the individual level estimates of persistent earnings ($z_{i,t|T}$) and the variance-covariance matrix ($M_{i,t|T}$).

$$R = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} l_{E,i,t}\left[\left(y_{i,t} - \hat{z}_{i,t|T}\right)^2 + [1,0]M_{i,t|T}[1,0]'\right]}{\sum_{i=1}^{N}\sum_{t=1}^{T} l_{E,i,t}}, \tag{56}$$

B.4.3 Updating Variance of Initial persistent Earnings Draw(Q_0)

In this section we discuss how we update the variance of the initial draw of persistent earnings. We can write the variance of initial persistent earnings as:

$$Q_0 = exp(l^0_{E,i,t}\Lambda_{u_{z0}})$$

where $l^0_{E,i,t}$ is a dummy variable that is equal to 1 if individual i is employed E for the first time in the sample in period t.

The relevant part of the negative log likelihood which depends on $\Lambda_{u_{z0}}$ is:

$$\Theta(\Lambda_{u_{z0}}) \equiv \sum_{i=1}^{N}\log(Q_0) + \sum_{i=1}^{N}\frac{(z_{i0})^2}{Q_0}$$

Taking the conditional expectation using the posterior distribution of latent states (given all of the missing data) returns:

$$\Theta(\Lambda_{u_{z0}}) \equiv \sum_{i=1}^{N}\log(Q_0) + \sum_{i=1}^{N}\frac{E\left[(z_{i0})^2|x_{it}, y_{it}, l_{it}\right]}{Q_0}$$

Similar to above, observe that this function is convex in $\Lambda_{u_{z0}}$. Therefore, if we take a second order approximation of the objective, we obtain the following:

$$\Theta(\Lambda_{u_{z0}}) - \Theta(\Lambda_{u_{z0},0}) \equiv (\Lambda_{u_{z0}} - \Lambda_{u_{z0},0})'\nabla\Theta + \tfrac{1}{2}(\Lambda_{u_{z0}} - \Lambda_{u_{z0},0})'\,\nabla^2\Theta(\Lambda_{u_{z0}} - \Lambda_{u_{z0},0}),$$

where the Jacobian matrix is defined as

$$\nabla\Theta(\Lambda_{u_{z0},0}) \equiv \sum_{i=1}^{N}\left[1 - \frac{E\left[(z_{i0})^2|x_{it}, y_{it}, l_{it}\right]}{Q_0}\right] l^0_{E,i,t}$$

and the Hessian matrix $\nabla^2\Theta$ is defined as

$$\nabla^2\Theta(\Lambda_{u_{z0},0}) \equiv \sum_{i=1}^{N}\left[\frac{E\left[(z_{i0})^2|x_{it}, y_{it}, l_{it}\right]}{Q_0}\right] l^0_{E,i,t} l^{0\prime}_{E,i,t}$$

Taking first order conditions, we obtain the familiar expressions for Newton's method:

$$\nabla^2\Theta\Lambda_{u_{z0}} = \nabla^2\Theta\Lambda_{u_{z0},0} - \nabla\Theta \qquad \Longleftrightarrow \qquad \Lambda_{u_{z0}} = \Lambda_{u_{z0},0} - \left[\nabla^2\Theta\right]^{-1}\nabla\Theta, \tag{57}$$

Implementation Note that we can write the conditional expectations term as:

$$E\left[(z_{i0})^2|x_{it}, y_{it}, l_{it}\right] = E\left[z_{i0}|x_{it}, y_{it}, l_{it}\right]^2 + var\left[z_{i0}|x_{it}, y_{it}, l_{it}\right]$$

B.5 Algorithm

In this section, we present the EM algorithm we use to recover the estimate of persistent earnings as well as the parameters which govern the income process.

1. Guess an initial set of parameters $\theta_0 = [F, Q_E, Q_U, B_E, B_U, R, Q_0]'$.

2. Using the parameter guess θ_0 use the Kalman filter for the state-space system in equations (2) and (1) to obtain an estimate of $\{\{z_{i,t}\}_{i=1}^{N}\}_{t=0}^{T}$, and estimate the log likelihood.

3. Using estimated persistent earnings $\{\{z_{i,t}\}_{t=0}^{T}\}_{i=1}^{N}$ and data $\{\{y_{i,t}\}_{t=0}^{T}, \{l_{i,t}\}_{t=0}^{T}, \{x_{i,t}\}_{t=0}^{T}\}_{i=1}^{N}$, update the parameter vector as follows:

   (a) Update F, B_U, B_E using equation (40).

   (b) Update the shocks to persistent earnings by iterating on equation (46).

   (c) Update the shocks to temporary earnings by iterating on equation (54).

   (d) Update the initial draw of persistent earnings by solving (57).

4. Repeat steps (2) and (3) until the log likelihood is maximized.

B.6 Monte Carlo Exercises

In this appendix, we perform a Monte Carlo (MC) exercise to examine how well our method can recover the path of persistent earnings at the individual level as well as the parameters of the income process. For this MC exercise, we simulate data for 2,500 individuals using the parameters of our baseline income process presented in Table 2. We simulate data for 2,500 individuals as this is a typical sample size in panel data sets such as the PSID. We vary the number of observations for each individual between 35 and 5 to examine the degree to which our method can recover the path of persistent earnings in panels with shorter and longer time dimensions. After simulating the data, we perform our estimation on the simulated data and recover the estimated parameters of the income process and the path of persistent earnings for each simulated individual. Let $\hat{z}_{i,t}$ denote the estimate of persistent earnings for individual i in period t, and let $z_{i,t}$ denote the true value of persistent earnings. To summarize the accuracy of the method, we estimate the following OLS regression:

$$z_{i,t} = \alpha + \beta^{MC} \hat{z}_{i,t} + \epsilon_{i,t} \tag{58}$$

If $\beta^{MC} \approx 1$ then we have evidence that our method is accurately recovering persistent earnings. For each time panel length T we repeat the process outlined above 50 times to examine the variability in our estimates. Table 5 summarizes the results of the MC exercise.

The first column of Table 5 presents the results from simulating data for $T = 35$ years. Across the 50 simulations the average β^{MC} from estimating equation 58 is 1.0007. This coefficient indicates that, on average, our estimates of persistent earnings differ from the true value by 0.07 percent. The fact that this coefficient is so close to 1 indicates that our method can accurately recover the path of persistent earnings at the individual level. The remaining columns of Table 5 show the results for estimations with different time series lengths. We continue to find that even as the panel dimension gets very small (5 observations) we are still able to recover persistent earnings with a high degree of accuracy. For example with 5 observations, on average, our estimate of persistent earnings differs from the true value by 0.4 percent.

We next use the MC exercise to examine the degree to which the estimation method can recover the parameters of the income process. Table 6 presents the average parameter value across the estimations as well as the t-statistic for the difference between the estimated parameter and the true parameter. The method is able to very accurately recover the parameters of the income process. Even when there are very few observations for each individual, the method

Table 5: Monte Carlo Exercise: Recovering Persistent Earnings

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T	35	30	25	20	15	10	5
Avg. β^{MC}	1.0007	1.0008	1.0006	1.0011	1.0012	1.0012	1.0041
	(0.6677)	(0.6780)	(0.4219)	(0.7766)	(0.7301)	(0.4824)	(1.1113)

Notes: Table presents the average coefficient β^{MC} from estimating equation 58. T denotes the number of periods simulated, and in each simulation 2500 indiviudals are simulated. We repeat the simulations 50 times. In parentheses we report the t-statistics that the average β^{MC} is statistically different from 1.

can recover the parameters within the 95% confidence interval.

Table 6: Monte Carlo Exercise: Recovering Parameters

	True Value	Estimated Parameters						
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
T		35	30	25	20	15	10	5
$Q_E^{1/2}$	0.2626	0.2610	0.2608	0.2609	0.2605	0.2602	0.2592	0.2495
		(-0.9691)	(-1.0218)	(-0.8837)	(-1.0029)	(-1.0076)	(-0.8842)	(-1.7387)
$Q_U^{1/2}$	0.5245	0.5252	0.5249	0.5237	0.5231	0.5236	0.5255	0.5153
		(0.0847)	(0.0590)	(-0.0869)	(-0.1460)	(-0.0851)	(0.0557)	(-0.2935)
$R^{1/2}$	0.1896	0.1911	0.1915	0.1914	0.1917	0.1922	0.1929	0.2003
		(1.1121)	(1.0620)	(0.9337)	(1.2441)	(1.3952)	(1.1062)	(1.7324)
B_E	0.0005	0.0003	0.0005	0.0007	0.0006	0.0004	0.0003	0.0004
		(-0.1815)	(0.0194)	(0.1397)	(0.0504)	(-0.0943)	(-0.1199)	(-0.0263)
B_U	-0.1878	-0.1865	-0.1866	-0.1865	-0.1852	-0.1893	-0.1856	-0.1985
		(0.1669)	(0.1134)	(0.1397)	(0.2349)	(-0.1249)	(0.1381)	(-0.3482)
$Q_0^{1/2}$	0.8206	0.8190	0.8158	0.8168	0.8142	0.8232	0.8195	0.8195
		(-0.1237)	(-0.4398)	(-0.2312)	(-0.5071)	(0.2066)	(-0.0772)	(-0.0700)
F	0.9316	0.9319	0.9322	0.9323	0.9327	0.9320	0.9328	0.9369
		(0.2102)	(0.4373)	(0.4635)	(0.6675)	(0.2209)	(0.4278)	(1.0809)

Notes: Table presents the average parameter values recovered by the estimation procedure in the simulated data. T denotes the number of periods simulated, and in each simulation 2,500 individuals are simulated. We repeat the simulations 50 times. In parentheses we report the t-statistics that the average value of the recovered parameter is statistically different from the true value.

B.6.1 What if shocks are non-normal?

In the above MC exercise, we are simulating data using the parameter estimates from the baseline model. In this simulation the shocks to temporary and persistent earnings are drawn from normal distributions. Recently, the literature has found that labor income changes have substantial deviations from a normal distribution (e.g., Guvenen et al. (2014), Guvenen et al. (2021)). Further, the literature finds that using a mixture of normal distributions can help recover the non-Gaussian features of the data. In this appendix, we examine how well our estimation procedure uncovers the path of persistent earnings when the labor income process has non-Gaussian shocks.

For this exercise, we need to simulate data that exhibits deviations from a Gaussian distribution. We simulate data from a distribution with non-Gaussian features by using the estimates from Guvenen et al. (2021). In particular, we use the parameter estimates from their estimation where there is a mixture distribution for both persistent and temporary earnings shocks (parameters presented in their Table 4, column (3)). In this estimation, the shock to persistent and temporary earnings are mixture distributions, where one distribution is drawn with a low probability and has a very negative mean and wide variance, while the other distribution has a near zero mean and a very tight variance. The latter distributions represents how most individuals have very small income fluctuations, while the first set of distributions induces severe negative shocks for a segment of the population. Using this labor income process, we repeat the MC exercise.

Table 7 presents the results of the MC exercise where the shocks to labor income are non-normal. Similar to the case of normally distributed shocks, our estimation routine is able to very accurately recover the path of persistent earnings at the individual level. For all time spans considered (35 observations per person to 5 observations) the average value of β^{MC} is not statistically different from one, indicating the method can successfully estimate persistent earnings when shocks are non-normal.

B.7 Misclassifications

For a small positive $\bar{y} > 0$ and arbitrary employment l_{it} process, there is a very small chance of "misclassification" of an individual as employed when in fact they earn less than $\bar{y}$. For small positive values of $\bar{y}$ used in this paper, we establish in simulated data that having an employed income below $\bar{y}$ is very rare. Using the simulated data from our quantitative model in Appendix D, we can examine how likely it is in our income process for an individual to be employed but have an income below $\bar{y}$. In our simulated data we find that less than 0.3% of

Table 7: Monte Carlo Exercise: Recovering Persistent Earnings with Non-normal Shocks

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
T	35	30	25	20	15	10	5
Avg. β^{MC}	0.9994	0.9993	0.9998	0.9999	1.0002	1.0012	1.0008
	(-0.2798)	(-0.2593)	(-0.0615)	(-0.0320)	(0.0388)	(0.1039)	(0.0327)

Notes: Table presents the average coefficient β^{MC} from estimating equation 58 where the data are simulated from an income process with non-normal shocks to persistent and temporary earnings. T denotes the number of periods simulated, and in each simulation 2500 indiviudals are simulated. We repeat the simulations 50 times. In parentheses we report the t-statistics that the average β^{MC} is statistically different from 1.

employed individuals satisfy this condition.

We then examine how this discrepancy impacts our estimates from the filtering exercise. Using the simulated data from our quantitative model in Appendix D we re-classify the less than 0.3% of individuals who are labeled employed but have earnings below the cutoff $\bar{y}$ as unemployed and run the filtering algorithm on this adjusted simulated data. By comparing the output of the filtering algorithm on this adjusted data to our simulated data where we make no adjustment, we can examine how this misclassification is impacting our results. We find that we obtain nearly identical distributions of shocks with the "adjusted" dataset and the baseline dataset. In Figure 9, we plot the CDF of the shocks to persistent earnings (Panel (a)) and temporary shocks (Panel (b)). In each figure the black, solid line with diamond markers, is the distribution from the baseline estimation while the red, dashed line with circle markers, is the distribution where we have adjusted the employment classification for individuals who are employed but have earnings below $\bar{y}$. The figures show that we obtain nearly identical distributions of shocks across the two specifications. We view this exercise as providing evidence that the potential for employed individuals to have earnings below $\bar{y}$ has a minuscule impact upon our results.

C Additional Details: Empirics

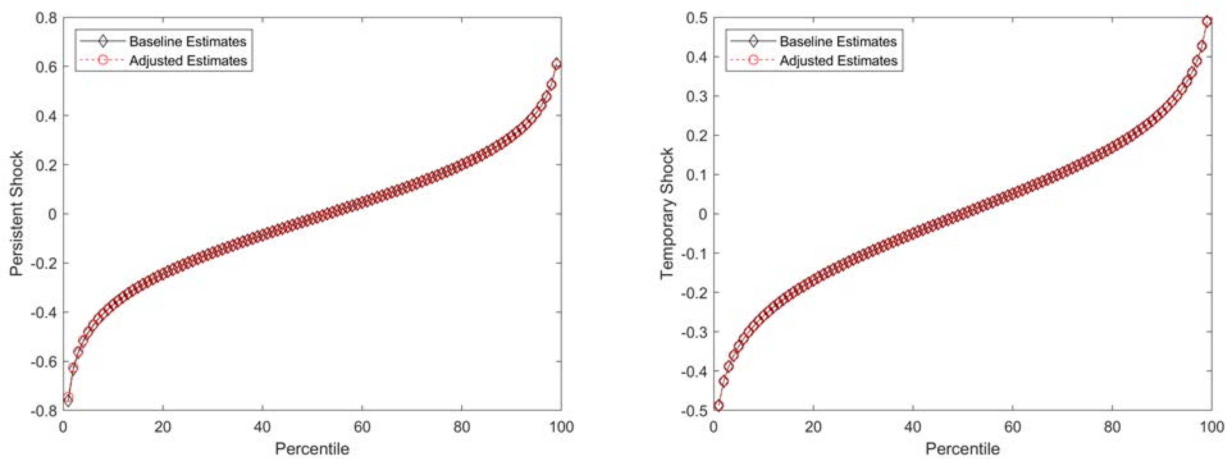
C.1 Residualizing Earnings

We remove the common age component of earnings (residualizing) as in Guvenen et al. (2014). Using all earnings observations from the base sample, we run a pooled regression of earnings on age and cohort dummies without a constant. This regression recovers the age profile of

Figure 9: CDF of Shocks to Persistent and Temporary Earnings

(a) Persistent Shock

(b) Temporary Shock



Note: The figure plots the CDF of the shocks to persistent earnings (Panel (a)) and temporary shocks (Panel (b)). In each figure the black, solid line with diamond markers, is the distribution from the baseline estimation while the red, dashed line with circle markers, is the distribution where we have adjusted the employment classification for individuals who are employed but have earnings below $\bar{y}$.

log earnings. We then scale the age dummies so as to match the average log earnings of 25-year-olds used in the regression. We then subtract the age dummies from earnings to recover residualized earnings.

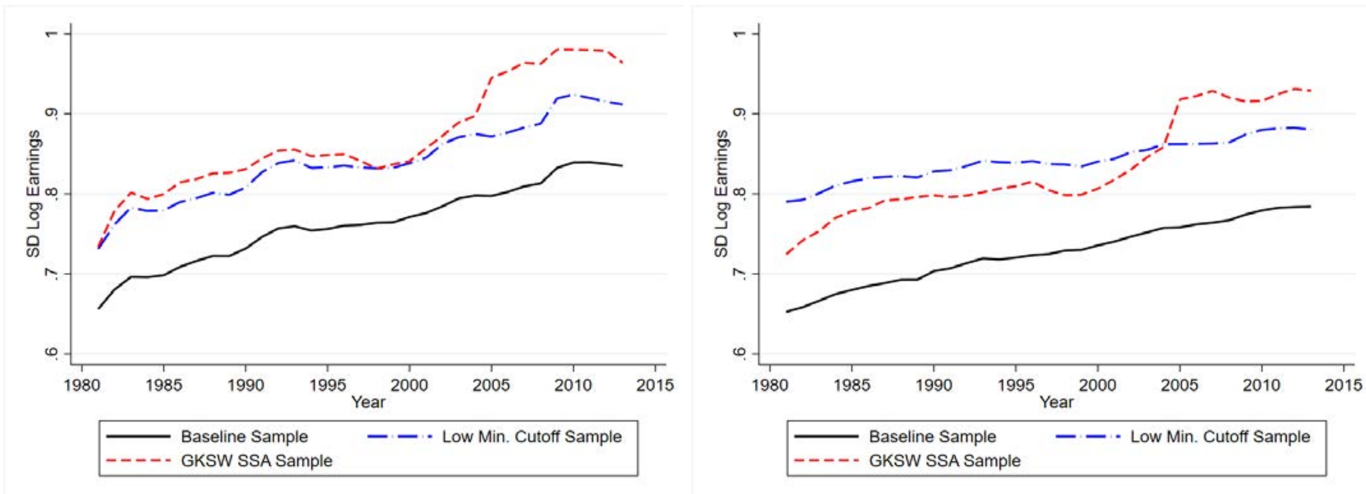
C.2 Representativeness

In this appendix, we assess the representativeness of our linked SSA-CPS sample. We show that moments from our sample align with estimates from (1) the full SSA database as well as (2) the publicly available ASEC.

Comparison to Full SSA database. We first show that median earnings and the standard deviation of earnings closely mirror those reported in Guvenen, Kaplan, Song, and Weidner (2022), hereafter referred to as GKSW, whose sample is the universe of SSA earnings records. Importantly, GKSW impose very similar earnings criteria for sample inclusion: (1) ages 25 to 55 during the panel period (1957-2013), (2) earnings are larger than a year specific minimum earnings criterion, denoted $Y_{min,t}$ in at least 15 years between the ages of 25 and 55, and (3) had a total lifetime earnings of at least $31Y_{min,t}$. GKSW set their minimum earnings criterion to the equivalent of working part-time at the real federal minimum wage for 1 quarter of the year. When the earnings history is truncated condition (2) is updated to 50% of years since 25, and

Figure 10: Representativeness of SSA-CPS linked sample vs. Guvenen et al. (2022)

(a) Std. dev. log earnings, men (b) Std. dev. log earnings, women



Note: Panels (a) and (b) report standard deviations of log earnings in levels for men and women, respectively. The GKSW data are taken from the accompanying supporting documents to Guvenen et al. (2022).

condition (3) is updated to number of years since age 25 times $Y_{min,t}$.

We first compare estimates of the standard deviation of log earnings by gender from our linked SSA-CPS sample to the moments reported in GKSW. Panels (a) and (b) of Figure 10 present the standard deviation of log earnings among men and women, respectively. The black solid line presents the standard deviation of log earnings among individuals in our baseline sample, while the red line presents the standard deviation from GKSW. One important difference between our setup and GKSW is the value of the minimum earnings criterion. The blue dashed-dotted line presents the standard deviation of log earnings from our sampling criteria when we impose the GKSW minimum earnings criterion. The figure shows that there is a similar trend for the standard deviation of log earnings over time across all samples for both men and women. The trend in the standard deviation of log earnings from our sample with the lower minimum earnings criterion is especially similar to the estimates from GKSW.[56] In results that are available upon request, we show that we obtain similar trends in both median earnings and the standard deviation of log earnings by selected age in our samples as well as in the GKSW sample.

The results of this section show that we obtain similar estimates in our linked sample of SSA-CPS earnings as in the full SSA database. We view these results as providing evidence that our linked SSA-CPS sample is representative of the full sample of SSA individuals.

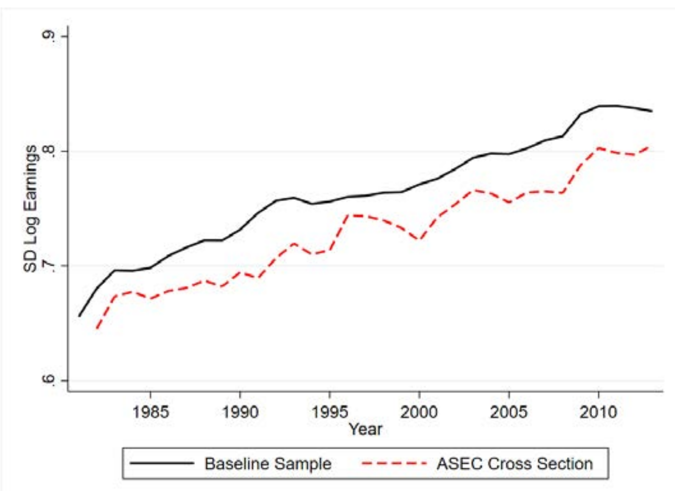
[56]The differences at the end of the sample have to do with a minor difference in sample construction, where we require at least 5 years of data to be included in our sample. GKSW do not have a corresponding sample condition.

Comparison to estimates from the ASEC We next compare the estimates of our sample to estimates from the cross-section of individuals in the ASEC over time in Figure 11. To create estimates from the ASEC, we impose our baseline minimum earnings criterion to be included in the sample. We additionally remove individuals with imputed earnings.

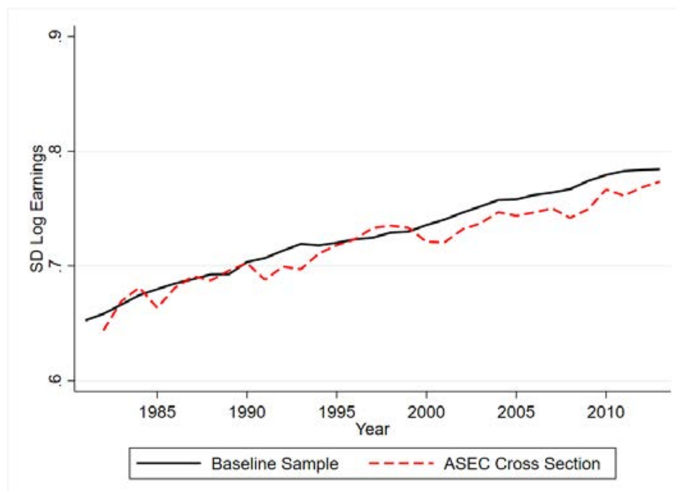
Panels (a) and (b) of Figure 11 plot the standard deviation of log earnings for men and women, respectively, in our baseline sample (black solid line) as well as in the ASEC-cross-section (red dashed line). The figure shows that we obtain very similar trends in the standard deviation of log earnings among the linked SSA-CPS sample as in the full cross-section of the ASEC.

Figure 11: Representativeness of SSA-CPS linked sample vs. ASEC

(a) Std. dev. log earnings, men



(b) Std. dev. log earnings, women



Note: Note: Panels (a) and (b) report standard deviations of log earnings in levels for men and women, respectively. The ASEC data was downloaded from IPUMS.

C.3 What drives persistent and temporary income shocks?

In this appendix, we examine how our filtered estimates of temporary and persistent earnings shocks align with job switches, and job loss in our SSA-CPS database. For each labor market event, we report the joint density of persistent and temporary shocks. We illustrate these joint densities as heatmaps whose colors correspond to the mass of individuals with a given combination of persistent and temporary shocks.

As we are showing results from the distribution of shocks to temporary and persistent earnings from our filtering method we must take into account *filtering uncertainty*. The Kalman smoother returns an estimate of persistent earnings for individual i in period t and the lag of persistent earnings, i.e., $\hat{\zeta}_{i,t|T} = \begin{bmatrix} \hat{z}_{i,t|T} & \hat{z}_{i,t-1|T} \end{bmatrix}'$. The Kalman smoother also produces an estimate about the uncertainty of this estimate, which is given by the MSE matrix $M_{i,t|T}$. To obtain an estimate of persistent earnings for individual i in period t taking into account filtering uncertainty, denoted $\hat{z}_{i,t}$, we draw normal noise, denoted $\xi_{i,t}$, from a bi-variate normal distribution with mean zero and variance-covariance matrix $M_{i,t|T}$. Let $\xi_{1,i,t}$ and $\xi_{2,i,t}$ denote the first and second elements of $\xi_{i,t}$, respectively. For each individual i in period t, we recover persistent earnings $\hat{z}_{i,t} = \hat{z}_{i,t|T} + \xi_{1,i,t}$, the persistent earnings innovation $\Delta\hat{z}_{i,t} = \hat{z}_{i,t|T} + \xi_{1,i,t} - (\hat{z}_{i,t-1|T} + \xi_{2,i,t})$ and the temporary earnings innovation $\hat{\omega}_{i,t} = y_{i,t} - \hat{z}_{i,t}$.[57]

C.3.1 Job Switching

We first plot heatmaps of the shocks to persistent and temporary earnings for individuals who remain at the same primary employer (EIN) across two consecutive years (Panel (a) of Figure 12) and for individuals who switch their primary employer (Panel (b) of Figure 12).[58] Panel (a) shows that among job stayers, the majority of individuals have small shocks to temporary and persistent earnings. These shocks likely reflect changes in hours and weeks worked, as well as promotions, and raises, etc. Conversely, Panel (b) shows that among job switchers, the mass of individuals spreads out of the middle of the distribution towards more extreme persistent and temporary shocks (either positive or negative). To facilitate comparison, Panel (c) of Figure 12 subtracts the joint density in Panel (a) from Panel (b). The resulting difference in densities more clearly illustrates that job switching is associated with larger shocks (both positive and negative) to persistent and temporary earnings. Among non-switchers, roughly 6% have the most extreme earnings outcomes (lowest or highest persistent and temporary shocks). Among switchers, over 16 percent have the most extreme earnings outcomes, representing nearly a threefold increase.

We further split job switchers by the type of job switch. Using data from the Longitudinal Business Database (LBD), we measure average earnings per employer.[59] We separate job switchers into those who move to an employer with average earnings per worker that are 25%

[57] Note that in every time period t, filtering uncertainty is not fully resolved for the current and lagged innovation. Filtering uncertainty is not iid, and any comparison of current and lagged values of persistent earnings requires a correlated draw of filtering uncertainty for the current and lagged persistent earnings. The bivariate filtering uncertainty distribution is over the current and lagged innovation. The primary benefit of including lagged persistent earnings in the state vector is that this bivariate distribution of filtering uncertainty is estimated by the Kalman filter.

[58] An individual's primary employer in a given year is the defined as the EIN where the individual had the largest share of earnings in that year.

[59] See Jarmin and Miranda (2002) for details on the construction of the LBD.

lower (higher) than their previous employer. Panel (d) (Panel (e)) of Figure 12 shows that when an individual moves to a lower (higher) paying employer, they become more likely to experience a large negative (positive) shock to persistent earnings. To facilitate comparison, Panel (f) of Figure 12 subtracts the joint density in Panel (d) from Panel (e). Panel (f) demonstrates that moving to a higher-paying firm is associated with positive shocks, especially to persistent earnings.

The results of this section demonstrate that the estimates of temporary and persistent shocks align with observable labor market events. In particular, the estimates align with job ladder models of the labor market in which job switching is associated with shocks to temporary and persistent earnings that are larger than those associated with remaining at the same employer. Further, the direction of the job switch (i.e., moving to a higher or lower paying employer) aligns with the notion of climbing up and falling down the job ladder.

C.3.2 Layoff

In this section, we examine temporary and persistent earnings shocks around layoff. While many papers have studied the average response of earnings to layoffs, we examine the heterogeneous behavior of temporary and persistent earnings following layoff. We document substantial heterogeneity in earnings following layoff and how it correlates with observable features of the layoff.

We identify layoffs using an individual's self-reported CPS responses. In particular, we define an individual to have been laid off in year t if they report having positive weeks on layoff in year t and zero weeks on layoff in year $t-1$. We impose the requirement that an individual have zero weeks on layoff in year $t-1$ so that we are able to accurately measure the inflow of individuals into unemployment. In Panel (a) of Figure 13, we plot the heatmap of persistent and temporary earnings shocks in year t for individuals we identify as laid off in the CPS. The figure shows that there is a large mass of individuals in the bottom left hand corner of the heatmap, which indicates that a sizeable mass of laid off individuals have large negative persistent and temporary shocks. Interestingly, there is also a large mass of individuals with small shocks, and there are even some individuals with positive shocks.

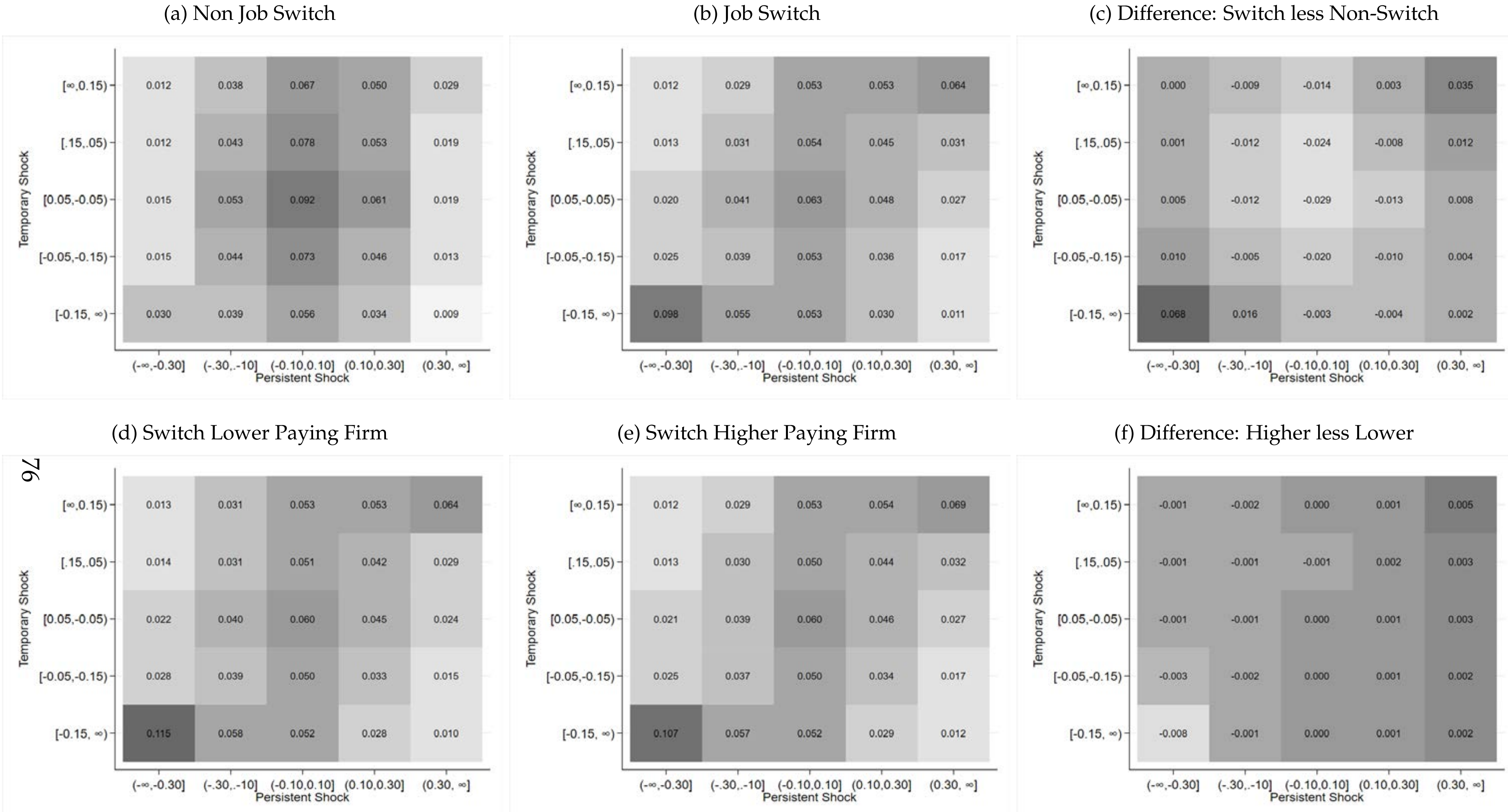
We investigate the heterogeneity in shocks around layoffs by distinguishing *recalls* from *non-recalls*. We define an individual to be recalled if their primary employer in the year before layoff is also their primary employer in the year after layoff.[60] We define an individual to be non-recalled if they have different primary employers in the years before and after layoff. Panel (b)

[60] As in section C.3.1, we define an individual's primary employer in a given year as the EIN where they have the largest labor earnings.

of Figure 13 plots the heatmap of persistent and temporary shocks among recalled individuals, while Panel (c) of Figure 13 plots the heatmap for non-recalled individuals, and Panel (d) illustrates the difference. Comparing Panels (b) and (c) shows that relative to individuals who are not recalled after layoff, individuals who are recalled exhibit much smaller negative shocks to temporary and persistent earnings, and they are more likely to have a positive shock.

Taking stock. The results of this section showed that our estimates of temporary and persistent earnings align with observable shocks that individuals face in the labor market. As the filter is unaware of the shocks individuals face, we view these results as a validation of the method.

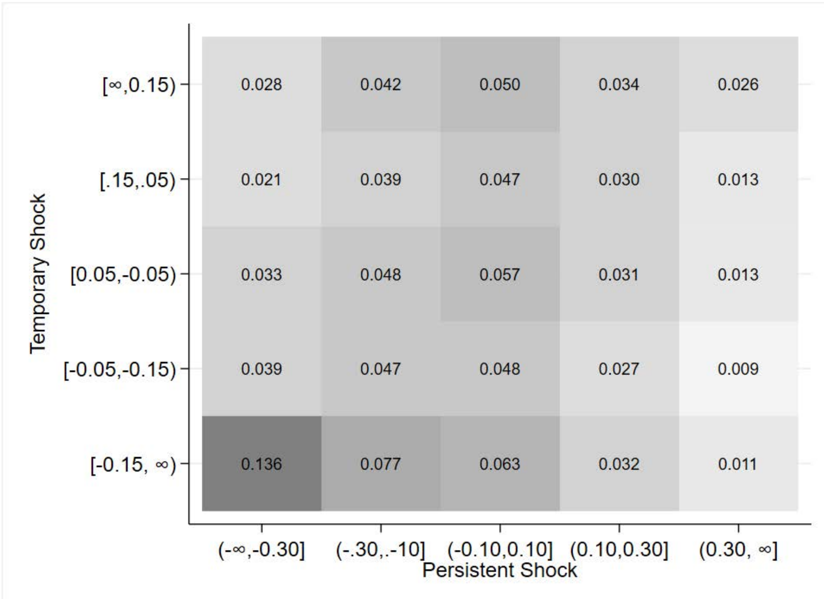
Figure 12: Shocks to temporary and persistent earnings around job switching



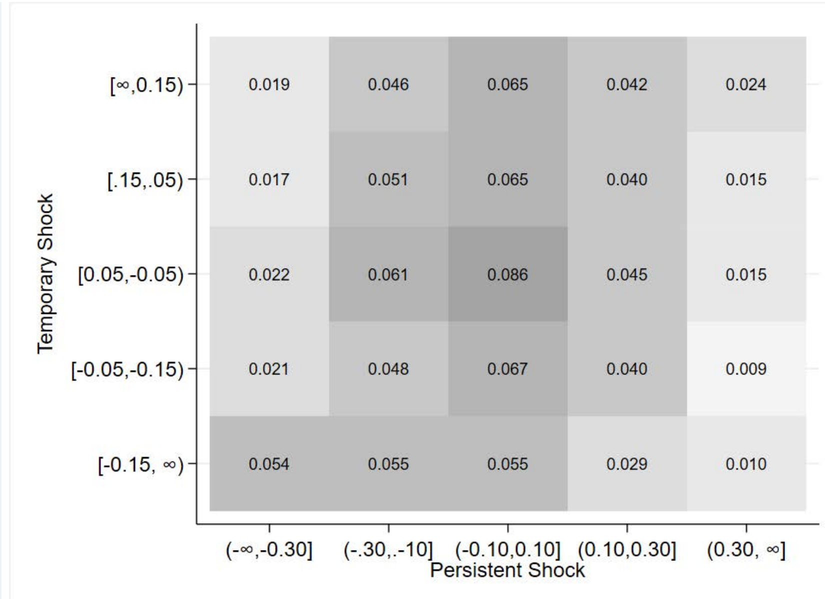
Note: Figure plots a heatmap of temporary and persistent shocks by observable labor market event. Higher (lower) paying firms are identified by moving to an employer with average earnings that are 25% above (below) an individuals current employer.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

Figure 13: Shocks to temporary and persistent earnings around layoff

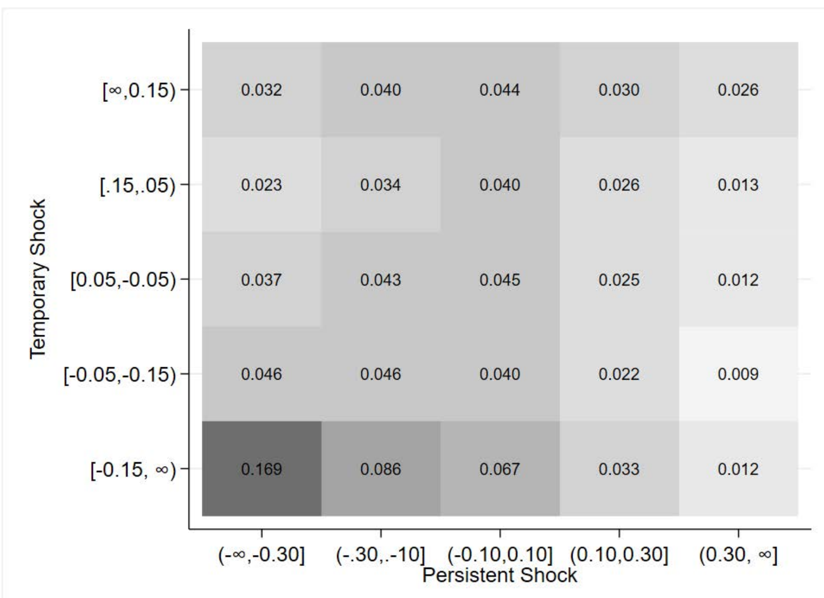
(a) Layoffs



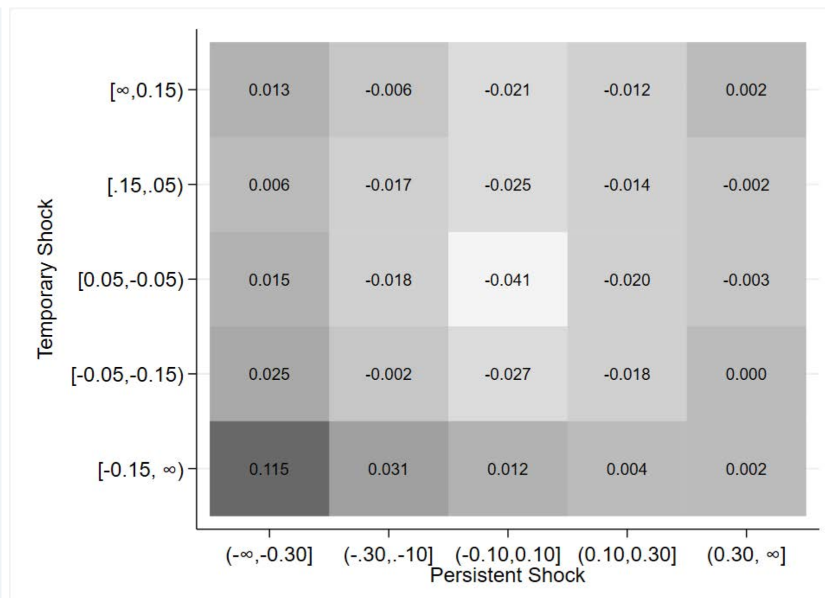
(b) Layoffs: Recalls



(c) Layoffs: Non-Recalls



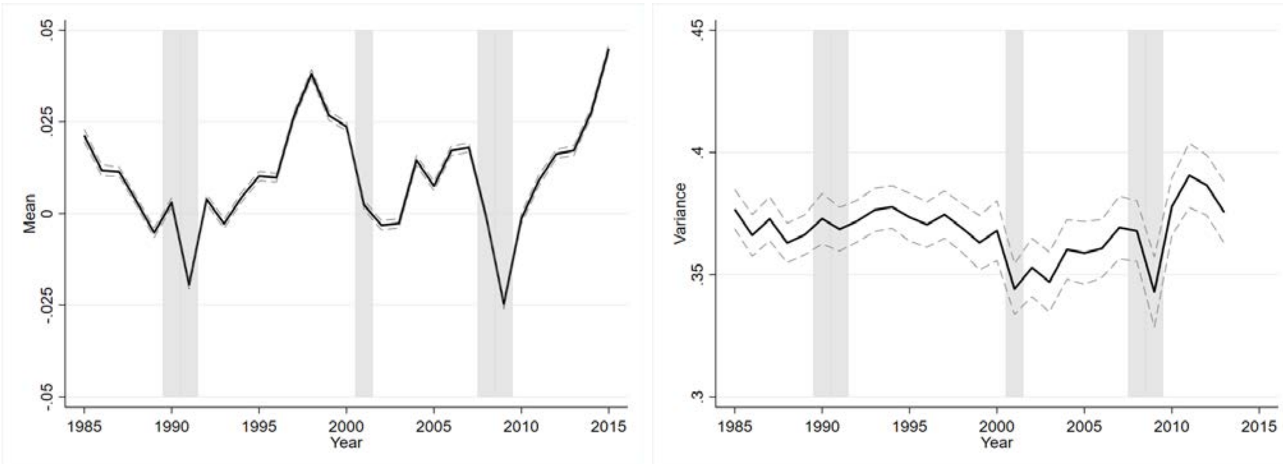
(d) Difference: Non-Recall less Recall



Note: Figure plots a heatmap of temporary and persistent shocks around layoff. Layoffs are identified using the CPS. Individuals are defined as "recalled" if their primary employer in the year after layoff is the same as the year before layoff.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

Figure 14: Additional time series from Model 2

(a) Drift Persistent Earnings (Emp.) (b) Variance Initial Draw Persistent Earnings



Note: Figure presents time series estimates from estimating Model 2. Panel (a) presents the mean (drift) of shocks to persistent earnings among the employed. Panel (b) presents the variance of the initial draw of persistent earnings. Dashed gray lines denote a 95% confidence interval. Gray bars denote NBER recessions.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.4 Additional Results: Estimating Model 2

In this appendix, we present additional results from the estimation of Model 2. We first present additional time series results, and then present the estimates of how risk evolves over the life-cycle.

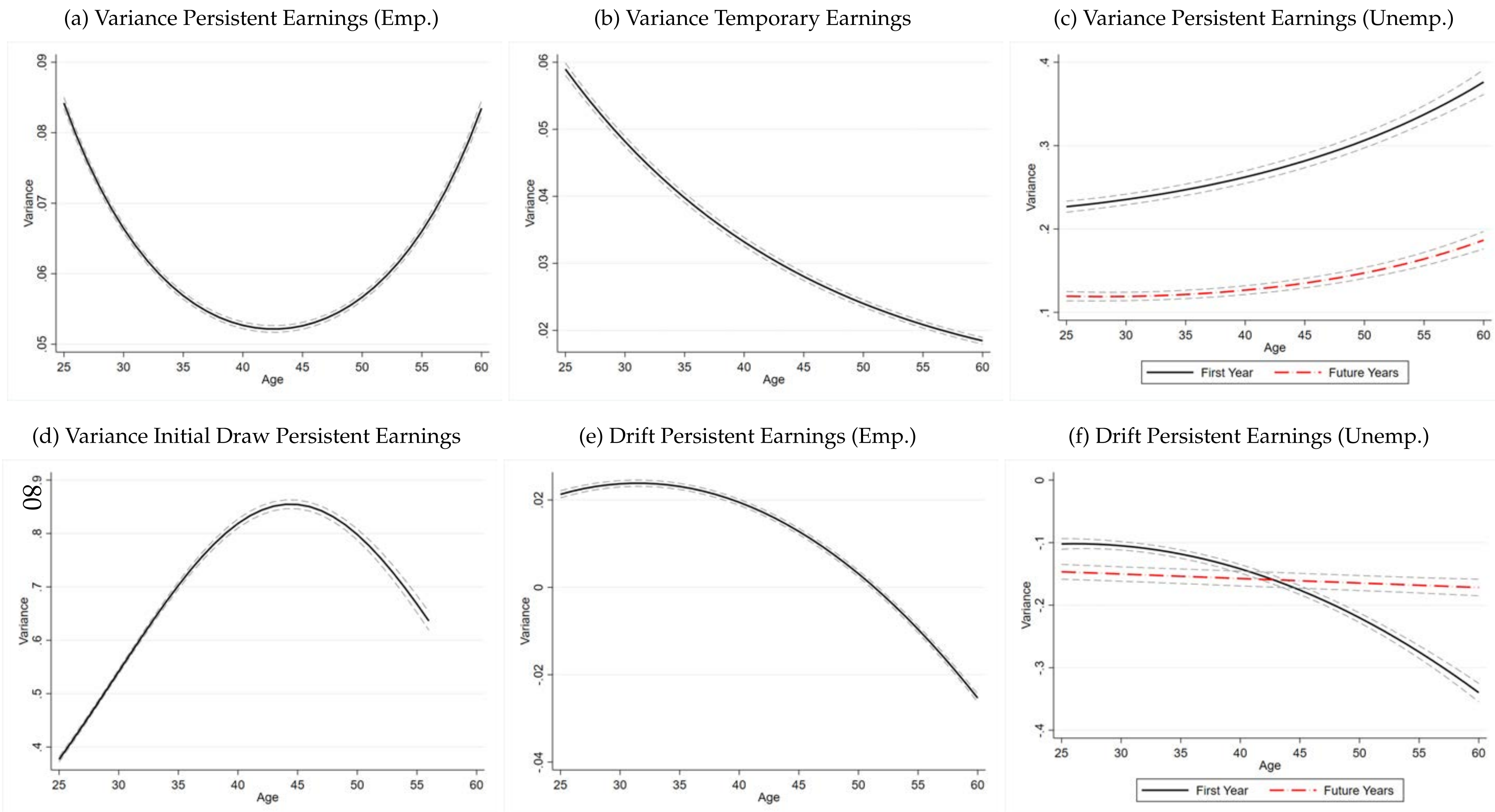
Additional time series results. In Figure 14, we present additional time series produced in the estimation of Model 2. Panel (a) of Figure 14 presents the mean (drift) of shocks to persistent earnings while employed by year. The figure shows that the mean of shocks to persistent earnings is highly cyclical, with the average shock to persistent earnings declining in recessions and increasing in expansions. This cyclical behavior aligns with work by Guvenen et al. (2014), who show that the distribution of shocks to earnings shifts to the left during recessions relative to expansions. In panel (b) of Figure 14, we plot the variance of the initial draw of persistent earnings. The figure shows that there has been a relatively stable level of the initial draw of persistent earnings since the mid-1980s.

Shocks by age. In Figure 15, we present the age profiles of the means and variances of shocks to temporary and persistent earnings over the life-cycle as estimated by Model 2. Panel (a) of Figure 15 shows that the variance of shocks to persistent earnings among the employed is "U-shaped" over the life-cycle.[61] The shape of this profile is in part influenced by how indiviudals transition across jobs over the life-cycle. We show in Appendix C.7 that the variance of shocks to persistent earnings is larger for job switchers relative to job stayers. Given that the likelihood of job switching decreases over the life-cycle, this contributes to the decline in persistent earnings risk over the first half of the life-cycle. We also show in Appendix C.7 that the increase in persistent risk among individuals over the age of 45 occurs among both job switchers and job stayers.

Panel (b) of Figure 15 presents the profile of temporary earnings risk over the life-cycle. The figure shows that temporary earnings risk decreases over the life-cycle. As for persistent risk, this shape is largely influenced by the declining nature of job switching over the life-cycle, as job switchers undergo much larger temporary shocks relative to job stayers (see Appendix C.7). Panel (c) shows that the variance of shocks to persistent earnings is increasing over the life-cycle in both the first year of unemployment (black, solid line) and in all future years of unemployment (red, dashed line). Panel (d) of Figure 15 shows that the variance of the initial draw of persistent earnings is increasing in the age in which an individuals shows up in our sample. Panel (e) of Figure 15 shows that the mean shock to persistent earnings is decreasing over the life-cycle. Finally, panel (f) of Figure 15 shows that the mean shocks to persistent earnings during the first year of unemployment is decreasing with age, i.e., older individuals see a larger decline from entering unemployment relative to younger individuals.

[61] Karahan and Ozkan (2013) also find that shocks to persistent earnings are U-shaped over the life-cycle using the PSID.

Figure 15: Persistent and temporary earnings risk by age



Note: Figure presents parameter estimates of the shocks to earnings over the life-cycle from estimating Model 2. Panel (a) plots the variance of persistent earnings among the employed. Panel (b) plots the variance of temporary shocks. Panel (c) plots the variance of persistent shocks to the unemployed in the first year of unemployment (black, solid line) and in all future years of unemployment (red, dashed line). Panel (d) plots the variance of the initial draw of persistent earnings. Panel (e) plots the mean (drift) of shocks to persistent earnings among the employed. Panel (f) plots the mean of persistent shocks to the unemployed in the first year of unemployment (black, solid line) and in all future years of unemployment (red, dashed line). Dashed gray lines denote a 95% confidence interval.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.5 Additional Results: Identifying risk among unemployed

In this appendix, we present additional results about how trends in the risk faced by unemployed workers evolve over time. In Appendix A.1, we show that the trends in the mean and variance of shocks to persistent earnings while unemployed can be identified from the time series of the mean and variance of the (quasi) change in log earnings over two-years where in the middle year an individual is unemployed. In Panel (a) of Figure 16, we plot the variance of the (quasi) earnings change over two-years for individuals who are unemployed in the middle year. The figure shows that the variance of these earnings changes has increases substantially over the sample period. Accordingly, panel (c) of Figure 16 shows that the implied variance of shocks to persistent earnings among the unemployed has increased over the sample period. In panel (b) of Figure 16, we plot the mean of the (quasi) difference in earnings over two years around an unemployment spell. The figure shows that the decline in earnings around these unemployment spells has gotten larger over our sample period. This acceleration in earnings declines around unemployment spells generates a larger mean decline in persistent earnings during unemployment, which we show in panel (d) of Figure 16.

C.6 Additional Results: Shocks by age over time

In this appendix, we examine how income risk has evolved over time across the age distribution. To do so, we estimate a fourth model of income risk, which we specify below:

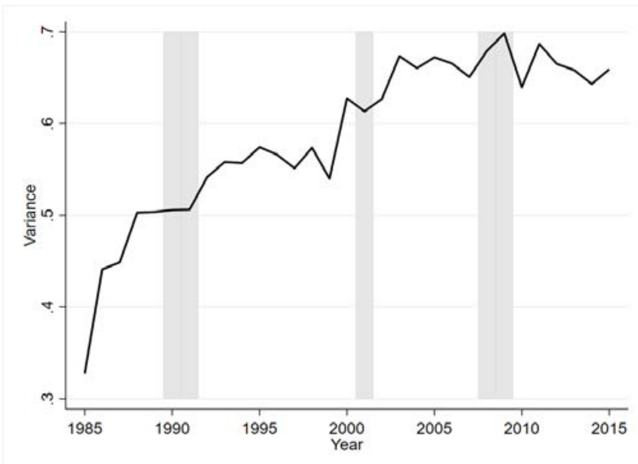
- **Model 4.** The income process presented in Section 1.1, extended so that unemployment spells are split into the first year of unemployment and all future years of unemployment. Additionally, we will allow the variance parameters (Q_E, Q_U, Q_0 and R) and mean parameters (B_E and B_U) to vary by age via an age quadratic, where the age quadratics are specific to a decade (e.g., 1981-1989, 1990-1999, etc.).

Figure 17 presents the results of estimating Model 4. Panel (a) of Figure 17 plots the variance of persistent shocks to the employed by age as well as by decade. Comparing the 2010s (red, dashed line) to the 1980s (black, solid line), we find that persistent earnings risk among the employed has increased the most among young workers. In particular, persistent earnings risk among the young increases by upwards of 40 percent. Alternatively for older workers in our sample there has been a small increase in persistent earnings risk, e.g., a 5 percent increase among 55-year olds.

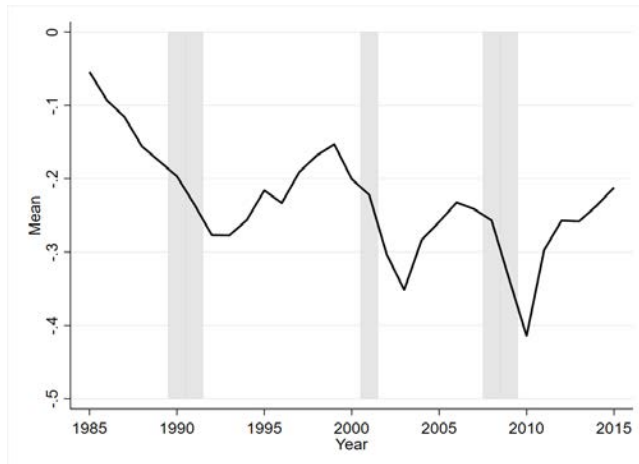
Panel (b) of Figure 17 plots the variance of shocks to persistent earnings among the unemployed in their first period of unemployment by age and decade. The figure shows that the

Figure 16: Identifying changes in risk over time among unemployed

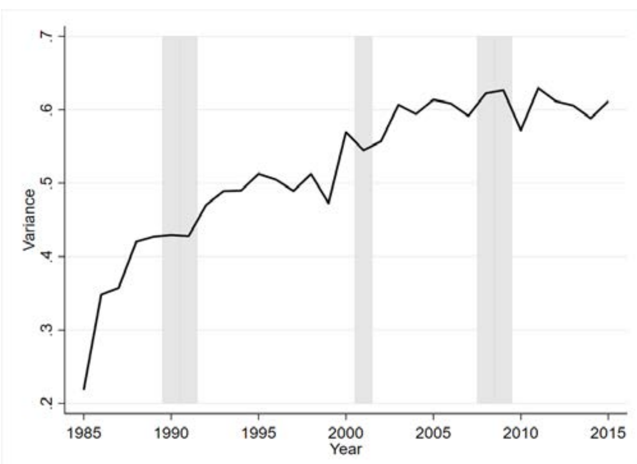
(a) Variance Earnings Changes



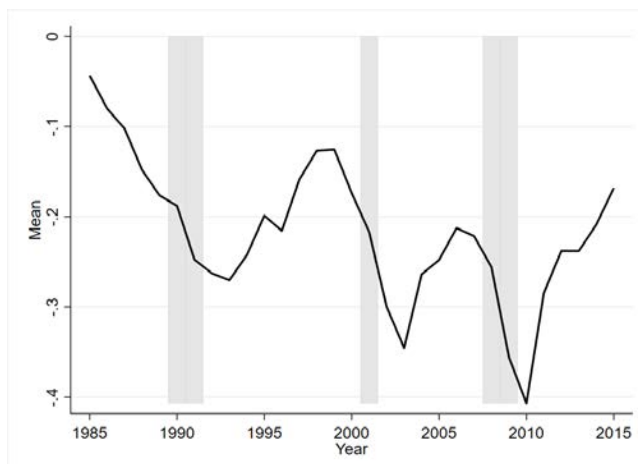
(b) Mean Earnings Changes



(c) Implied Variance of Persistent Shock



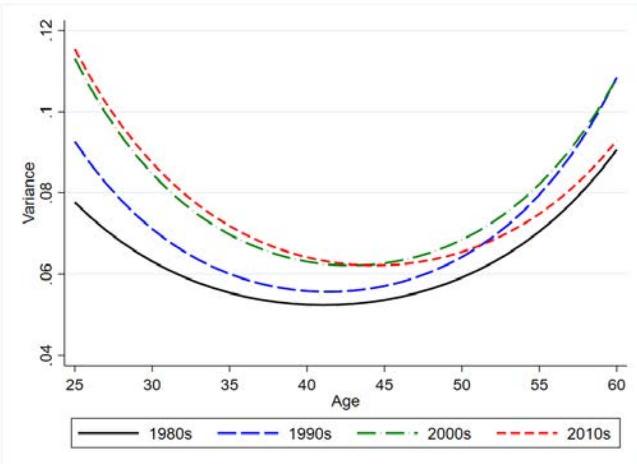
(d) Implied Mean of Persistent Shocks



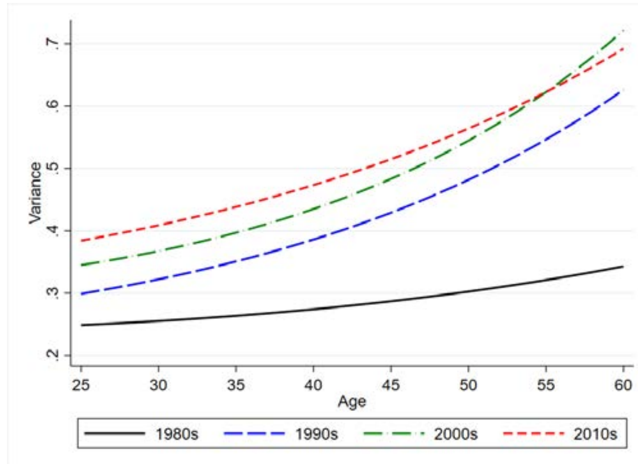
Note: Panel (a) plots the variance of the quasi-difference of log earnings over a two year horizon where the individual was unemployed in the middle year. Panel (b) plots the mean of the quasi-difference of log earnings over a two year horizon where the individual was unemployed in the middle year. Panel (c) plots the implied path of the variance of shocks to persistent earnings among the unemployed, while panel (d) plots the implied path of the mean of shocks to persistent earnings among the unemployed using the identification argument in Appendix A.1. Gray bars denote NBER recession dates.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

Figure 17: Persistent and temporary risk over time by age

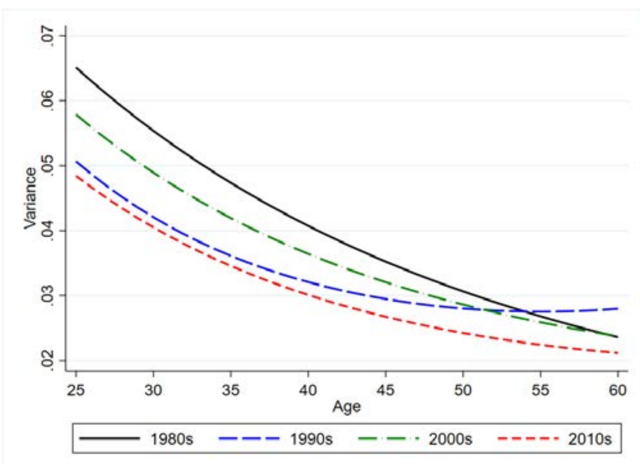
(a) Variance Persistent Shocks Employed



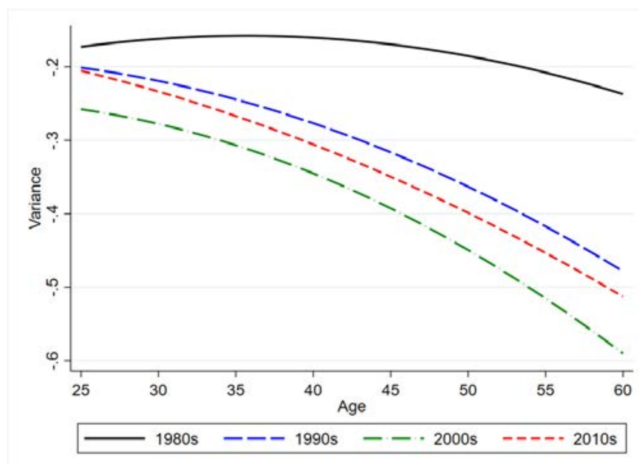
(b) Variance Persistent Shocks Unemp., First Period



(c) Variance Temporary Shocks



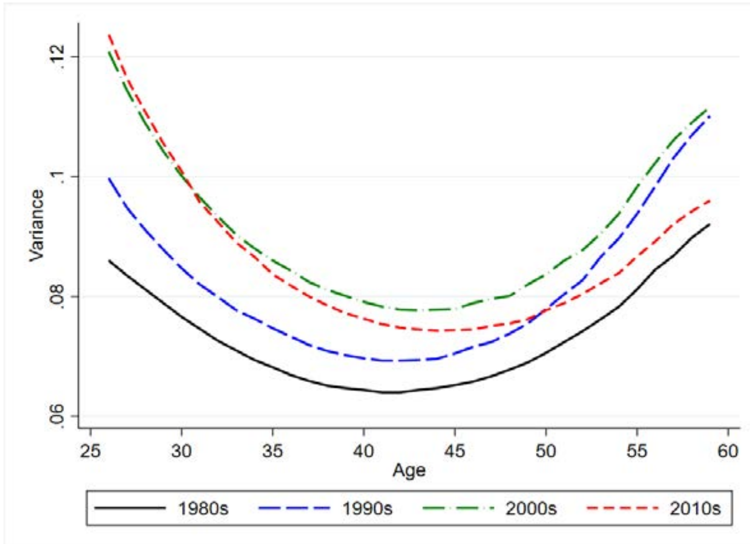
(d) Mean Persistent Shocks Unemployed, First Period



Note: Figure presents the results of estimating Model 4. Panel (a) plots the variance of shocks to persistent earnings among the employed. Panel (b) plots the variance of shocks to persistent earnings among the unemployed, in their first period of unemployment. Panel (c) plots the variance of shocks to temporary earnings. Panel (d) plots the mean of shocks to persistent earnings among the unemployed, in their first period of unemployment. The black, solid line corresponds to 1980s, the blue, long dashed line corresponds to the 1990s, the green, dashed-dotted line corresponds to the 2000s, and the red, dashed line corresponds to the 2010s.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

increase in persistent income risk among the unemployed over time has occurred for workers of all ages but is most pronounced among older workers. In panel (c) of Figure 17, we show the evolution of temporary risk over time by age. The figure shows that the decline in temporary earnings risk between the 1980s and 2010s has occurred for workers of all ages, but is most

Figure 18: Combined persistent risk over time by age



Note: Figure presents combined persistent risk over time by age and decade. The black, solid line corresponds to 1980s, the blue, long dashed line corresponds to the 1990s, the green, dashed-dotted line corresponds to the 2000s, and the red, dashed line corresponds to the 2010s.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

pronounced among the young. Finally, panel (d) of Figure 17 presents the drift in persistent earnings among the unemployed in their first period of unemployment. The figure shows that the decline in persistent earnings during unemployment spells has become larger for workers of all ages since the 1980s, but that the decline has accelerated the most among older workers.

Using the parameter estimates from Model 4 and the shares of workers across employment/unemployment by age and decade we can compute how combined persistent earnings risk has evolved over time. Figure 18 plots combined persistent risk by age over the life-cycle for each decade from the 1980s to the 2010s. Between the 2010s and the 1980s, combined persistent income risk has increased for workers of all ages, but the increase has been most pronounced among younger workers.

C.7 Additional Results: Role of job switching and staying

In this appendix, we examine how income risk has changed over time among job stayers and job switchers across the age distribution. We define an individual to be a *job stayer* in a year t, if they have the same primary employer (EIN) in year t and year $t-1$, while a job switcher in a year t has a different primary employer (EIN) in year t and $t-1$. To examine how income risk has evolved over the age distribution among switcher and stayers, we estimate a version of Model 4, where we split the employed in job switchers and stayers.[62]

We first present estimates for how shocks to persistent earnings have evolved over time among switchers and stayers. Panel (a) of Figure 19 presents the variance of shocks to persistent earnings among job stayers by age for the 1980s (black, solid line) and the 2010s (red, dashed line). The figure shows that persistent earnings risk among stayers has increased between the 2010s and 1980s for workers under the age of 50. Panel (b) of Figure 19 presents the variance of shocks to persistent earnings among job switchers by age for the 1980s (black, solid line) and the 2010s (red, dashed line). First, observe that the variance of shocks to job switchers is substantially larger than for job stayers. Second, the figure shows that persistent risk among job switchers has increased from the 1980s to the 2010s across the age distribution with the largest increases being for younger and older workers.[63]

We next examine how shocks to temporary earnings have evolved over time among switchers and stayers. Panel (c) of Figure 19 presents the variance of temporary earnings shocks among job stayers by age in the 1980s (black, solid line) and the 2010s (red, dashed line). The figure shows that there has been a decline in temporary earnings risk among job stayers that is nearly uniform across the age distribution. Finally, panel (d) of Figure 19 presents the variance of temporary earnings shocks among job switchers by age in the 1980s (black, solid line) and the 2010s (red, dashed line). As for persistent shocks, job switchers face substantially more dispersion in temporary shocks relative to job stayers. Across time, we find that the variance of shocks to temporary earnings among job stayers has declined between the 1980s and 2010s for workers who are less than 50-years old.

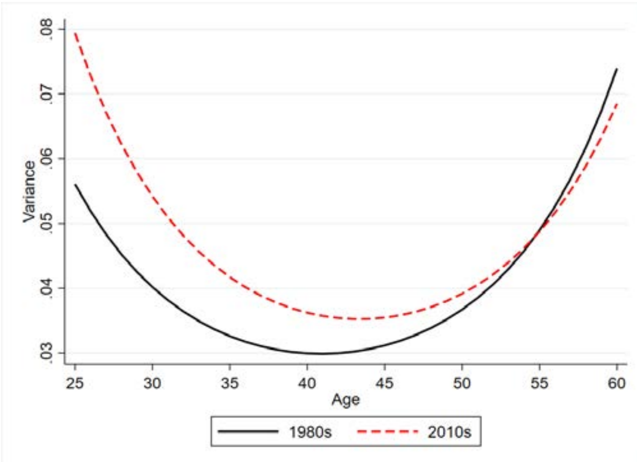
Finally, using the parameters from estimating Model 4 with the employed split into job switcher and stayers as well as data on labor market flows by age and decade we can compute combined persistent risk by age for each decade. In Figure 20, we present how combined persistent risk when taking into account job switching. Similar, to the results presented in Appendix C.6, we find that combined persistent income risk has increased across the age distribution with the largest increase occurring among younger workers.

[62]We examine job switchers and stayers using Model 4, which allows for means and variance of persistent and temporary shocks to evolve by age and decade, given that the likelihood job switching is strongly predicted by age. We find that the likelihood of job switching over the life-cycle declines from nearly 35% (in a year) for younger workers to just over 10% for older workers.

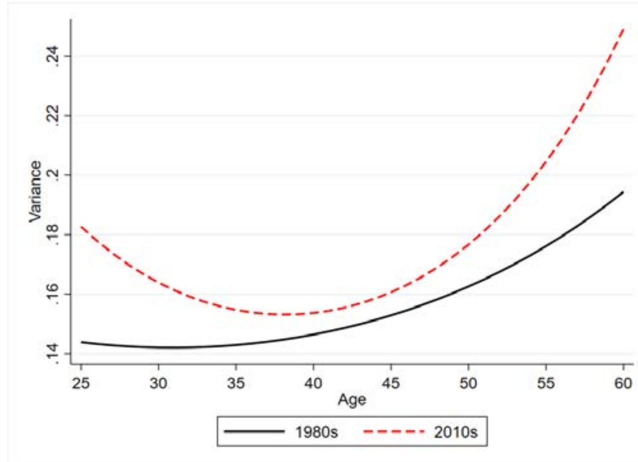
[63]Changes in persistent earnings risk among the unemployed evolve in the same manner as presented in Appendix C.6. These results are available upon request.

Figure 19: Persistent and temporary risk over time for job switchers and stayers

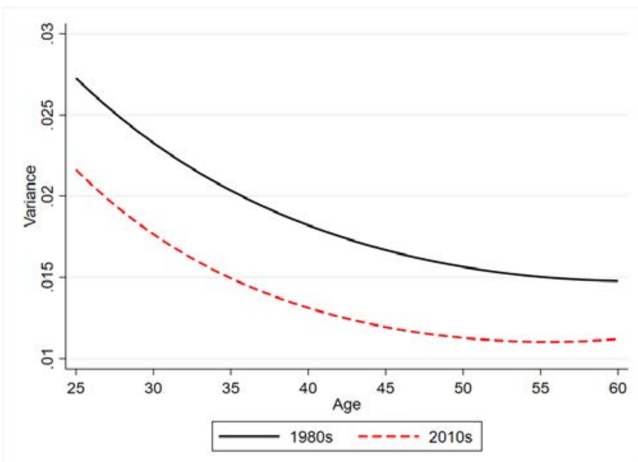
(a) Variance Persistent Shocks, Emp. Job Stayers



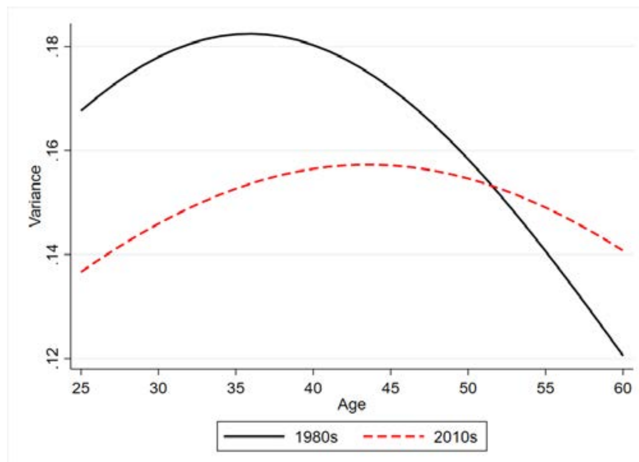
(b) Variance Persistent Shocks, Emp. Job Switchers



(c) Variance Temporary Shocks, Job Stayers



(d) Variance Temporary Shocks, Job Switchers

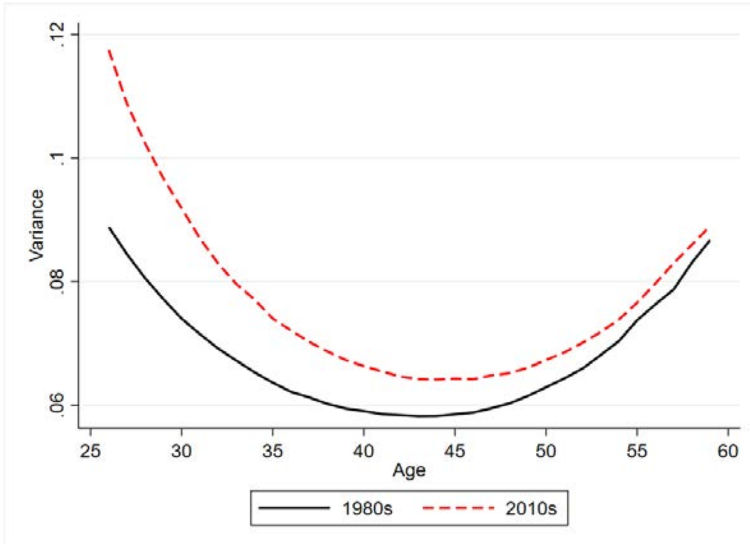


Note: Figure presents the results of estimating Model 4, where the employed are split into job switchers and stayers. Panel (a) plots the variance of shocks to persistent earnings employed, job stayers. Panel (b) plots the variance of shocks to persistent earnings among employed, job switchers. Panel (c) plots the variance of shocks to temporary earnings among job stayers. Panel (d) plots the variance of shocks to temporary earnings among job switchers. The black, solid line corresponds to 1980s, while the red, dashed line corresponds to 2010s.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.8 Additional Results: Gender

In this appendix, we examine how income risk has evolved for men and women separately. Figure 21 presents the results of estimating Model 2 separately for men and women. The figure shows that we see similar increases in persistent earnings risk for both men (black, solid line)

Figure 20: Combined persistent risk over time by age with job switcher/stayer



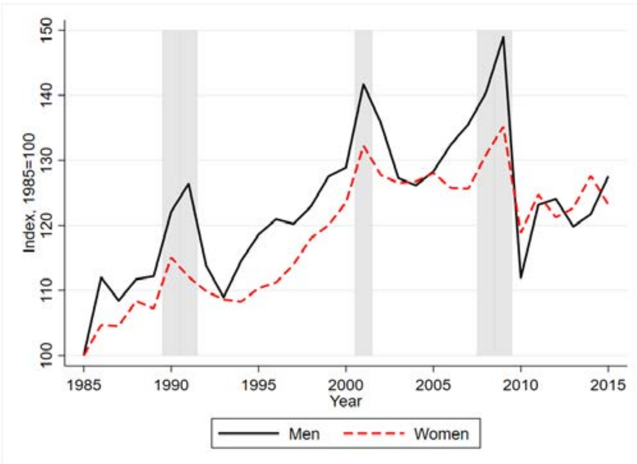
Note: Figure presents combined persistent risk from the estimation of Model 4 with the employed split into job stayers and switchers. The black, solid line corresponds to 1980s, and the red, dashed line corresponds to the 2010s. Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

and women (red, dashed line) while employed (panel (a)) and while unemployed (panel (b)). Similarly, there is a decline in temporary earnings risk between 1985 and 2015 for both men and women (panel (c)). Finally in panel (d), we find that declines in persistent earnings during spells of unemployment have become larger for both men and women, however, men have been more exposed to this trend relative to women. While men and women have seen similar trends in these outcomes, the time series also show that men's income risk experiences more cyclical variation relative to women. This results aligns with work by Doepke and Tertilt (2016), who argue that men work in jobs that tend to be more exposed to the business cycle.

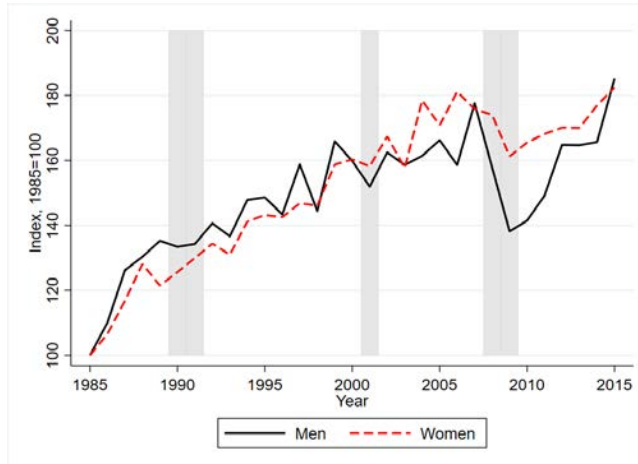
Finally, we examine how combined persistent earnings risk has evolved over time for both men and women. Figure 22 plots the path of combined persistent earnings risk for men (black, solid line) and women (red, dashed line). The figure shows that combined persistent income risk has increased for both men and women, and that by the 2010s both men and women have seen approximately a 20% increase in risk since 1985. One notable difference between genders is their exposure to the business cycle. The evolution of combined persistent income risk for men shows greater cyclical variation, with spikes in each of the recessions during our sample period.

Figure 21: Persistent and temporary risk over time by gender

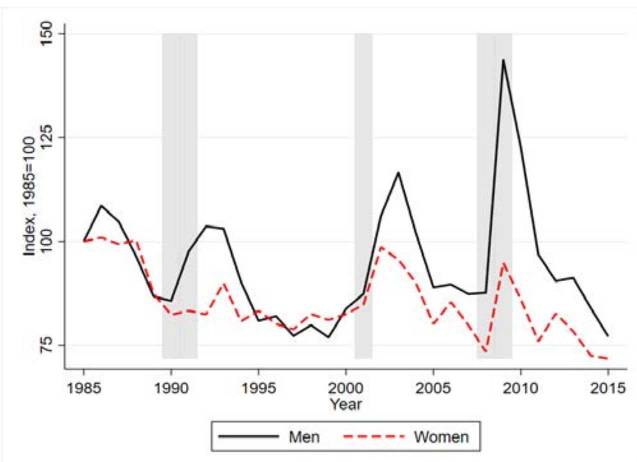
(a) Variance Persistent Shocks Employed



(b) Variance Persistent Shocks Unemp., First Period



(c) Variance Temporary Shocks



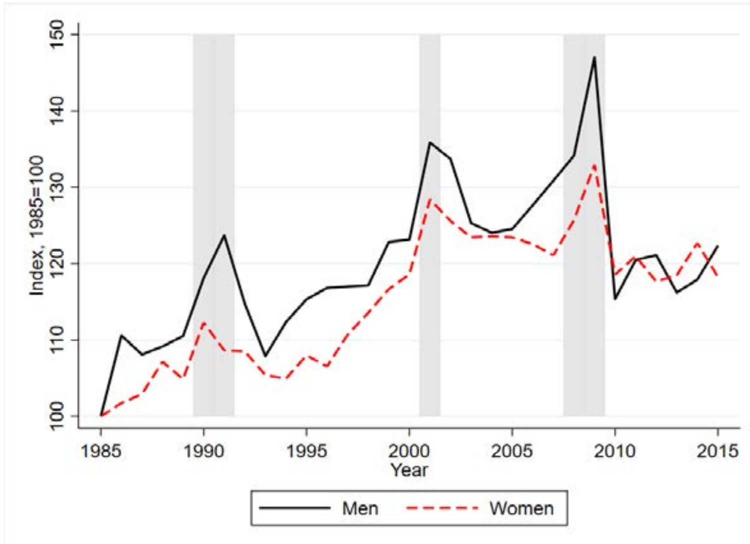
(d) Mean Persistent Shocks Unemployed, First Period

Note: The figure presents the results of estimating Model 3, where the parameters vary by gender. Panel (a) plots the variance of shocks to persistent earnings among the employed. Panel (b) plots the variance of shocks to persistent earnings among the unemployed, in their first period of unemployment. Panel (c) plots the variance of shocks to temporary earnings. Panel (d) plots the mean of shocks to persistent earnings among the unemployed, in their first period of unemployment. The black, solid line corresponds to men, while the red, dashed line corresponds to women.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.9 Additional Results: Minimum earnings cutoff

In this appendix, we examine the robustness of our results to alternative values of the minimum earnings cutoff. We find that we obtain similar time trends if we decrease the value of the minimum earnings criteria to the value from Guvenen et al. (2014) ($1473), which we refer to as

Figure 22: Combined persistent risk over time by gender



Note: Figure presents combined persistent risk over time for men (black, solid line) and women (red, dashed line). Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

the *low cutoff*, or increase the value of the minimum earnings criteria to the value from Juhn et al. (1993) and Autor et al. (2008) ($5893), which we refer to as the *high cutoff*. Figure 23 presents the parameters of our income process for different levels of the minimum earnings cutoff, and Figure 24 presents combined persistent income risk across different values of the minimum earnings cutoff. Putting these results together, we find that we obtain similar trends of rising persistent risk and declining temporary risk across values of the minimum earnings cutoff.

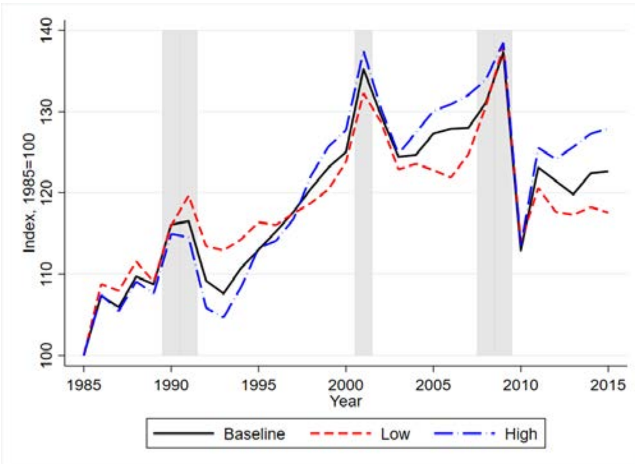
C.10 Additional Results: Sample start and end dates

In this appendix, we show that our results are robust to changing the start/end dates of our sample. In Figure 25, we present our baseline estimates of the parameters of the income process as well as estimates from the 1985-2019 time period (red, dashed line), which we refer to as "Start 1985", and from the 1981-2015 time period (blue, long dash-dotted line).[64] The figure shows that we obtain virtually identical results when we change the start/end year of the our sample.

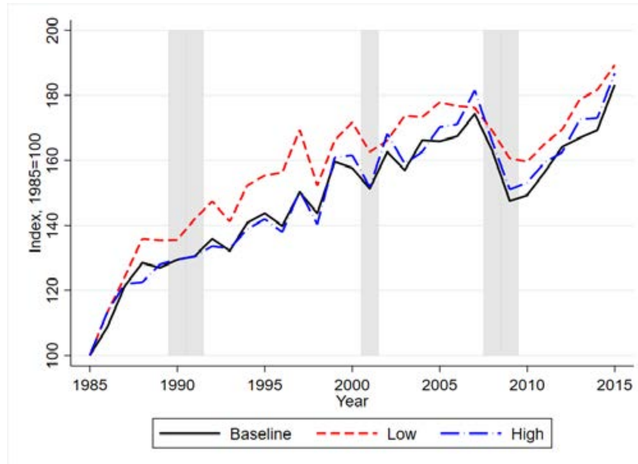
[64]Note that because of the way we bin together the first four years and last four years in creating time fixed effects the Start 1985 series is presented from 1989-2015, and similarly teh End 2015 series is presented from 1985-2011.

Figure 23: Persistent and temporary risk over time by minimum earnings cutoff

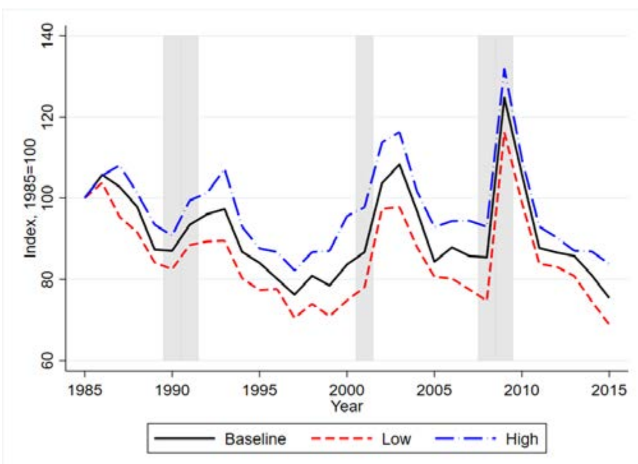
(a) Variance Persistent Shocks Employed



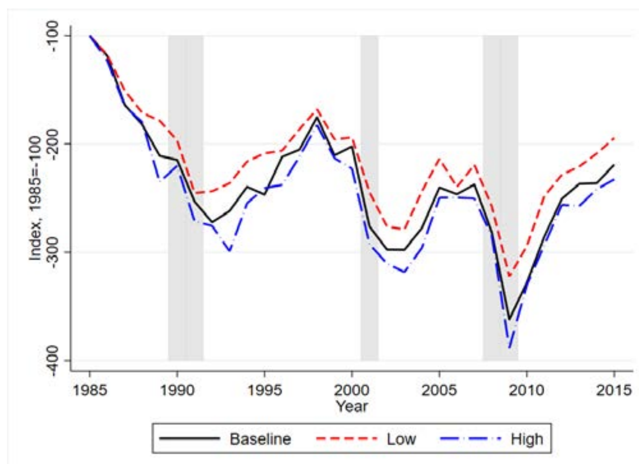
(b) Variance Persistent Shocks Unemp., First Period



(c) Variance Temporary Shocks

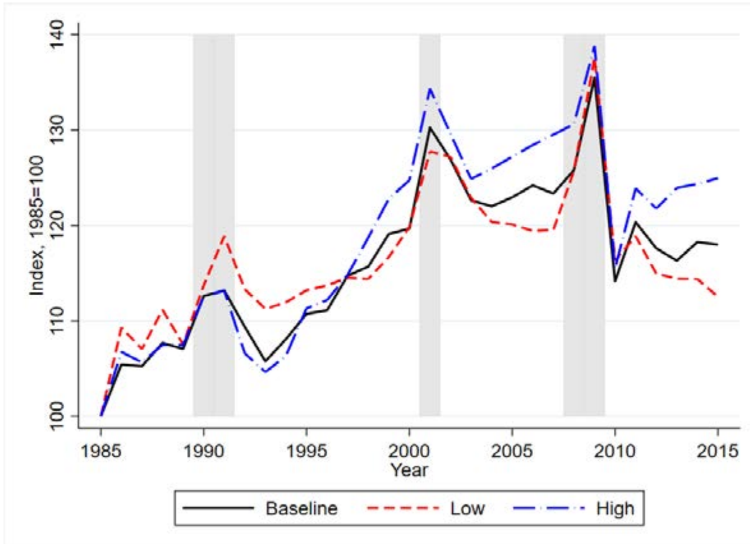


(d) Mean Persistent Shocks Unemployed, First Period



Note: The figure presents the results of estimating Model 2 across different values of the minimum earnings cutoffs. Panel (a) plots the variance of shocks to persistent earnings among the employed. Panel (b) plots the variance of shocks to persistent earnings among the unemployed, in their first period of unemployment. Panel (c) plots the variance of shocks to temporary earnings. Panel (d) plots the mean of shocks to persistent earnings among the unemployed, in their first period of unemployment. The black, solid line corresponds to baseline value of the minimum earnings cutoff ($3,350). The red, dashed line corresponds to low earnings cutoff ($1,473), and the blue, long dash-dotted line corresponds to the high earnings cutoff ($5,893). All dollars amounts are in 2005 PCE dollars.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

Figure 24: Combined persistent risk over time by minimum earnings cutoff



Note: Figure presents combined persistent risk over time for baseline minimum earnings cutoff (black, solid line), the low minimum earnings cutoff (red, dashed line), and the high minimum earnings cutoff (blue, long dashed-dotted line). See notes to Figure 23 for dollar amounts associated with each minimum earnings cutoff.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.11 Additional Results: Geographic Variation

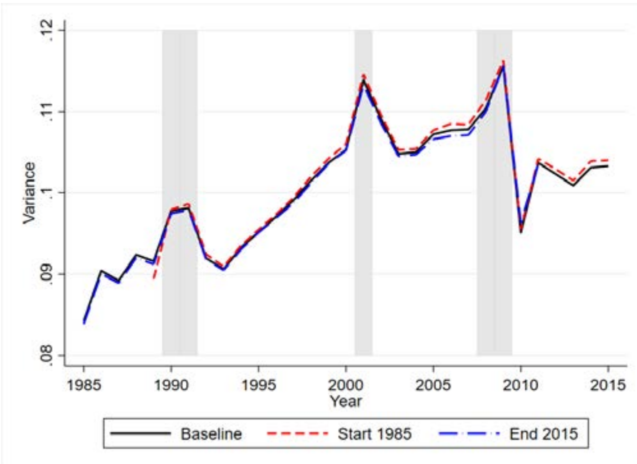
In this appendix, we test the second hypothesis that rising persistent earnings risk is related to the declines in manufacturing employment and union coverage. To do so, we start by estimating Model 3 by state. Given the large number of parameters, we use 5-year windows for the time fixed effects. We then relate changes in the parameters of the income process over time to changes in union coverage and manufacturing employment in a given state. Let X_s be the change in union membership (manufacturing employment) in state s between 1985-1989 and 2010-2015. The share of employed workers that are members of a union is measured in the CPS, while manufacturing employment is based upon Fort and Klimek (2016) industry classifications in the DER data. Let $\Delta Y_s = Y_{s,(2010-2015)} - Y_{s,(1985-1989)}$ denote the change in parameter Y (e.g. the variance of shocks to persistent earnings among employed etc.) for state s between $2010 - 2015$ and $1985 - 1989$. The specification we use is of the form,

$$\Delta Y_s = \alpha + \eta X_s + \epsilon_s \tag{59}$$

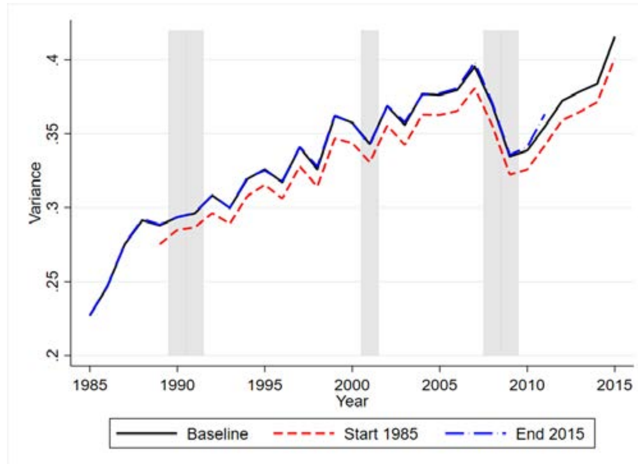
The parameter of interest is η which reports the correlation between the change in union coverage (manufacturing employment) in a state and measures of earnings risk in that state. If $\eta < 0$,

Figure 25: Persistent and temporary risk over time by sample start/end date

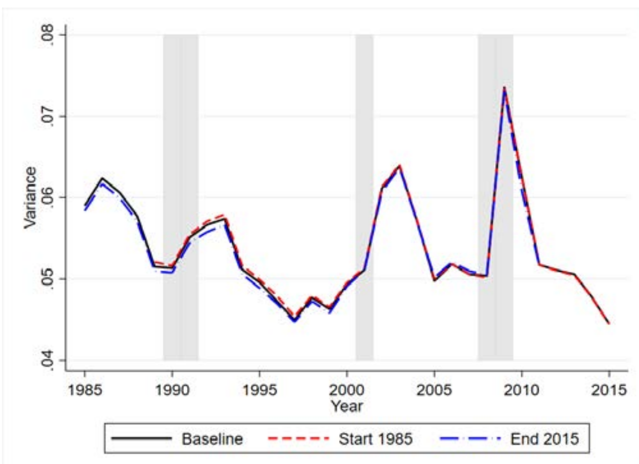
(a) Variance Persistent Shocks Employed



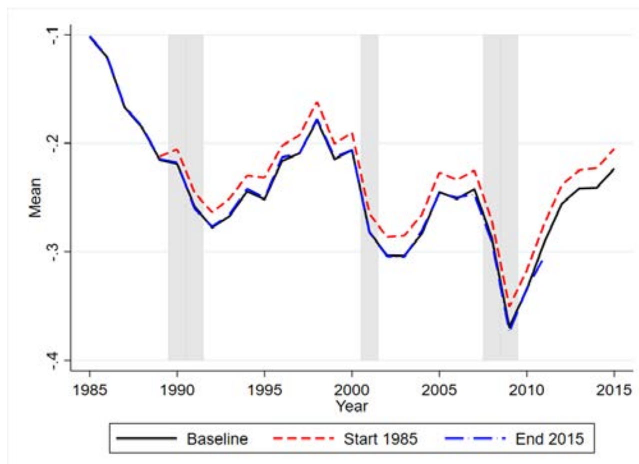
(b) Variance Persistent Shocks Unemp., First Period



(c) Variance Temporary Shocks



(d) Mean Persistent Shocks Unemployed, First Period



Note: The figure presents the results of estimating Model 2 for different starting and ending dates. Panel (a) plots the variance of shocks to persistent earnings among the employed. Panel (b) plots the variance of shocks to persistent earnings among the unemployed, in their first period of unemployment. Panel (c) plots the variance of shocks to temporary earnings. Panel (d) plots the mean of shocks to persistent earnings among the unemployed, in their first period of unemployment. The black, solid line corresponds to our baseline estimate which uses a sample from 1981-2019. The red, dashed line uses a sample from 1985-2019, while the blue, long dashed-dotted line uses a sample from 1981-2015.
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

then we have evidence that in states with larger declines in union coverage (manufacturing employment) there have been larger increases in earnings risk.

Tables 8 and 9 present the results of estimating equation (59) for changes in union coverage

Table 8: Change in union coverage and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Change Union Coverage	-0.00133	-0.00128	-0.00601	0.00332
	(0.00136)	(0.00125)	(0.00846)	(0.00762)
Round N (States)	100	100	100	100
R-squared	0.020	0.020	0.009	0.004

*Note: Table presents parameter results of estimating equation (59), where the independent variable is the change in union coverage in a state between 1985-1989 and 2010-2015 as measured in the CPS. Changes in union coverage are normalized to be mean zero and have unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.*
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

Table 9: Change in manufacturing employment and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Change in Manufacturing Emp.	-0.00265	-0.00275*	0.0144	0.00309
	(0.00173)	(0.00155)	(0.0119)	(0.00766)
Round N (States)	100	100	100	100
R-squared	0.064	0.076	0.042	0.003

*Note: Table presents parameter results of estimating equation (59), where the independent variable is the change in manufacturing employment in a state between 1985-1989 and 2010-2015. Changes in manufacturing employment are normalized to be mean zero and have unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.*
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

and manufacturing employment, respectively. The tables show that changes in union coverage and manufacturing employment are largely uncorrelated with changes in earnings risk. We view these results as providing strong evidence against the second hypothesis that the increase in persistent income risk is related to the declines in manufacturing employment and union coverage.

C.12 Additional Results: Occupations

In this appendix, we present addition results where we estimate the income process by occupation. We first present results for routine occupations, and then present results for alternative

Table 10: Routine skills and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Routine Skills	-0.00109	-0.00139	-0.0301***	0.0359***
	(0.00138)	(0.00140)	(0.0102)	(0.0106)
Round N (Occupations)	300	300	300	300
R-squared	0.002	0.005	0.027	0.058

*Note: Table presents parameter results of estimating equation (18), where the independent variable is the degree of routine skills in an occupation as measured by Acemoglu and Autor (2011). Routine skills are normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.*
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

measures of high skill occupations.

C.12.1 Routine Occupations

In this section, we examine the third hypothesis that the increase in persistent income risk is occurring in routine occupations. To do so, we split occupations by their routine task content as measured in Acemoglu and Autor (2011). Acemoglu and Autor (2011) provide a measure of the routine manual as well as routine cognitive task content of an occupation. We combine their estimates into a single measure of the routine task content of an occupation by averaging the two measures. As in Acemoglu and Autor (2011) we normalize the index to be mean zero and have unit variance. Table 10 presents the results of estimating equation (18) where the independent variable is the routine task content of an occupation. The results presented in Table 10 shows that the degree of routine task content is not correlated with changes in combined persistent risk (column (1)), or changes in persistent risk among the employed (column (2)). Column (3) shows that occupations with more routine task content have seen declines in the persistent risk from entering unemployment, which goes in the opposite direction of the aggregate trends presented in Section 3. Similarly, column (4) shows that occupations with more routine task content have seen smaller declines in persistent earnings from entering unemployment, which also goes in the opposite direction of the aggregate trends presented in Section 3. We view these results as providing strong evidence against the third hypothesis that the increase in persistent income risk is related to the declines in routine employment.

Table 11: Mean earnings and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Mean Earnings	0.00407**	0.00372**	0.0535***	-0.0149
	(0.00161)	(0.00180)	(0.00680)	(0.0119)
Round N (Occupations)	300	300	300	300
R-squared	0.051	0.053	0.139	0.016

*Note: Table presents parameter results of estimating equation (18), where the independent variable is the mean earnings in an occupation in the years 1985-1989. Mean earnings are normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.*

Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.12.2 High Skill Occupations

We next present results for two additional measures of high skill occupations.

Mean Earnings. In this section, we split occupations by their mean earnings for the $1985 - 1989$ time period. To ease the interpretation we normalize the statistic to have mean zero and standard deviation equal to one. Table 11 presents the results of estimating equation (18) where the independent variable is mean earnings in the occupation between 1985 and 1989. The results in Table 11 show that workers in higher paying occupations have experienced a larger increase in combined persistent earnings risk, as well as persistent earnings risk while employed and unemployed.

Years of Education. In this section, we split occupations by their average years of education in the $1985 - 1989$ time period. To ease the interpretation we normalize the statistic to have mean zero and standard deviation equal to one. Table 12 presents the results of estimating equation (18) where the independent variable is the mean years of completed education in the occupation between 1985 and 1989. The results in Table 12 show that workers in occupations with greater mean years of education have experienced a larger increase in combined persistent income risk, persistent earnings risk while employed and unemployed, as well as larger declines in persistent earnings while unemployed.

Table 12: Mean years of education and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(6) ΔB_U
Mean Yrs. Education	0.00540***	0.00610***	0.0735***	-0.0275**
	(0.00164)	(0.00166)	(0.00909)	(0.0118)
Round N (Occupations)	300	300	300	300
R-squared	0.067	0.104	0.194	0.041

*Note: Table presents parameter results of estimating equation (18), where the independent variable is the mean years of education in an occupation in the years 1985-1989. Mean years of education are normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.*
Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.13 Additional Results: Industry

In this appendix, we show that our results for the high skill hypothesis in Section 4.3 are robust to using detailed industry instead of detailed occupation. We repeat the analysis of Section 4.3 for industries, by splitting the sample by an individuals 4-digit industry in their first CPS year. We are able to obtain an individual's industry by using industry classification from the LBD at an individual's EIN, which we are able to observe as part of the DER. At the industry level, we are able to measure the change in mean earnings as well as mean years of education.[65]

Table 13 shows the results of estimating equation 18 using industry level variation where the independent variable is mean earnings in an industry between 1985-1989. The table shows that industries with higher initial earnings have seen larger increases in combined persistent earnings, persistent earnings risk among the employed and unemployed, as well as a larger decline in persistent earnings from entering unemployment. Table 14 shows the results of estimating equation 18 using industry level variation where the independent variable is mean years of education in an industry between 1985-1989. The table shows that industries with higher initial years of education have seen larger increases in combined persistent earnings, and persistent earnings risk among the employed and unemployed.

[65]Note that the degree of non-routine cognitive task content is measured in the ONET database, which is based upon occupations. We are unaware of any attempts to create an ONET style database by industry.

Table 13: Mean earnings by industry and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Mean Earnings	0.00373***	0.00307**	0.0267***	-0.0212**
	(0.00123)	(0.00133)	(0.00998)	(0.00898)
Round N (Industries)	300	300	300	300
R-Squared	0.036	0.029	0.031	0.035

Note: Table presents parameter results of estimating equation (18), where the independent variable is the mean years of education in an industry in the years 1985-1989. Mean years of education are normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.

Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

Table 14: Mean years of education by industry and changes in earnings risk

	(1) ΔQ	(2) ΔQ_E	(3) ΔQ_U	(4) ΔB_U
Mean Years of Education	0.00458**	0.00449*	0.0488***	-0.0149
	(0.00223)	(0.00248)	(0.00857)	(0.00910)
Round N (Industries)	300	300	300	300
R-Squared	0.070	0.079	0.134	0.022

Note: Table presents parameter results of estimating equation (18), where the independent variable is the mean years of education in an industry in the years 1985-1989. Mean years of education are normalized to be mean zero and unit standard deviation. Changes in income risk are measured between 1985-1989 and 2010-2015. Robust standard errors in parentheses. ***$p < 0.01$, **$p < 0.05$,*$p < 0.1$.

Source: 1973, 1979, 1981-1991, 1994, and 1996-2020 Current Population Survey Annual Social and Economic Supplement linked to the Detailed Earnings Record for 1981 to 2019.

C.14 Additional Details: ONET Job Zones

In this appendix, we provide additional details about ONET job zones. We first present the description of each ONET job zone that is included in the ONET package. We then present summary statistics about the mean level of earnings and education by ONET job zone as reported in the 2005 ACS.

C.14.1 Description and examples of each ONET Job Zone

Job Zone 1: Little or No Preparation Needed

- **Experience.** Little or no previous work-related skill, knowledge, or experience is needed for these occupations. For example, a person can become a waiter or waitress even if he/she has never worked before.

- **Education.** Some of these occupations may require a high school diploma or GED certificate.
- **Job Training.** Employees in these occupations need anywhere from a few days to a few months of training. Usually, an experienced worker could show you how to do the job.
- **Job Zone Examples.** These occupations involve following instructions and helping others. Examples include taxi drivers, amusement and recreation attendants, counter and rental clerks, construction laborers, continuous mining machine operators, and waiters/waitresses.

Job Zone 2: Some Preparation Needed

- **Related Experience.** Some previous work-related skill, knowledge, or experience is usually needed. For example, a teller would benefit from experience working directly with the public.
- **Education.** These occupations usually require a high school diploma.
- **Job Training.** Employees in these occupations need anywhere from a few months to one year of working with experienced employees. A recognized apprenticeship program may be associated with these occupations.
- **Job Zone Examples.** These occupations often involve using your knowledge and skills to help others. Examples include sheet metal workers, forest fire fighters, customer service representatives, physical therapist aides, salespersons (retail), and tellers.

Job Zone 3: Medium Preparation Needed

- **Related Experience.** Previous work-related skill, knowledge, or experience is required for these occupations. For example, an electrician must have completed three or four years of apprenticeship or several years of vocational training, and often must have passed a licensing exam, in order to perform the job.
- **Education.** Most occupations in this zone require training in vocational schools, related on-the-job experience, or an associate's degree.
- **Job Training.** Employees in these occupations usually need one or two years of training involving both on-the-job experience and informal training with experienced workers. A recognized apprenticeship program may be associated with these occupations.

- **Job Zone Examples.** These occupations usually involve using communication and organizational skills to coordinate, supervise, manage, or train others to accomplish goals. Examples include food service managers, electricians, agricultural technicians, legal secretaries, interviewers, and insurance sales agents.

Job Zone 4: Considerable Preparation Needed

- **Related Experience.** A considerable amount of work-related skill, knowledge, or experience is needed for these occupations. For example, an accountant must complete four years of college and work for several years in accounting to be considered qualified.

- **Education.** Most of these occupations require a four-year bachelor's degree, but some do not.

- **Job Training.** Employees in these occupations usually need several years of work-related experience, on-the-job training, and/or vocational training.

- **Job Zone Examples.** Many of these occupations involve coordinating, supervising, managing, or training others. Examples include accountants, sales managers, database administrators, teachers, chemists, environmental engineers, criminal investigators, and special agents.

Job Zone 5: Extensive Preparation Needed

- **Related Experience.** Extensive skill, knowledge, and experience are needed for these occupations. Many require more than five years of experience. For example, surgeons must complete four years of college and an additional five to seven years of specialized medical training to be able to do their job.

- **Education.** Most of these occupations require graduate school. For example, they may require a master's degree, and some require a Ph.D., M.D., or J.D. (law degree).

- **Job Training.** Employees may need some on-the-job training, but most of these occupations assume that the person will already have the required skills, knowledge, work-related experience, and/or training.

- **Job Zone Examples.** These occupations often involve coordinating, training, supervising, or managing the activities of others to accomplish goals. Very advanced communication and organizational skills are required. Examples include librarians, lawyers, aerospace engineers, wildlife biologists, school psychologists, surgeons, treasurers, and controllers.

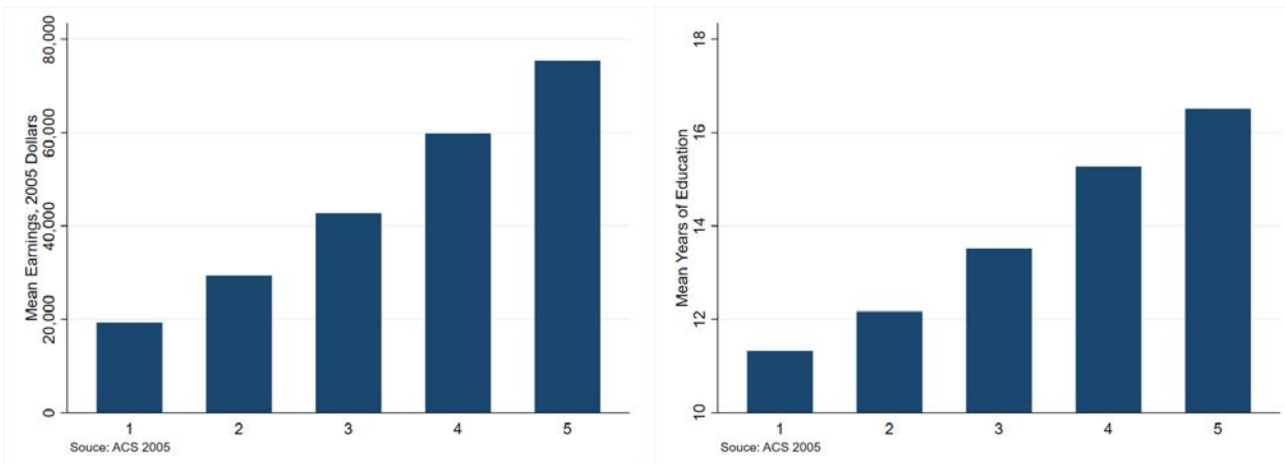
C.14.2 Summary statistics by ONET Job Zone

We next present summary statistics by ONET job zone from the 2005 ACS. Panel (a) of Figure 26 shows mean annual earnings by ONET job zone. The figure shows that the average level of earnings is steeply increasing in job zone. Panel (b) shows that mean years of education in the 2005 ACS by ONET job zone. The figure shows that the average level of education is steeply increasing by ONET job zone.

Figure 26: Summary statistics by ONET Job Zone

(a) Average earnings

(b) Average years of education



Note: Figure presents average annual earnings (panel (a) and average years of education (panel (b)) by ONET job zone.).

D Welfare effects of changing earnings risk

In this appendix, we use a finite lifecycle Bewley-Huggett-Aiyagari model to examine the welfare effects of changes in earnings volatility between the 1980s and 2010s.

D.1 Steady State Model

In this section, we introduce a steady state version of a finite lifecycle Bewley-Huggett-Aiyagari model. We assume there are $T \geq 2$ overlapping generations of agents, and let $t \in \{1, ..., T\}$ denote the age of an agent. Agents exit the model exogenously at age T, and there is no retirement.

Agents are heterogeneous along several dimensions. Let $e \in \{E, U\}$ denote the employment status of an agent, where $e = E$ denotes employed and $e = U$ denotes unemployed. Let $b \in \mathbb{R}$

denote the net asset position of an agent. When $b > 0$, the agent is saving, and when $b < 0$, the agent is borrowing. The agent's asset choice is constrained by a borrowing limit $\underline{b}$. Agents save and borrow at the risk-free rate, denoted r_f. Let $z \in \mathbb{R}$ denote an agent's persistent earnings. Let $\epsilon \in \mathbb{R}$ denote an agent's temporary shock to earnings.

Let $w_t(z, \epsilon, e)$ be a function that maps an individual's (i) age, (ii) persistent earnings, (iii) temporary shock, and (iv) employment status into a wage. We define the wage $w_t(z, \epsilon, e)$ such that

$$w_t(z, \epsilon, e) = \begin{cases} \exp(\kappa_t + z + \epsilon) & \text{if } e = E \\ \gamma \exp(\kappa_t + z) & \text{if } e = U, \end{cases}$$

where κ_t is a deterministic age profile of log earnings. $\gamma \in [0, 1]$ can be thought of as a replacement rate of persistent income for the unemployed.[66] Wages are subject to labor income taxation. Let $\widetilde{w}_t(z, \epsilon, e)$ denote the after tax income for an age t agent with persistent earnings z, temporary shock ϵ and employment status e. We model taxes following Heathcote et al. (2017), where after-tax income is given by

$$\widetilde{w}_t(z, \epsilon, e) = \lambda w_t(z, \epsilon, e)^{1-\alpha}.$$

The parameter $\alpha > 0$ governs the degree of tax progressivity.

At the start of each period, agents observe their employment status, as well as the shocks to persistent and temporary earnings. Define $y = z + \epsilon$ as residual earnings for an employed individual, and let $\delta(y, e) \in [0, 1]$ denote the probability that an agent becomes unemployed. The probability that an agent becomes unemployed depends upon their (residual) earnings and employment status from the prior period. In Section D.2, we discuss how we estimate the function $\delta(y, e)$ using prior earnings and employment status. Finally, when an agent enters into the labor market, they start as an employed agent and they draw their persistent earnings from a normal distribution with mean zero and variance Q_0.

Value Functions. We next define the value function for agents in the model. We write the value function for agents after the shocks to employment status as well as those to temporary and persistent earnings have been realized. Let $V_t(b, z, \epsilon, e)$ denote the value of being an age t agent with employment status e, persistent earnings z and temporary earnings ϵ.[67] The agent makes a consumption savings decision in the current period, taking into account the set of

[66] Alternatively, one can model home production as proportional to z, consistent with our empirical interpretation of z as capturing "income risk" among the unemployed. See Appendix A.3 for such a model.

[67] Note for unemployed the value of temporary earnings ϵ is irrelevant.

potential income shocks next period. The value function for an age t agent is given by,

$$V_t(b,z,\epsilon,e) = \max_{c,b' \geq \underline{b}} u(c) + \beta \mathbb{E}_{z',\epsilon',e'}\left[V_{t+1}(b',z',\epsilon',e')\right] \quad \forall t \leq T$$
$$V_{T+1}(b,z,\epsilon,e) = 0,$$

subject to the budget constraint,

$$c + b' \leq b(1+r_f) + \widetilde{w}_t(z,\epsilon,e);$$

the law of motion for employment status,

$$e' = \begin{cases} E & \text{w. prob } 1-\delta(y,e) \\ U & \text{w. prob } \delta(y,e); \end{cases}$$

and the law of motion for persistent earnings,

$$z' = \begin{cases} Fz + \nu_{E,t+1} & \text{if } e' = E \,\&\, e = E \\ Fz + \nu_{U,t+1} & \text{if } e' = U \,\&\, e = E \\ Fz + \nu_{N,t+1} & \text{if } e' = U \,\&\, e = U, \end{cases}$$

where $\nu_{e,t+1} \sim N(B_{e,t+1}, Q_{e,t+1})$.[68] Note that the mean and variance to the shock depends on the agent's employment status and age. Finally, the law of motion for temporary earnings is given by,

$$\epsilon' = \begin{cases} \epsilon_{t+1} & \text{if } e' = E \\ 0 & \text{if } e' = U, \end{cases}$$

where $\epsilon_{t+1} \sim N(0, R_{t+1})$. The variance of the shock to temporary earnings depends on the agent's age $t+1$.

D.2 Estimation

We next discuss the estimation of the model.[69] Some parameters are assigned using estimates from the literature, while others are calibrated to be consistent with the U.S. labor market in the

[68] As in Section C.6, we allow the shocks an individual draws while unemployed to depend on whether they are in their first period of unemployed ($e' = U \,\&\, e = E$) or after their first year of unemployment ($e' = U \,\&\, e = U$).

[69] We solve the model using VFI on discrete grids. In Appendix D.4, we discuss how we discretize the income process.

1980s.

Demographics and preferences. To align with the sample in Section 2.1, agents enter the model at age 25 ($t = 1$) , and work until age 60 ($T = 36$). When agents enter the model, they begin with zero assets and are employed.

Agents receive utility from consumption, with preferences given by,

$$u(c) = \frac{c^{1-\sigma}}{1-\sigma}.$$

We set the risk aversion parameter to a standard value, $\sigma = 2$. Agents discount the future at rate $\beta = 0.964$. The parameter β is calibrated to match the average ratio of net worth to income. As in Kaplan and Violante (2010), we target a value of 2.5.

Income process. Agents receive wages that are a function of their age, persistent earnings, and temporary earnings. The fixed age component of earnings is estimated as part of the residualization process in Section 2.2. Panel (a) of Figure 27 plots the deterministic path of earnings that is used in the model.

When agents are unemployed they receive (pre-tax and transfers) a fraction $\gamma \in [0, 1]$ of their persistent earnings. The parameter γ can be thought as the replacement rate of unemployment insurance. We set $\gamma = 0.4$, as in Shimer (2005). We next discuss the estimation of the stochastic process that governs how earnings evolve in the model.

Shocks to income. We model shocks to labor income as a function of an individual's age as in Appendix C.6. In our baseline estimation, we use the shocks to labor labor income that correspond to the 1980s. In the welfare experiment in Section D.3, we sequentially adjust the parameters of the income process to their 2010 values. Panels (d)-(h) of Figure 27 present the age profiles of shocks to temporary and persistent earnings that are used in the quantitative model.

Probability of unemployment. We model the likelihood that an employed individuals becomes unemployed as a non-linear function of their prior earnings as in Section 5.1. In particular, we use the parameter estimates from estimating equation 19 to specify the following non-linear probability of being unemployed based upon prior earnings,

$$\delta(y, E) = \alpha_E + \mathbb{I}\{y \geq 0\} \left[\sum_{k=0}^{2} \alpha_{k,E}^{+} y^k \right] + \mathbb{I}\{y < 0\} \left[\sum_{k=0}^{2} \alpha_{k,E}^{-} y^k \right] \tag{60}$$

To be consistent with the labor market in the 1980s, we include a constant term α_E in equation 60. We calibrate the constant (α_E) to match the probability that an individual transitioned from employment to unemployment in the 1980s in our sample, which we find to be 3.9%. Panel (c) of Figure 27 presents the profile for the probability of entering into unemployment as a function of prior earnings. We set the likelihood that an unemployed individual remains unemployed to be a constant ($\delta(y, U)) = \alpha_U$) equal to 56.3%.

Taxes. We model taxes as in Heathcote et al. (2017). As in Heathcote et al. (2017), we set the tax progressivity (α) parameters to be equal to 0.181. In addition to financing the UI system, we model the government as having exogenous expenditures G that are equal to share $g \in [0, 1]$ of before-tax labor income. Using NIPA data on personal income and government consumption expenditure and investment, we set $g = 0.260$. We set the level parameter (λ) so that government revenue from taxes is equated to government spending on transfers and the exogenous government spending. Panel (b) of Figure 27 presents the implied tax function in the model economy. Agents with pre-tax incomes below approximately \$10K receive transfers from the government, while individuals with pre-tax incomes greater than \$10K pay labor income taxes.

Asset Markets. Agents are able to save and borrow at the risk-free rate of 4%. We set the borrowing limit $\underline{b}$ to the natural borrowing limit.[70] Setting the borrowing limit at the natural borrowing limit represents an upper bound to the extent to which agents can use borrowing to smooth shocks to income.

Table 15 and Figure 27 present the parameters that govern model economy. In the next section, we conduct the welfare experiment of adjusting labor income risk as documented in Section 3.

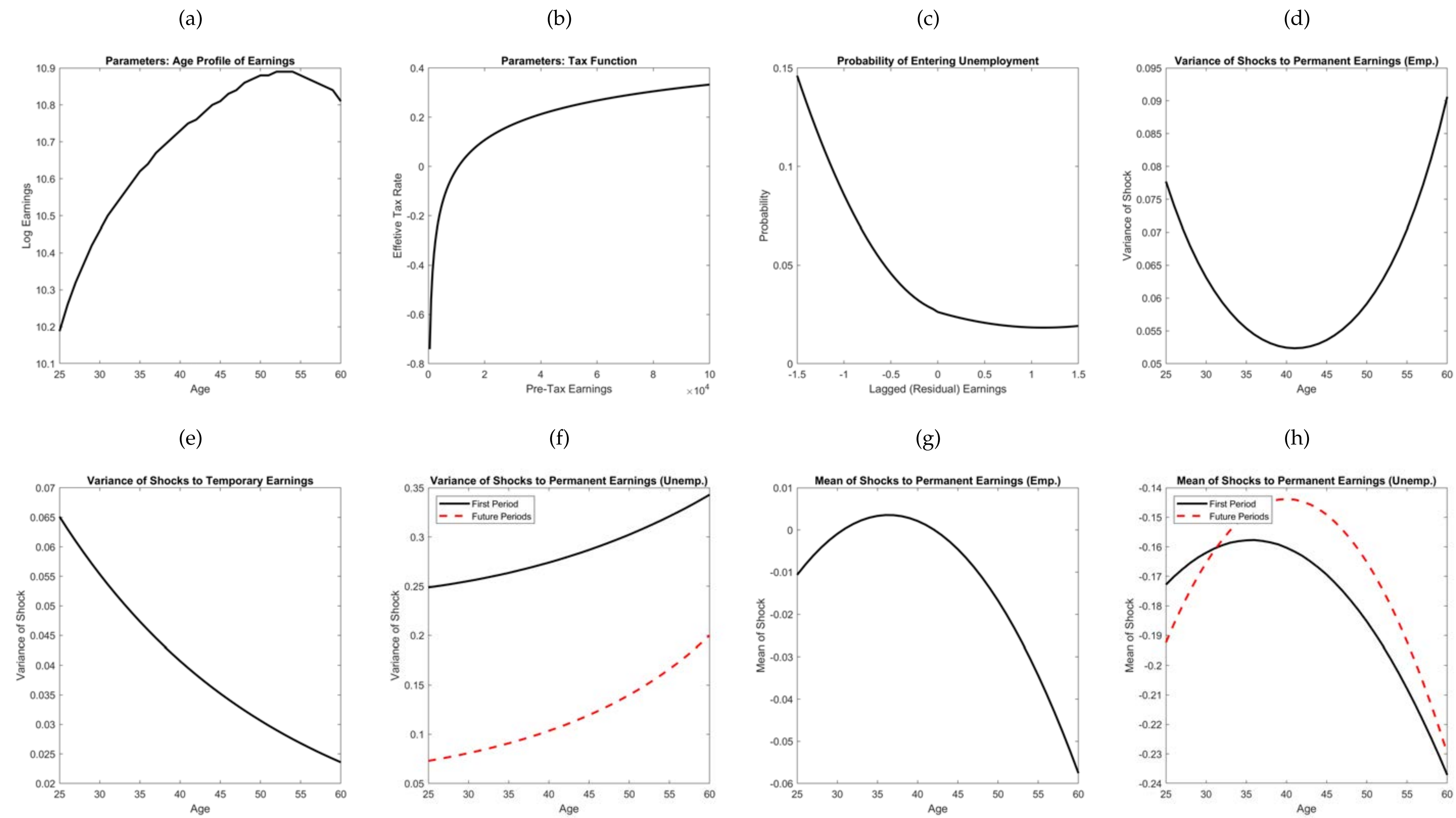
[70]In our model, implementing the natural borrowing limit is equivalent to requiring that individuals die with zero debt.

Table 15: Model parameters

Variable	Value	Description
β	0.964	Discount factor
r_f	4%	Risk-free interest rate
σ	2	Coefficient of relative risk-aversion
α	0.181	Progessivity of tax function
γ	0.4	Replacement Rate UI
g	0.260	Ratio of government expenditure to pre-tax income
α_E	0.0125	Constant in unemployment probability
$\alpha^+_{0,E}$	0.0138	Constant in unemployment probability, positive prior earnings
$\alpha^+_{1,E}$	-0.0141	Linear term in unemployment probability, positive prior earnings
$\alpha^+_{2,E}$	0.0062	Square term in unemployment probability, positive prior earnings
$\alpha^-_{0,E}$	0.0143	Constant in unemployment probability, negative prior earnings
$\alpha^-_{1,E}$	-0.0173	Linear term in unemployment probability, negative prior earnings
$\alpha^-_{2,E}$	0.0415	Square term in unemployment probability, negative prior earnings
α_U	0.563	Probability unemployed remain unemployed

Note: Table presents model parameters for the baseline estimation of the quantitative model.

Figure 27: Model parameters



Note: Figure plots parameters used in the baseline estimation of the quantitative model.

D.3 Welfare implications of changing earnings risk

In this section, we examine the welfare implications of rising persistent income risk. We measure the welfare effects of changing income risk using lifetime consumption equivalents behind the veil-of-ignorance (i.e., before initial draws of persistent earnings are drawn). In Section 3, we documented three facts about the changing nature of income persistent risk: (1) persistent earnings risk among the employed increased, (2) persistent earnings risk among the unemployed increased, (3) the scaring effect (decline in persistent earnings) from entering unemployment accelerated. We also showed that the rate of entry into unemployment declined and that temporary earnings risk declined. We model these changes in our quantitative model by using the estimates for the 2010s from the income process in Appendix C.6 for Q_E, Q_U, B_U and R.[71] To match the change in the rate of entry into unemployment, we adjust α_E in the unemployment transition probability (equation 60) to match the EU rate from 2010-2015, which we measure to be 2.5%. Column (2) of Table 16 presents the results of modeling these changes in the quantitative model. We find that the increase in persistent earnings risk has reduced welfare by 5.1% of lifetime consumption as individuals increase their precautionary savings in response to the increase in risk.[72]

Two factors mitigated welfare losses from rising persistent earnings risk since the 1980s: (1) declining temporary risk, and (2) declining rates of entry into unemployment. We sequentially remove these mitigating factors and examine the implications of rising persistent risk. Column (2) of Table 16 shows the effects of rising persistent risk without the decline in temporary risk. We find that keeping temporary earnings risk at its 1980s value had a very modest role in mitigating the effects of rising persistent risk. If temporary earnings risk has stayed at its 1980s levels, then the welfare losses from the rise in persistent earnings risk would have been 5.25% of lifetime consumption.

Finally, we examine the role of reducing the entry into unemployment. Column (4) of Table 16 shows the effects of keeping the entry rate into unemployment at its 1980s level. We find that if the entry rate into unemployment had remained at its 1980s level, the rise in persistent earnings risk would have caused nearly an 8.5% welfare loss. Hence, the decline in entry into unemployment played a substantial role in mitigating the impact of rising persistent earnings risk within employment/unemployment spells.

[71] Given that we work with the log of earnings, changes in risk can change mean earnings due to Jensen's inequality. To highlight the role of risk, we adjust the age earnings profile κ so that mean earnings are held constant in the welfare experiment.

[72] In results that are available upon request, we find that the main driver of the welfare loss is the rise in persistent earnings risk among the employed Q_E.

Table 16: Welfare experiment: changes in earnings risk

	(1) Baseline	(2) $Q_E, Q_U, B_U, R, \alpha_E$	(3) Q_E, Q_U, B_U, α_E	(4) Q_E, Q_U, B_U, R
Welfare chg. from baseline	-	-5.0964%	-5.2437%	-8.4839%
R	0.0412	0.0316	0.0412	0.0316
Q_E	0.0649	0.0774	0.0774	0.0775
Q_U	0.2937	0.5357	0.5358	0.5348
B_U	-0.1804	-0.3555	-0.3545	-0.3539
Q	0.0796	0.0949	0.095	0.1062
EU Rate	0.0391	0.0252	0.0254	0.042

Note: Table presents the results of the welfare experiment in Section D.3. Welfare is measured as a percent of lifetime consumption.

D.4 Computational Details

In this appendix, we discuss how we solve the lifecycle Bewley model. We solve the model using value function iteration on grids. Below we discuss the process for discretizing income shocks.

D.4.1 Discretization Process (Persistent Earnings)

In this section, we outline our process for discretizing shocks to persistent earnings. For ease of presentation, we abstract from allowing the drift and variance of shocks to vary by age. For an income process that allows the mean and variance of shocks to vary by age, simply repeat these steps for each age level.

At the start of the period an agent draws whether or not they will be employed for the period. Recall that the process for persistent earnings is given by:

$$z^{'} = \rho z + B_e + \nu_e$$

where $e \in \{E, U\}$ denotes employment status, B_e denotes the drift of persistent earnings while in employment status e, and ν_e is the shock to persistent earnings while in employment status e. We assume that the drifts to persistent earnings and the variance of the shocks to persistent earnings differ by employment status. That is $\nu_U \sim N(0, Q_U)$, and $\nu_E \sim N(0, Q_E)$. Finally, the parameter ρ governs the degree of persistence in the process.

Define a transition matrix for agents classified as employed, denoted π^E, and a transition matrix for agents classified as unemployed, denoted π^U. The elements of π^e_{jk} defines the probability that an agent with employment status e, moves from state j **today** to state k **tomorrow**.

Assume for now that we have specified a grid of values for z with N grid points, which are given by $[z_1, z_2, ..., z_N]$. Let the points be evenly spaced, with distance between grid points denoted by d.[73] The transition probability of going from state j **today** to state k **tomorrow** for an individual with employment status e is given by

$$\begin{aligned}\pi^e_{jk} &= P(\widetilde{z}_t = z_k | \widetilde{z}_{t-1} = z_j | e) \\ &= P(z_k - \frac{d}{2} < \rho z_j + B_e + \nu_e < z_k + \frac{d}{2}) \\ &= P(z_k - \frac{d}{2} - \rho z_j - B_e < \nu_e < z_k + \frac{d}{2} - \rho z_j - B_e) \end{aligned} \tag{61}$$

For an interior point on the grid, the probability in equation (61) is given by:

$$\pi^e_{jk} = F\left(\frac{z_k + \frac{d}{2} - \rho z_j - B_e}{\sigma_{\nu,e}}\right) - F\left(\frac{z_k - \frac{d}{2} - \rho z_j - B_e}{Q_e^{1/2}}\right)$$

where $F(\cdot)$ is the standard normal distribution. For the end points of the grid, define the probabilities using:

$$\begin{aligned}\pi^e_{j1} &= F\left(\frac{z_1 + \frac{d}{2} - \rho z_j - B_e}{Q_e^{1/2}}\right) \\ \pi^e_{jN} &= F\left(\frac{z_N - \frac{d}{2} - \rho z_j - B_e}{Q_e^{1/2}}\right)\end{aligned}$$

Discretization Process (Temporary Earnings). To discretize the process for temporary earnings, we use Tauchen's method with the persistence of the shock set to zero.

D.5 Additional quantitative results: Higher order moments

In this section, we briefly discuss estimates of the higher order moments from the income process using in our quantitative model.

Guvenen et al. (2021), hereafter referred to as GKOS, showed that labor income changes exhibit substantial deviations from a normal distribution and that the scope of these deviations varies by an individual's prior earnings, which they refer to as recent earnings. In estimating the higher order moments from the simulated data of our income process we closely follow the

[73] In practice, we define the endpoints of the grid using $z_N = m\left(\frac{Q_U}{1-\rho}\right)^{\frac{1}{2}}$, setting $m = 3$, and $z_1 = -z_N$.

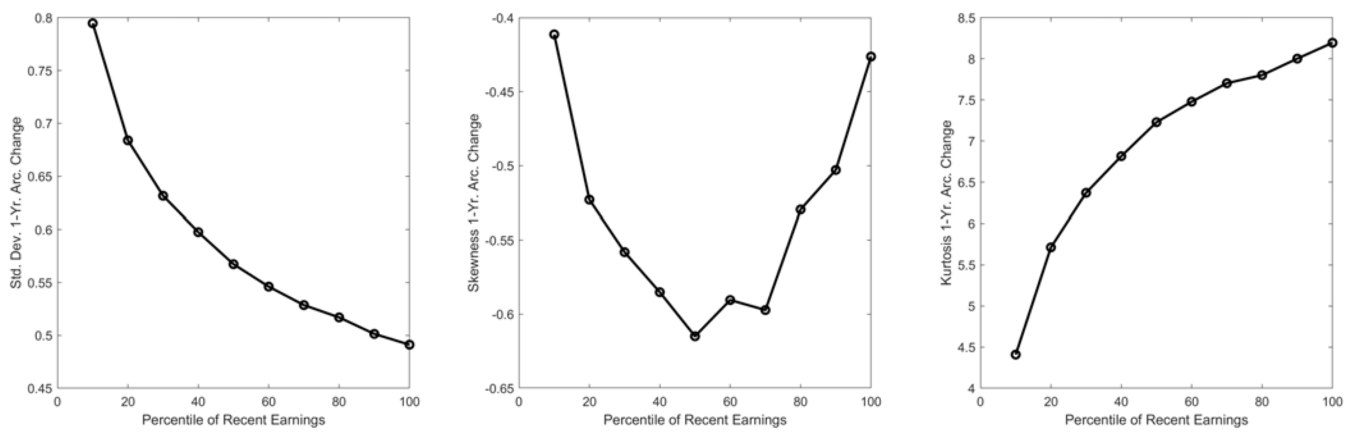
setup of GKOS. First for each simulated individual, we measure their recent earnings, which is the sum of their earnings over the prior 5 years and then remove the age specific component. To align with GKOS, in measuring recent earnings in a year $t-1$ we require that the individual have earnings above the minimum cutoff in year $t-1$ as well as in at least two of the years between $t-2$ and $t-5$. We then measure moments of the distribution of changes in earnings between t and $t+1$ by decile of recent earnings. To align with the estimated income process in GKOS, we use arc percent changes in earnings.

Before discussing the trends in higher order moments for our income process, we briefly review the findings of GKOS. They find that the standard deviation of 1 year earnings changes decreases in prior earnings up to the 80^{th} percentile of recent earnings distribution and increases slightly with recent earnings at the top of the distribution. GKOS find that labor income shocks are negatively skewed, and as recent earnings increase the shocks become more negatively skewed up to the 80th percentile of recent earnings and then exhibit an uptick at the top of the distribution. Finally, GKOS show that labor income shocks have excess kurtosis and that kurtosis increases up to the 80^{th} percentile of recent earnings and then declines at the top of the distribution.

In Figure 28 below, we present estimates of the 2^{nd}, 3^{rd}, and 4^{th} moments of the 1-year arc change in labor earnings and find that the higher order moments of our income process qualitatively align with the pattern documented by GKOS. The far left panel of the figure show the standard deviation of 1-year arc earnings changes by decile of recent earnings. In our simulated data, the standard deviation of earnings changes decreases with recent earnings. The middle panel presents the skewness of earnings changes in our simulated data. Our income process generates negative skewness and generates a U-shaped pattern of negative skewness, with the greatest negative skewness at the median of the distribution. Finally, the right panel plots the kurtosis of the simulated earnings changes and shows that in the simulated data kurtosis increases with recent earnings.

Figure 28: Higher order moments for baseline income process

(a) Std. dev. by recent earnings (b) Skewness by recent earnings (c) Kurtosis by recent earnings



Note: Panel (a) plots the standard deviation of log residualized arc income changes from the model simulation based on the income process from Section D. Panels (b) and (c) report the skewness and kurtosis of model simulated log residualized arc income changes, respectively.