What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Dynamics?*

Fatih Guvenen Fatih Karahan Serdar Ozkan Jae Song

February 12, 2019

Abstract

We study individual earnings dynamics over the life cycle using panel data on millions of U.S. workers. Using nonparametric methods, we first show that the distribution of earnings changes exhibits substantial deviations from lognormality, such as negative skewness and very high kurtosis. Further, the extent of these non-normalities varies significantly with age and earnings level, peaking around age 50 and between the 70th and 90th percentiles of the earnings distribution. Second, we estimate nonparametric impulse response functions and find important asymmetries: positive changes for high-income individuals are quite transitory, whereas negative ones are very persistent; the opposite is true for low-income individuals. Third, we turn to long-run outcomes and find substantial heterogeneity in the cumulative growth rates of earnings and total years individuals spend nonemployed between ages 25 and 55. Finally, by targeting these rich sets of moments, we estimate stochastic processes for earnings, that range from the simple to the complex. Our preferred specification features normal mixture innovations to both persistent and transitory components, and includes long-term nonemployment shocks with a realization probability that varies with age and earnings.

JEL Codes: E24, J24, J31.
Keywords: Earnings dynamics, higher-order earnings risk, kurtosis, skewness, non-Gaussian shocks, normal mixture.

*Guvenen: University of Minnesota, FRB of Minneapolis, and NBER; guvenen@umn.edu; Karahan: FRB of New York; fatih.karahan@ny.frb.org; Ozkan: University of Toronto; serdar.ozkan@utoronto.ca; Song: Social Security Administration; jae.song@ssa.gov. Special thanks to Gerald Ray at the Social Security Administration for his help and support. For helpful critiques and comments, we thank participants at the 2012 CASEE Conference on Quantitative Macroeconomics, 2013 and 2014 AEA Winter Meetings, 2013 RRC Conference, 2013 JDP Dilemmas on Inequality Lecture at Princeton, 2014 Sverige Riksbank Conference on Inequality, 2014 NBER Summer Institute, 2015 NBER *EF&G* Meetings, as well as at research seminars at numerous universities and institutions. Special thanks to Joe Altonji, Andy Atkeson, Mark Bils, Richard Blundell, Chris Busch, Bob Hall, Michael Keane, Giuseppe Moscarini, Luigi Pistaferri, Fabien Postel-Vinay, Anthony Smith, Kjetil Storesletten, and Gianluca Violante for comments. The views expressed herein are those of the authors and do not represent those of the Social Security Administration or the Federal Reserve Banks of Minneapolis and New York. Guvenen acknowledges financial support from the National Science Foundation (SES–1357874).

1 Introduction

The goal of this paper is to characterize the most salient properties of individual earnings dynamics over the life cycle, focusing on *nonnormalities* and *nonlinearities*. First, by studying its higher-order moments (specifically, skewness and kurtosis), we investigate the distribution of earnings changes, and whether it can be well approximated by a normal distribution. Second, we explore mean reversion patterns of earnings changes that may differ between positive and negative changes as well as by size. Finally, we study how these properties vary over the life cycle and across the earnings distribution.

The extent and nature of these nonnormalities and nonlinearities are difficult to predict beforehand, and strong parametric assumptions can mask those features, making it difficult to uncover them. With these considerations in mind, we first employ a fully nonparametric approach and take "high-resolution pictures" of individuals' earnings histories. To this end, we use administrative panel data from the U.S. Social Security Administration (SSA) covering a long time span from 1978 to 2013, with a substantial sample size (10% random sample of males aged 25–60).[1] Next, using the facts uncovered in this descriptive analysis, we estimate nonlinear and non-Gaussian earnings processes.

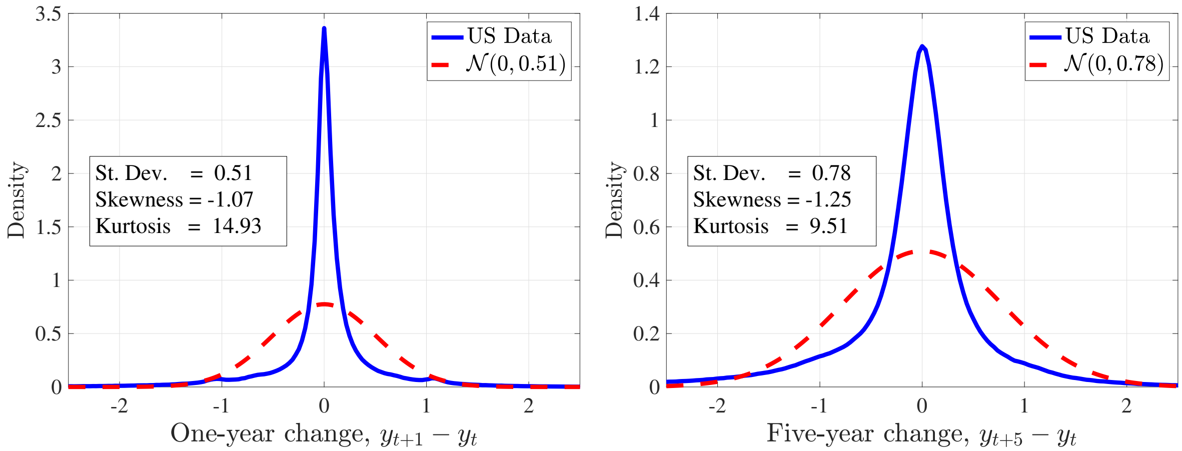
Our descriptive analysis covers (i) the properties of the distributions of earnings changes, (ii) the extent of mean reversion during the 10 years following earnings changes, and (iii) workers' long-term outcomes covering their entire working lives such as cumulative earnings growth and the incidence of nonemployment.

Starting with the distribution of earnings changes, we find that it is left- (negatively) skewed, and this left-skewness becomes more severe as individuals get older, or their earnings increase (or both). For example, workers aged 45–55 and earning about $100,000 per year (in 2010 dollars) face a five-year log earnings change distribution with a lower tail (the gap between the 50th and 10th percentiles) 2.5 times longer than the upper tail (50th to 90th percentiles). In contrast, low-income young workers face an almost symmetric distribution. The rise in left-skewness over the life cycle is entirely due to a reduction in opportunities for large gains from ages 25 to 45 and to the increasing likelihood of a sharp fall in earnings after age 45.

In addition, earnings growth displays a very high kurtosis relative to a Gaussian density (Figure 1). There are far more people in the data with very small and extreme

[1]In this paper, we focus on men for comparability with earlier work. Our analysis for women found qualitatively similar patterns. These results are reported in an online appendix on the authors' websites.

FIGURE 1 – Histograms of One- and Five-Year Log Earnings Changes



Notes: This figure plots the empirical densities of one- and five-year earnings changes superimposed on Gaussian densities with the same standard deviation. The data are for all workers in the base sample defined in Section 2 and $t = 1997$.

earnings changes and fewer people with middling ones. For example, 31% of the annual earnings changes are less than 5%, compared to only 8% under the Gaussian distribution. Also, a typical worker sees a change larger than three standard deviations with a 2.4% chance, which is about one-ninth as likely under a normal distribution. Importantly, the *average* kurtosis masks significant heterogeneity. For example, five-year earnings growth of males aged 45–55 and earning \$100,000 has a kurtosis of 18, compared to 5 for younger workers earning \$10,000 (and 3 for a Gaussian distribution).

What are the sources of nonnormalities? While a full-blown investigation is beyond the scope of this paper, we shed some light on the question by using data from the Panel Study of Income Dynamics (PSID), which contains information not available in the SSA data. We find that hourly wage changes exhibit little left-skewness but an excess kurtosis with a magnitude and life-cycle variation similar to earnings changes. Furthermore, we find that wage changes are at least as important as hours changes, even in the tails of the distribution. For example, a group of workers with an earnings decline of around 165 log points experience an average wage decline of 101 log points. Furthermore, workers experiencing these extreme changes are likely to have gone through nonemployment, job or occupation changes, or experienced health shocks, suggesting that the tails are not purely a statistical artifact or measurement error in survey data.

Next, we characterize the mean reversion patterns of earnings changes by estimating nonparametric impulse response functions conditional on recent earnings and on the size and sign of the change. We find two types of asymmetry. First, fixing the size of the change, positive changes to high-earnings individuals are quite transitory while negative ones are persistent, in contrast, the opposite is true for low-earnings individuals. Second, with a fixed level of earnings, the strength of mean reversion differs by the size of the change: large changes tend to be much more transitory than small ones. These asymmetries are difficult to detect in a covariance matrix, in which all sorts of earnings changes—large, small, positive, and negative—are masked by a single statistic.

Finally, we document two facts regarding long-term outcomes covering individuals' entire working lives. First, the cumulative earnings growth over the life cycle varies systematically and substantially across groups of workers with different lifetime earnings. For example, average earnings rise by 60% from age 25 to age 55 for the median lifetime earnings group, by 4.8-fold for the 95th percentile, and by 27.8-fold for the top 1 percent groups.[2] Second, there is substantial variation in individuals' lifetime nonemployment rate—which we define as the fraction of lifetime (ages 25 to 60) spent as (full-year) nonemployed. For example, 40% of men experience at most one year of nonemployment, while 18% spend *more than half* of their working years as nonemployed. These numbers imply an extremely high persistence in the long-term nonemployment state.

While the nonparametric approach allows us to establish key features of earnings dynamics in a transparent way, a tractable parametric process is indispensable because (i) it allows us to connect earnings *changes* to underlying *innovations* or *shocks* to earnings, and (ii) it can be used as an input to calibrate quantitative models with idiosyncratic risk.[3] Therefore, in Section 6, we target the empirical moments described above to estimate a range of income processes. We start with the familiar linear-Gaussian framework and build on it incrementally until we arrive at a rich, yet tractable benchmark specification that can capture the key features of the data. Along the way, we discuss which aspect of the data each feature helps capture so that researchers can judge the trade-offs between matching a particular moment and the additional complexity it brings.

Our preferred benchmark process extends the linear-Gaussian framework to contain (i) an AR(1) process with innovations from a mixture of normals; (ii) an i.i.d. normal

[2]A positive relationship between lifetime earnings and life-cycle earnings growth is to be expected. What is surprising is that the magnitudes are so large that they cannot be explained by simple processes.

[3]Although we follow the common practice in the literature of referring to innovations as income "shocks," individuals are likely to have more information about them than what we—as econometricians—can identify from earnings data alone. Separating expected from unexpected changes requires either survey data on expectations (e.g., Pistaferri (2003)) or economic choices (e.g., Cunha *et al.* (2005) and Guvenen and Smith (2014)). Tackling this important question is beyond the scope of this paper.

mixture transitory shock; and (iii) heterogeneous income profiles. More importantly, we estimate (iv) a long-term nonemployment shock with a realization probability that depends on age and earnings, which helps capture life-cycle and income variation of the moments. This *heterogeneous* nonemployment risk generates recurring nonemployment with long-term scarring effects concentrated among young and low-income individuals. Our empirical facts require non-Gaussian features in *persistent* innovations; these can be achieved by such nonemployment shocks or non-Gaussian shocks to the persistent component, but not by a uniform nonemployment risk that is transitory in nature.[4]

Related Literature. The earnings dynamics literature has a long history, dating back to seminal papers by Lillard and Willis (1978), Lillard and Weiss (1979), and MaCurdy (1982). Until recently, this literature focused on linear ARMA-type time series models identified from the variance-covariance matrix, thereby abstracting away from nonlinearities and nonnormalities of the data.[5]

In an important paper, Geweke and Keane (2000) modeled earnings innovations using normal mixture distributions and found important deviations from normality.[6] More recently, using earnings data from France, Bonhomme and Robin (2009) estimate a flexible copula model for the dependence patterns over time and a mixture of normals for the transitory component that displays excess kurtosis. Bonhomme and Robin (2010) use a nonparametric deconvolution method and find excess kurtosis in permanent and transitory shocks. We go beyond the overall distribution and document that nonnormalities in transitory and persistent components vary substantially with earnings levels and age. Another related paper is Guvenen *et al.* (2014b), which shows that earnings growth becomes more left skewed in recessions, however, it abstracts away from life-cycle variation. We also go further by analyzing kurtosis and how it varies across the income distribution, asymmetries in mean reversion, and the heterogeneity in life-cycle income growth rates and lifetime nonemployment rates, which are all absent from that paper.

---

[4] A transitory nonemployment shock (or any transitory shock for that matter) cannot generate left skewness in earnings growth, because each nonemployment spell contributes same-sized negative and positive earnings change observation (one when the worker becomes nonemployed and one when he returns to work), consequently, stretching both tails of the distribution, leaving the symmetry unaffected.

[5] Quantitative macroeconomists use income processes usually by discretizing them using Tauchen or Rowenhorst methods. Typical implementations of the Tauchen method assume the shocks are normally distributed. Rowenhorst method does not rely on this assumption, but it focuses on matching the first two moments of earnings levels and ignores the mechanical implications for the distribution of changes.

[6] In an even earlier contribution, Horowitz and Markatou (1996) showed how an income dynamics model can be estimated nonparametrically and found evidence of nonnormality in the error components.

In contemporaneous work, Arellano *et al.* (2017) explore nonlinear earnings dynamics. They propose a new approach based on estimating the conditional quantiles of earnings, which allows the persistence of earnings to vary with the size and sign of the shock. They find asymmetries in mean reversion and non-Gaussian features that are consistent with our results. They also show that the consumption response to earnings shocks displays nonlinearities, which we do not study. Relative to that paper, we provide a more in-depth analysis of the conditional skewness and kurtosis of earnings, document how they differ between job-stayers and switchers, examine systematic variation in lifecycle earnings profiles and lifetime employment rates. Overall, the two papers complement each other.

Finally, our work is also related to Altonji *et al.* (2013), who estimate a joint process for earnings, wages, hours, and job changes, targeting a rich set of moments via indirect inference. Browning *et al.* (2010) also employ indirect inference to estimate a process featuring "lots of heterogeneity." However, neither paper explicitly focuses on higher-order moments, their evolution over the life cycle, or asymmetries in mean reversion.

2 Data and Variable Construction

2.1 The Data Set

We draw a representative 10% panel sample of the U.S. population from the Master Earnings File (MEF) of the SSA. The MEF combines various datasets that go back as far as 1978. For our purposes, the most important variables include labor income from W-2 forms (for each job held by the employee during the year), self-employment income (obtained from the Internal Revenue Service (IRS) tax form Schedule SE), and various demographics (date of birth, sex, and race).[7] We focus on *total annual labor earnings*, which is the sum of total annual wage income plus the labor portion (2/3) of self-employment income.

Wage income is not top coded throughout our sample, whereas self-employment income was capped at the SSA taxable limit until 1994. Although this top coding affects only a small number of individuals who make substantial income from self-employment, we restrict our analysis to the 1994–2013 period to ensure that our analysis of higher-order moments is not affected by this issue. The only exception is our use of the entire

[7]The measure of earnings on the W-2 form (Box 1) includes all wages and salaries, tips, restricted stock grants, exercised stock options, severance payments, and other types of income considered remuneration for labor services by the IRS. It does not include any pre-tax payments to individual retirement accounts (IRAs), retirement annuities, child care expense accounts, or other deferred compensation.

1978 to 2013 period in Section 5, where we analyze long-term outcomes of workers for which a longer time series is essential, such as the cumulative income growth over the life cycle. For robustness, we impute self-employment income above the cap for the years before 1994 using quantile regressions. Only a small number of individuals are affected by this imputation, so the effect on our results is minimal. The details are provided in Appendix A.1. Finally, we convert nominal values to real values using the personal consumption expenditure (PCE) deflator, taking 2010 as the base year (see Appendix A for further details of the construction of our sample and variables).

Despite the advantages noted, the dataset also has some important drawbacks, such as limited demographic information, the absence of capital income, and the lack of hours (and thus hourly wage) data. To overcome some of these limitations, we supplement our analysis with survey data whenever possible. Another important limitation is the lack of household-level data. Even though a large share of quantitative models focus on individual earnings fluctuations (which we study), household earnings dynamics are key for some economic questions, for which we have little to say about in this paper.[8]

2.2 Sample Selection

Our *base sample* is a revolving panel consisting of males with some labor market attachment that is designed to maximize the sample size (important for precise computation of higher-order moments in finely defined groups) and keep the age structure stable over time. First, in order for an individual-year income observation to be *admissible* to the *base sample*, the individual (i) must be between 25 and 60 years old (the working lifespan) and (ii) have earnings above the minimum income threshold $Y_{\min,t}$, that is equivalent to one quarter of full-time work (13 weeks at 40 hours per week) at half of the legal minimum wage in year t (e.g., approximately $1,885 in 2010). The revolving panel for year t then selects individuals that are admissible in $t-1$ and in at least two more years between $t-5$ and $t-2$. This ensures that the individual was participating in the labor market and we can compute a reasonable measure of average recent income—a variable widely used extensively in the paper—which we describe next.

Recent Earnings. The average income of a worker i between years $t-1$ and $t-5$ is given by $\hat{Y}^i_{t-1} = \frac{1}{5}\sum_{j=1}^{5} \max\left\{\tilde{Y}^i_{t-j}, Y_{\min,t}\right\}$, where $\tilde{Y}^i_t$ denotes his earnings in year t. We then control for age and year effects by regressing $\hat{Y}^i_{t-1}$ on age dummies separately for

[8] Arellano *et al.* (2017) show that household earnings dynamics in the PSID also display non-Gaussian and nonlinear features.

each year, and define the residuals as **recent earnings** (hereafter RE), $\bar{Y}^i_{t-1}$. In Sections 3 and 4, we will group individuals by age and by $\bar{Y}^i_{t-1}$ to investigate how the properties of income dynamics vary over the life cycle and by income levels.

3 Cross-Sectional Moments of Earnings Growth

In this section, we study the distribution of earnings growth rates by analyzing its second to fourth moments. We start by describing our nonparametric method.

3.1 Empirical Methodology: A Graphical Construct

Our main focus is on how the moments of earnings growth vary with recent earnings and age. To this end, for each year t, we divide individuals into six groups based on their age in $t-1$ (25–29, 30–34, 35–39, 40–44 and 45–54), and then within each age group, sort individuals into 100 percentile groups by their recent earnings $\bar{Y}^i_{t-1}$. If these groupings are done at a sufficiently fine level, we can think of all individuals within a given age/RE group to be ex ante identical (or at least very similar). Then, for each such group, the cross-sectional moments of earnings growth between t and $t+k$ can be viewed as the properties of earnings changes that workers within that group expect to face looking ahead (see Figure 2). In our figures, we plot the average of these moments for each age/RE group over years between 1997 and 2013-k. This approach allows us to compute higher-order moments precisely as each bin contains a very large number of observations (see Table A.1 for sample size statistics). To make the figures more readable, we aggregate the six age groups into three: 25–34, 35–44, 45–54.[9]

Growth Rate Measures. In our analysis we use two measures of income change, each with its own distinct advantages and trade-offs. The first measure is *log growth rate* of income between t to $t+k$, $\Delta^k_{\log} y^i_t \equiv y^i_{t+k} - y^i_t$, where $y^i_t = \tilde{y}^i_t - d_{t,h(i,t)}$ denote the log income

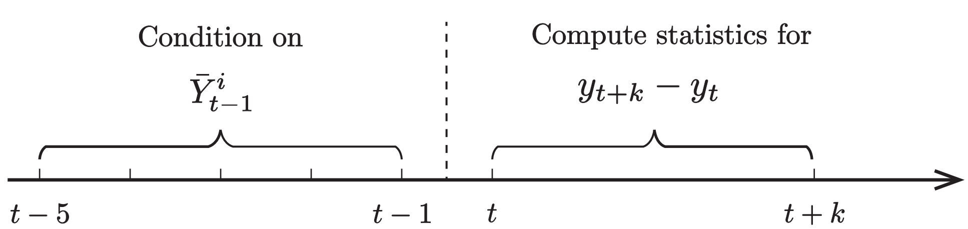


Figure 2 – Timeline for Rolling Panel Construction

[9]To be more precise, the first age group consists of individuals aged 28–34 in year $t-1$ due to the base sample requirement of having a minimum three years of admissible observations.

($\tilde{y}_t^i$) of individual i in year t at age $h(i,t)$ net of age and year effects $d_{t,h(i,t)}$. $\{d_{t,h}\}_{h=25}^{60}$ are obtained by regressing $\tilde{y}_t^i$ on a full set of age dummies separately in each year. While its familiarity makes the log change a good choice for the descriptive analysis, it has a well-known drawback that observations close to zero need to be dropped or winsorized at an arbitrary value. When we use $\Delta_{\log}^k y_t^i$, we drop individuals from the sample with earnings less than $Y_{\min}$ in t or $t+k$, and lose information in the extensive margin.

Our second measure of income growth—*arc-percent change*—is not prone to this caveat and is commonly used in the firm-dynamics literature, where firm entry and exit are key margins (e.g., Davis *et al.* (1996)). We define $\Delta_{\text{arc}}^k Y_t^i = \frac{Y_{t+k}^i - Y_t^i}{(Y_{t+k}^i + Y_t^i)/2}$, where earnings level $Y_t^i = \frac{\tilde{Y}_t^i}{\tilde{d}_{t,h(i,t)}}$ is net of average earnings in age h and year t, $\tilde{d}_{t,h(i,t)}$. Because of its familiarity, we use the log change measure in this section and report the results for arc-percent change in Appendix C.2, which show qualitatively similar patterns.

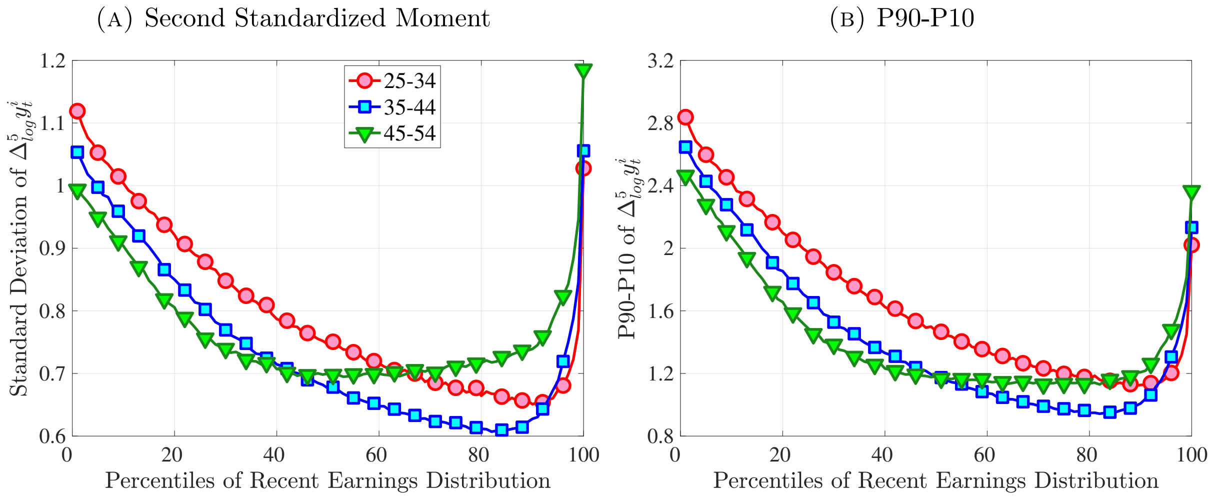
Transitory vs. Persistent Income Changes. As is well understood, longer-term earnings changes (i.e., $\Delta_{\log}^k y_t^i$ with larger k) reflect more persistent innovations. To see this intuition, consider the commonly used random-walk permanent/transitory model in which permanent (η_t^i) and transitory (ε_t^i) innovations are drawn from distributions F_η and F_ε, respectively. We denote the variance, skewness and excess kurtosis of distribution F_x, $x \in \{\eta, \varepsilon\}$ by σ_x^2, $\mathcal{S}_x$, and $\mathcal{K}_x$, respectively. Then the second to fourth moments of $k-$year log income growth $\Delta_{\log}^k y_t^i$ are given by (see Appendix B for the derivations):

$$\begin{aligned}
\sigma^2(\Delta_{\log}^k y_t^i) &= k\sigma_\eta^2 + 2\sigma_\varepsilon^2, \\
\mathcal{S}(\Delta_{\log}^k y_t^i) &= \underbrace{\frac{k \times \sigma_\eta^3}{(k\sigma_\eta^2 + 2\sigma_\varepsilon^2)^{3/2}}}_{<1} \mathcal{S}_\eta, \\
\mathcal{K}(\Delta_{\log}^k y_t^i) &= \underbrace{\frac{k \times \sigma_\eta^4}{(k\sigma_\eta^2 + 2\sigma_\varepsilon^2)^2}}_{<1} \mathcal{K}_\eta + \underbrace{\frac{2 \times \sigma_\varepsilon^4}{(k\sigma_\eta^2 + 2\sigma_\varepsilon^2)^2}}_{<1} \mathcal{K}_\varepsilon.
\end{aligned} \tag{1}$$

Equation 1 shows that as k increases the variance and kurtosis of $k-$year log change $\Delta_{\log}^k y_t^i$ reflect more of the distribution of η_t^i than that of ε_t^i. Also, skewness is solely driven by permanent changes.[10] Finally, the distribution of $\Delta_{\log}^k y_t^i$ is closer to normal than the underlying distributions of F_η and F_ε, because as innovations η_t^i and ε_t^i accumulate, the distribution of $\Delta_{\log}^k y_t^i$ converges toward Gaussian, per the central limit theorem.

[10] A time-varying distribution of transitory innovations can lead to asymmetry in earnings growth.

Figure 3 – Dispersion of Five-Year Log Earnings Growth



With these considerations in mind, we document the moments of one-year ($k = 1$) and five-year ($k = 5$) residual earnings growth to capture properties of transitory and persistent changes, respectively. As persistent changes have a greater effect on economic decisions compared with easier-to-insure transitory ones, we present the results for $k = 5$ in this section. The figures for $k = 1$ in Appendix C.1 show the same qualitative patterns.

3.2 Second Moment: Variance

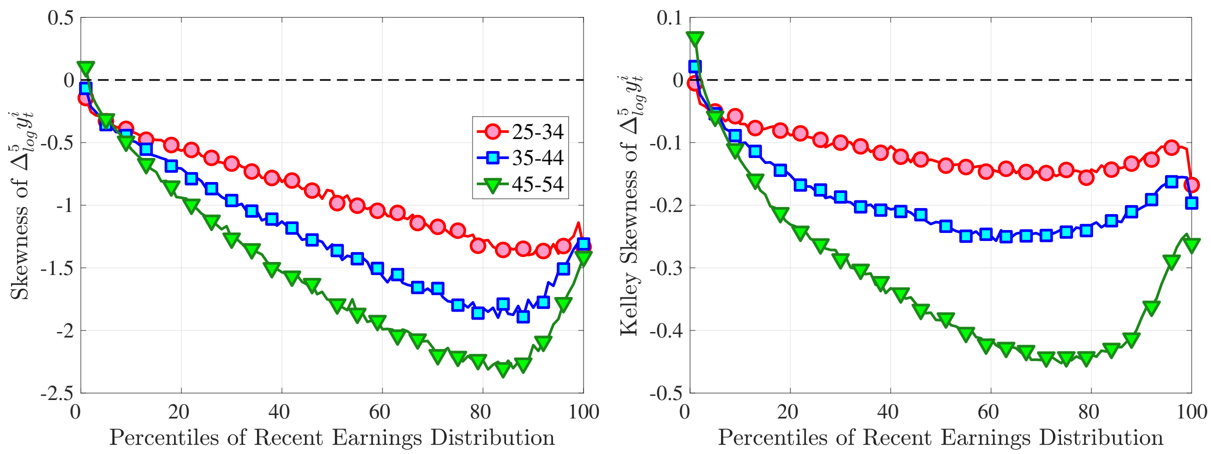
Figure 3a plots the standard deviation of five-year residual earnings growth by age and recent earnings groups (for clarity we use one marker for every 4th RE percentile group). In the right panel, we also report a quantile-based dispersion measure—the difference between the 90th and 10th percentiles of log earnings changes—denoted P90-P10, which is robust to outliers. By both measures and for every age group, there is a pronounced U-shape pattern across income groups. For example, for 35–44-year-olds, the standard deviation falls from 1.05 for the lowest RE group to 0.6 for the 90th percentile, and then rises rapidly to 1.05 for the top 1%.

Baker and Solon (2003) and Karahan and Ozkan (2013) have estimated a U-shaped life-cycle profile for the variance of persistent shocks. Our analysis reveals a more intricate life-cycle variation, as we also condition on recent earnings: Dispersion declines with age for the bottom one third of the RE distribution, is U-shaped until the 95th percentile, and monotonically increases for the top earners. However notice that the magnitude of variation with age is quite a bit smaller compared with the RE variation.

One important observation is that the highest earners (the top 5% or so) are strikingly

Figure 4 – Skewness of Five-Year Log Earnings Growth

(a) Third Standardized Moment　　(b) Kelley Measure



different from other high earners—even those just below the 95th percentile. The same theme will emerge again in our analysis of higher-order moments.

3.3 Third Moment: Skewness (Asymmetry)

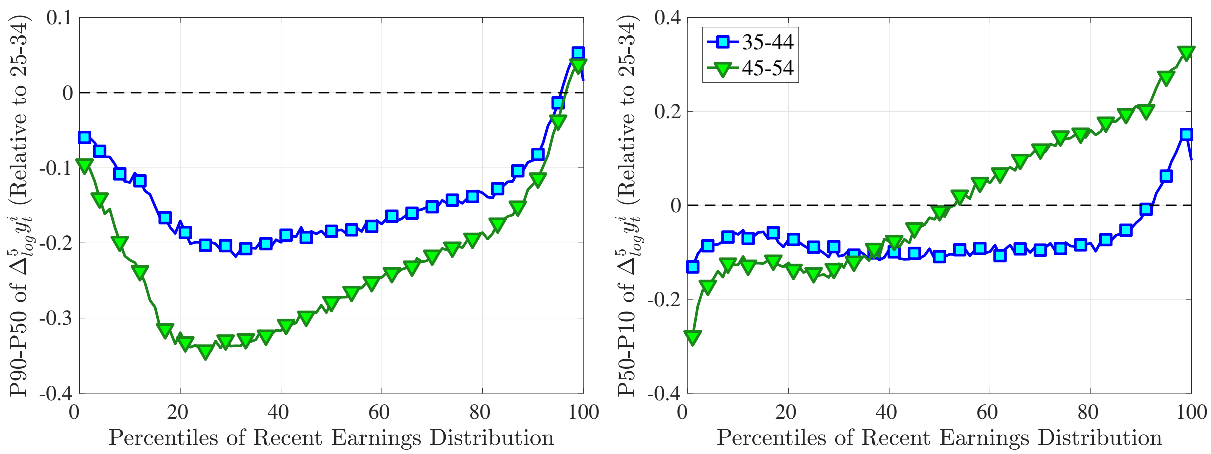
Figure 4a plots the skewness of five-year earnings growth, measured as the third standardized moment. First, notice that earnings changes are negatively (left) skewed at every stage of the life cycle and for (almost) all earnings groups. Second, skewness is increasingly more negative for individuals with higher earnings and as individuals get older. Thus, it seems that the higher an individual's current earnings, the more room they have to fall and the less room they have left to move up. Note that the variation in skewness with age is more muted for individuals at the bottom or top of the (recent) earnings distribution (similar to the dispersion patterns above).

Is negative skewness as measured by the third central moment driven by extreme observations? While the information on tails is important (and becomes especially valuable in estimating income processes in Section 6), we also look at Kelley (1947) skewness, $\mathcal{S}_{\mathcal{K}} = \frac{(\text{P90-P50})-(\text{P50-P10})}{\text{P90-P10}}$, which is robust to observations above the 90th or below the 10th percentile of the distribution. Basically, $\mathcal{S}_{\mathcal{K}}$ measures the relative fractions of the overall dispersion (P90–P10) accounted for by the upper and lower tails. Specifically, $\mathcal{S}_{\mathcal{K}} < 0$ implies that the lower tail (P50-P10) is longer than the upper tail (P90-P50).

Kelley's skewness exhibits essentially the same pattern as that of the third moment (Figure 4b). Thus, the asymmetry is prevalent across the entire distribution rather than being driven just by the tails. Furthermore, the magnitude of negative skewness is substantial. For example, a Kelley measure of –0.44 (for 45–54-year-old workers at the 80th percentile of the RE distribution) implies that P90-P50 accounts for 28% of P90-P10, far removed from the 50-50 of a normal distribution.

FIGURE 5 – Skewness Decomposed: P90-P50 and P50-P10 Relative to Age 25–34

(A) P90-P50 (B) P50-P10



Notes: The y-axes show the change in P90-P50 and P50-P10 from the youngest age group to the two older age groups.

Another question is whether skewness becomes more negative over the life cycle because of a compression of the upper tail (fewer opportunities for large gains) or because of an expansion in the lower tail (higher risk of large declines). To answer this question, we investigate how the P90-P50 and P50-P10 change over the life cycle from their levels in ages 25–34 (Figure 5). Up until age 44, both the P90-P50 and P50-P10 decline with age across most of the RE distribution. However, the upper tail compresses more strongly than the lower tail, which leads to the increasing left skewness. After age 45, the P90-P50 keeps shrinking, but the bottom end opens up for workers with above median RE (large declines become more likely). Top earners are again an exception to this pattern: the upper tail does not compress with age, but the bottom end opens up monotonically.

A natural question is whether the negative skewness is simply due to unemployment spells. First, notice that unemployment can generate negative skewness in earnings growth only if it has persistent effects: a transitory unemployment spell contributes one negative and one positive earnings change of similar size, leaving the symmetry unaffected (as shown by equation 1 for an arbitrary distribution F_ε). Jacobson *et al.* (1993) and Von Wachter *et al.* (2009) show that workers' earnings indeed experience large scarring effects after mass layoffs. We revisit this point in Section 6, where we link earnings changes to the underlying shocks. Second, negative skewness is stronger for upper middle-income

TABLE I – Fraction of Individuals within Selected Ranges of Log Earnings Change

	Prob($\lvert\Delta^1_{\log} y_t\rvert \in S$)		
$S:$	Data	$\mathcal{N}(0, 0.51)$	Ratio
$[-0.05, 0.05]$	30.6	7.7	3.88
$[-0.10, 0.10]$	48.8	15.4	3.27
$[-0.20, 0.20]$	66.5	30.2	2.23
$2\sigma+$: $[1.02, \infty)$	6.64	4.55	1.46
$3\sigma+$: $[1.53, \infty]$	2.37	0.027	8.77

Notes: The empirical distribution used in this calculation is for 1997-98, the same as in Figure 1.

and older workers, for whom unemployment risk is relatively small, implying that decline in hours is not the main driver. Finally, as noted above, the shift toward more negative skewness is mostly coming from the compression of the right tail up to age 45, which is unlikely to be related to unemployment.

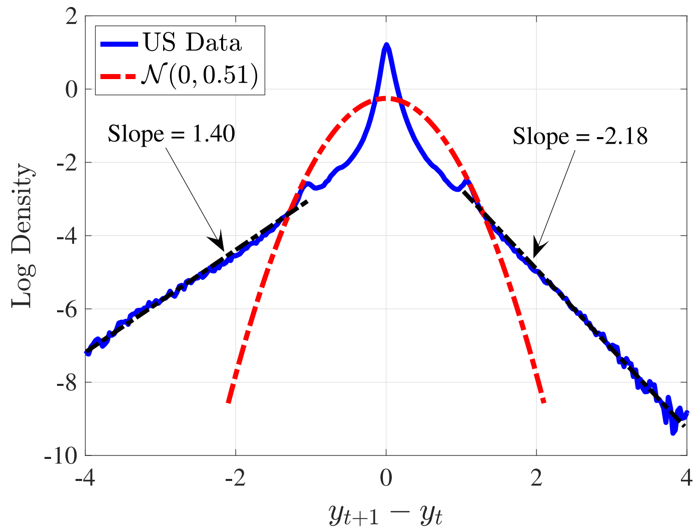
3.4 Fourth Moment: Kurtosis (Peakedness and Tailedness)

We begin by discussing what kurtosis measures. One can think of it as the tendency of a probability distribution to stay away from $\mu \pm \sigma$ (see Moors (1986)). Thus, a leptokurtic distribution often has a sharp/pointy center, long tails, and little mass near $\mu \pm \sigma$. A corollary to this description is that with excess kurtosis, the usual way we think about standard deviation—as representing the size of the typical earnings change—is not very useful: most realizations will be either close to the center (median) or in the tails.

To illustrate this point, we calculate concentration measures for earnings growth. Table I reports the fraction of individuals experiencing an absolute log earnings change less than a threshold $x = 0.05, 0.10$, and so on. In the data 31% of workers experience an earnings change of less than 5%, whereas if innovations were drawn from a Gaussian density with the same standard deviation as the data, only 8% of individuals would experience such changes. Furthermore, extreme events are more likely in the data: A typical worker experiences a change larger than three standard deviations (153 log points) once in a lifetime—with a 2.4% annual chance—whereas this probability is almost one-ninth that size under a normal distribution. These values suggest that the Gaussian assumption vastly overstates the typical earnings growth and misses the extreme changes received by a non-negligible share of the population.

The high likelihood of extreme events in the data motivate us to take a closer look at the tails of the earnings growth distribution by examining its empirical log density versus

Figure 6 – Double-Pareto Tails of the U.S. Annual Earnings Growth Distribution



Notes: The empirical distribution in this figure is for 1997-98, the same as in Figure 1 but with the y-axis now in logs.

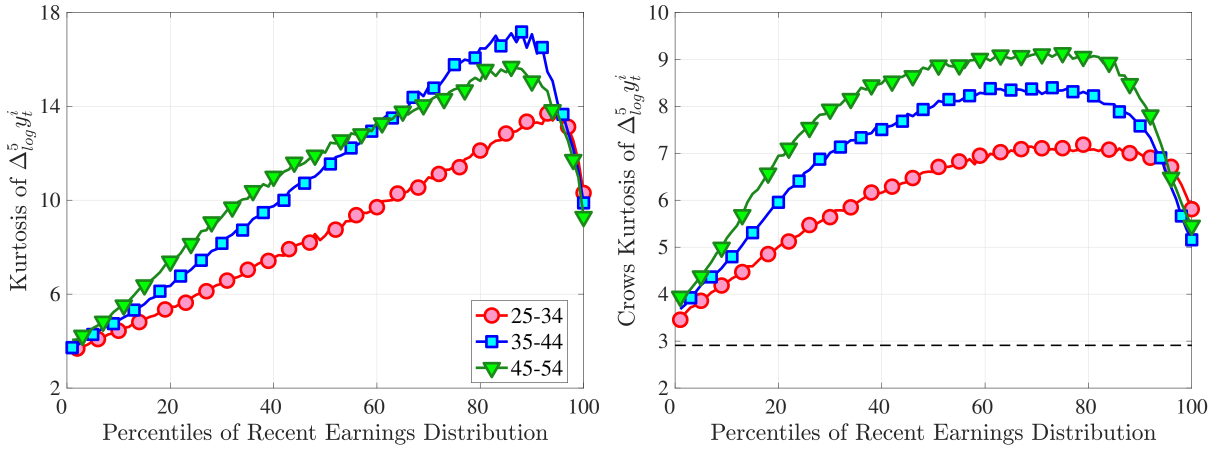
the Gaussian log density (which is an exact quadratic). First, in line with our previous discussion, the data has much thicker *and* longer tails compared with a normal distribution (Figure 6). Second, the tails decline almost linearly, implying a Pareto distribution at both ends. Third, they are asymmetric, with the left tail declining much more slowly than the right, which contributes to the left skewness documented above. In fact, fitting linear lines to each tail (in the regions $\pm[1, 4]$) yields a tail index of 1.18 for the right tail and 0.40 for the left tail—the latter showing especially high thickness. We highlight that while the Pareto tail in the earnings levels distribution is well known—indeed, going as far back as Pareto (1897)—the two Pareto tails emerge here in the earnings *growth* distribution. To our knowledge, the present paper is the first to document this fact.

Finally, we quantify the extent of kurtosis by looking at two measures. The first is the familiar fourth standardized moment of the data. For the same reasons given above for dispersion and skewness, the second one is the quantile-based Crow and Siddiqui (1967) measure, $\kappa_{\text{C-S}} = \frac{P97.5-P2.5}{P75-P25}$, which is equal to 2.91 for a Gaussian distribution.

Both measures of kurtosis increase monotonically up to the 80th to 90th percentiles of RE for all age groups (Figure 7). That is, high earners experience *even smaller* changes of either sign, and few experience very large changes. Also, kurtosis tends to increase with age for all RE levels (except the top 5%). It peaks over RE at 14 for the youngest group and at 17.5 for the 35–44-year-olds. The Crow-Siddiqui measure also shows very high kurtosis levels, indicating that the excess kurtosis is not driven by outliers.

FIGURE 7 – Kurtosis of Five-Year Log Earnings Growth

(A) Fourth Standardized Moment (B) Crow-Siddiqui Measure



Overall, these findings show that earnings changes in the U.S. data exhibit important deviations from lognormality, and the extent of these deviations varies both over the life cycle and with the income level of individuals.

An Alternative Approach to Isolating Persistent Changes As we argued earlier, five-year log income change acts as a proxy for persistent innovations. However, it still contains the transitory innovations in years t and $t+5$. Another approach is to look at the growth between average future (over t to $t+4$) and average past earnings (over $t-1$ to $t-5$), which should partially purge transitory changes; i.e., $\overline{\Delta}^5_{\log}(\bar{y}_t{}^i) \equiv \bar{y}^i_{t+4} - \bar{y}^i_{t-1}$, where $\bar{y}^i_{t+4} \equiv \log(\bar{Y}^i_{t+4})$ and $\bar{y}^i_{t-1} \equiv \log(\bar{Y}^i_{t-1})$. $\bar{Y}^i_{t+4}$ is analogous to recent earnings $\bar{Y}^i_{t-1}$ but is calculated over the period t to $t+4$.

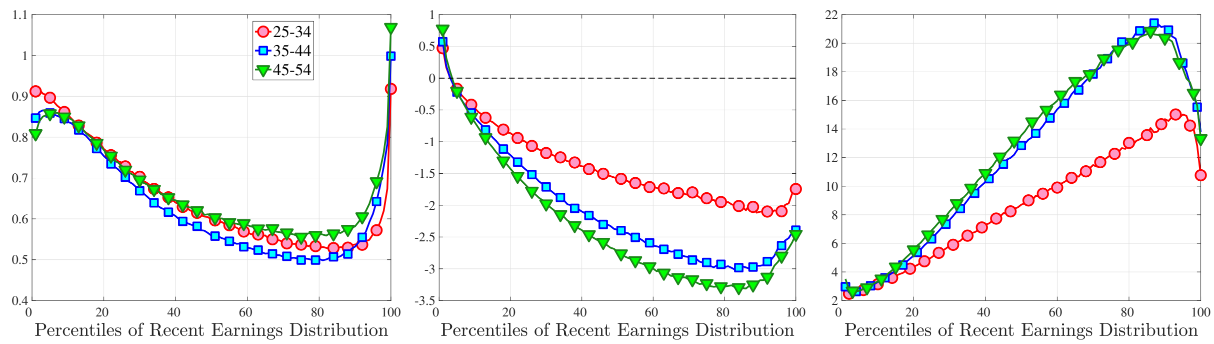
If nonnormalities are present only in the transitory changes, we should expect this alternative measure to be closer to a Gaussian distribution. Centralized moments of this measure display similar patterns to those in our baseline figures, confirming strong non-Gaussian features in persistent changes (Figure 8). In fact, this measure displays a slightly higher left-skewness and excess kurtosis compared to the 5-year log change measure, which suggests that nonnormalities are stronger in persistent innovations than transitory ones. A more formal exploration in Section 6 will reach similar conclusions.

3.5 Job-Stayers and Job-Switchers

Economists have documented that the earnings changes of workers who stay with the same employer (job-stayers) are notably different from the changes of workers who switch jobs (job-switchers) (see, Topel and Ward (1992), Low *et al.* (2010), and Bagger *et al.*

FIGURE 8 – Centralized Moments of Persistent Earnings Changes, $\overline{\Delta}^5_{\log}(\bar{y}_t^{\,i})$

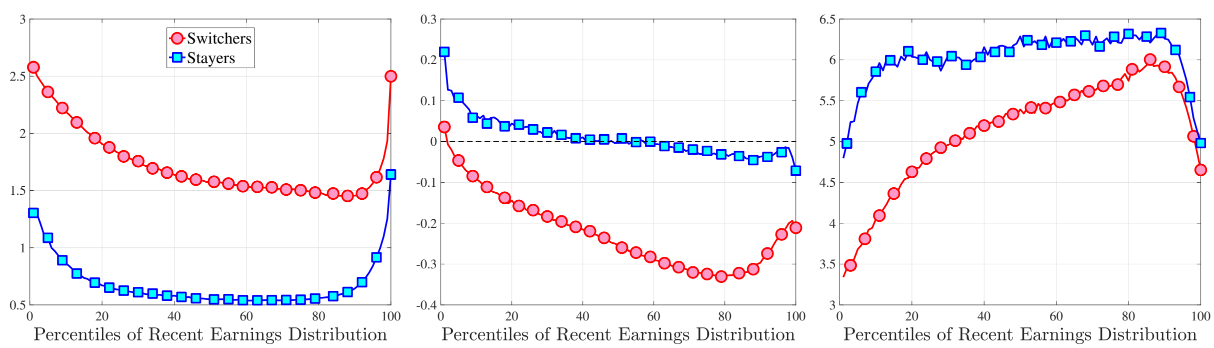
(A) Standard Deviation (B) Skewness (C) Kurtosis



(2014)). This literature has focused on the average change, whereas we examine how the higher-order moments of earnings growth vary between job-stayers and job-switchers.

FIGURE 9 – Higher-Order Moments of Earnings Growth, $\Delta^5_{\log} y^i_t$: Stayers vs Switchers

(A) P90-10 (B) Kelley's Skewness (C) Crow-Siddiqui Kurtosis



The SSA dataset contains employer identification numbers (EINs) that allow us to match workers to firms. However, the annual frequency of the data, together with the fact that some workers hold multiple jobs in a given year, poses a challenge for a precise identification of job-stayers and job-switchers. We have explored several plausible definitions for stayers and switchers and found qualitatively similar results. Here, we describe one reasonable definition: A worker is said to be a job-stayer between years t and $t+1$ if he has a W-2 form from the same firm in years $t-1$ through $t+2$, and that firm is the main employer by providing at least 80% of his total annual earnings in years t and $t+1$. A worker is defined as a job-switcher if he is not a job-stayer.

We show in Figure 9 how the quantile-based second to fourth moments of five-year earnings growth for stayers and switchers vary with recent earnings. Relative to job-

Table II – Higher-Order Moments of Two-Year Changes in the PSID

		All			25–39		40–55	
	Normal	Earnings	Wages		Earnings	Wages	Earnings	Wages
Skewness	0.0	–0.26	–0.14		–0.17	–0.20	–0.34	–0.09
Kelley Skew.	0.0	–0.02	–0.02		0.03	0.016	–0.06	–0.04
Kurtosis	3.0	12.26	13.65		10.44	9.00	14.01	17.10
Crow Kurt.	2.91	6.83	5.59		6.33	5.02	7.33	6.11

Note: Wages are obtained by dividing annual earnings of male heads of households by their annual hours in the PSID using data over the period 1999–2013, during which data are biennial.

switchers, job-stayers experience earnings changes that have a smaller dispersion (about one-third for median-income workers), and are more leptokurtic, especially for low-RE workers. Changes are symmetric or slightly right skewed for stayers and left skewed for switchers. The age profiles are broadly similar across switchers and stayers, and figures for annual changes and centralized moments display similar patterns (Appendix C.4).

3.6 What Are the Sources of Nonnormalities in Earnings Growth?

So far, our analysis has focused on the distribution of *annual earnings changes* and remained silent on what may be behind the nonnormalities. For example, are the left skewness and excess kurtosis also present in the wage growth distribution? What are the life-cycle events associated with extreme income changes? The lack of information in the SSA data other than annual earnings does not allow us to investigate these questions, which, in turn, we study using the PSID.

3.6.1 Separating Earnings, Wages, and Hours

For many economic questions, it is important to know the extent to which nonnormalities in earnings dynamics are driven by wages versus hours. For example, if nonnormalities come from changes in hours and not wages, this would suggest focus on hours to identify its underlying sources, e.g., preferences for work or shocks to labor supply (health shocks, involuntary layoffs, etc.). If, instead, nonnormalities are also present in wage changes, that would point to a different set of factors to focus on. To shed light on this question, we analyze the wage growth distribution in the PSID using a sample that closely mimics the SSA sample (see Appendix C.7 for the construction of the sample).[11]

We start by investigating the non-Gaussian features of two-year earnings changes in the PSID (Table II). The centralized third moment and the Kelley measure point to a

[11] Another approach is to estimate structural models of endogenous labor supply (such as those in Low *et al.* (2010) and Heathcote *et al.* (2010b)) to study the role of hours versus wages in higher-order moments of earnings growth.

weakly left skewed distribution, possibly due to added noise in the PSID to the extent that measurement error is symmetric. Excess kurtosis is a more striking feature: Both measures of kurtosis from the PSID are quite close to their SSA counterparts. The age patterns are also broadly in line with those from administrative data. In addition, De Nardi *et al.* (forthcoming) document the income variation in higher order moments of earnings growth from the PSID and find similar patterns to those in the SSA data.

Turning to hourly wage growth, negative skewness in the overall sample is even less pronounced than that of earnings. Unlike skewness, excess kurtosis of wage growth, and its life-cycle variation are roughly similar to those features of earnings growth.[12] This evidence suggests the leptokurtic property of earnings growth cannot be driven entirely by the hours margin. We also conducted an analogous analysis using data from the Current Population Survey (CPS), which has a larger sample size and reached similar conclusions, specifically a weak left skewness (possibly due to measurement error) and strong excess kurtosis in earnings and wage growth (see Appendix C.7).

Motivated by the importance of extreme earnings changes for excess kurtosis, we investigate the roles of hours and wages in the tails of the earnings growth distribution. For this purpose, we distribute workers into six groups based on their two-year residual earnings change. As in the SSA data, most workers experience only small earnings changes (col. 1 of Table III). For each group, we compute the average change in residual earnings, hours and wages (Table III, cols. 2-4).[13] Our results show that wage changes are at least as important as hours changes. For example, the bottom group with the 165 log points average earnings decline experiences a 101 log points drop in wages. Clearly, extensive margin events (e.g. layoffs) can lead to large declines in hours and wages at the same time. Moreover, wage changes seem to be even more important for smaller earnings changes (e.g. more than 70% of $|\Delta y| < 0.25$ can be attributed to wages).

3.6.2 Linking Earnings Changes to Life-cycle Events

In this section, we link large earnings changes to various life-cycle events. We start with a natural suspect: nonemployment spells. The group with the largest earnings

[12]It is well known that there is significant measurement error in hours in the PSID, which leads to "division bias" when constructing hourly wages (see Bound *et al.* (2001) and Heathcote *et al.* (2010b)). This measurement error attenuates left skewness (if it is classical) and excess kurtosis (if it is Gaussian) implying that our estimates are lower bounds (see Halvorsen *et al.* (2018)).

[13]Log earnings, hours and wages have been residualized by similar but separate regressions. Therefore, at the individual level, residual hours and wages do not exactly add up to residual earnings. However, as cols. 2–4 of Table III show, the discrepancy is negligible.

Table III – Important Life Cycle Events and Earnings Changes

Group $\Delta y \in$	Share % (1)	Mean Δy (2)	Mean Δw (3)	Mean Δh (4)	Δ wks not empl. (5)	Occup. switch % (6)	Employer switch % (7)	Disab. Flow in % (8)
$(-\infty, -1)$	3.8%	–1.65	–1.01	–0.64	10.01	26.1	45.6	9.2
$[-1, -0.25)$	14.4%	–0.48	–0.34	–0.14	1.62	14.9	29.1	4.4
$[-0.25, 0)$	31.2%	–0.11	–0.08	–0.03	0.17	6.9	13.3	3.5
$[0, 0.25)$	31.1%	0.11	0.08	0.03	-0.03	5.3	9.7	2.8
$[0.25, 1)$	16.5%	0.47	0.34	0.13	-1.30	8.6	16.9	2.9
$(1, \infty)$	3.0%	1.64	1.06	0.58	-7.51	18.0	30.7	3.8

Notes: This table shows hours and wage growth (Δh and Δw, respectively) and the various life-cycle events for people in different biennial earnings change (Δy) groups. In column 5, "weeks not employed" is the sum of weeks unemployed and out of the labor force. Columns 6 and 7 show the fraction of workers that switch occupation and employer within each earnings change group, respectively. Column 8 shows the fraction of workers that become disabled in that period.

decline also reports the largest increase in the incidence of nonemployment—10 weeks (Table III, col. 5). Similarly, the group with the largest earnings increase reports the largest decline in nonemployment.[14] These results underline the importance of the extensive margin for the tails of the earnings change distribution.

Next, we study occupation and job mobility, both of which are known to be associated with large changes in earnings. The likelihood of occupation and employer switches follows a distinct U-shape pattern with earnings changes (Table III, cols. 6 and 7, respectively). Compared to the workers with small changes ($|\Delta y| < 0.25$), the top and bottom earnings-change groups are three to four times more likely to make these switches. The sources of mobility are possibly very different at the top and the bottom earnings-change groups. For example, the switches at the top are likely associated with promotions or outside offers, whereas moves at the bottom are probably necessitated by job losses. We also looked into involuntary geographic moves and found that they are associated with large earnings changes too (Appendix C.7).

Finally, we investigate health shocks, which are known to have large effects on earnings (see Dobkin *et al.* (2018)). We focus on disabilities that affect individuals' work performance (see Appendix C.7 for a detailed description). We find higher transition rates into disability for workers with earnings declines, with the highest transition (9.2%) in the bottom earnings-change group (Table III, col. 8). These results suggest that the extreme earnings changes are not purely a statistical artifact or measurement error.

[14]We reached similar conclusions analyzing unemployment and out of labor force separately. The incidence of unemployment is somewhat lower in the PSID than it is in the CPS: 6.8% of our sample reports some unemployment in the previous year, compared with around 10% in a similar CPS sample.

Disability Income If health shocks are an important source of earnings changes, how important is disability insurance as a safety net? To answer this question, we add individuals' Social Security Disability Income (SSDI) from the SSA to their labor income and construct a "total income" measure. Our results in Appendix C.9.1 show that the cross-sectional moments of total income overlap with their labor income counterparts, mainly because the share of SSDI recipients is small, ranging from 1.3% in 1978 to 4.1% in 2013. However, SSDI makes a noticeable (albeit slight) difference only for the oldest group of workers, who constitute the majority of the recipients.

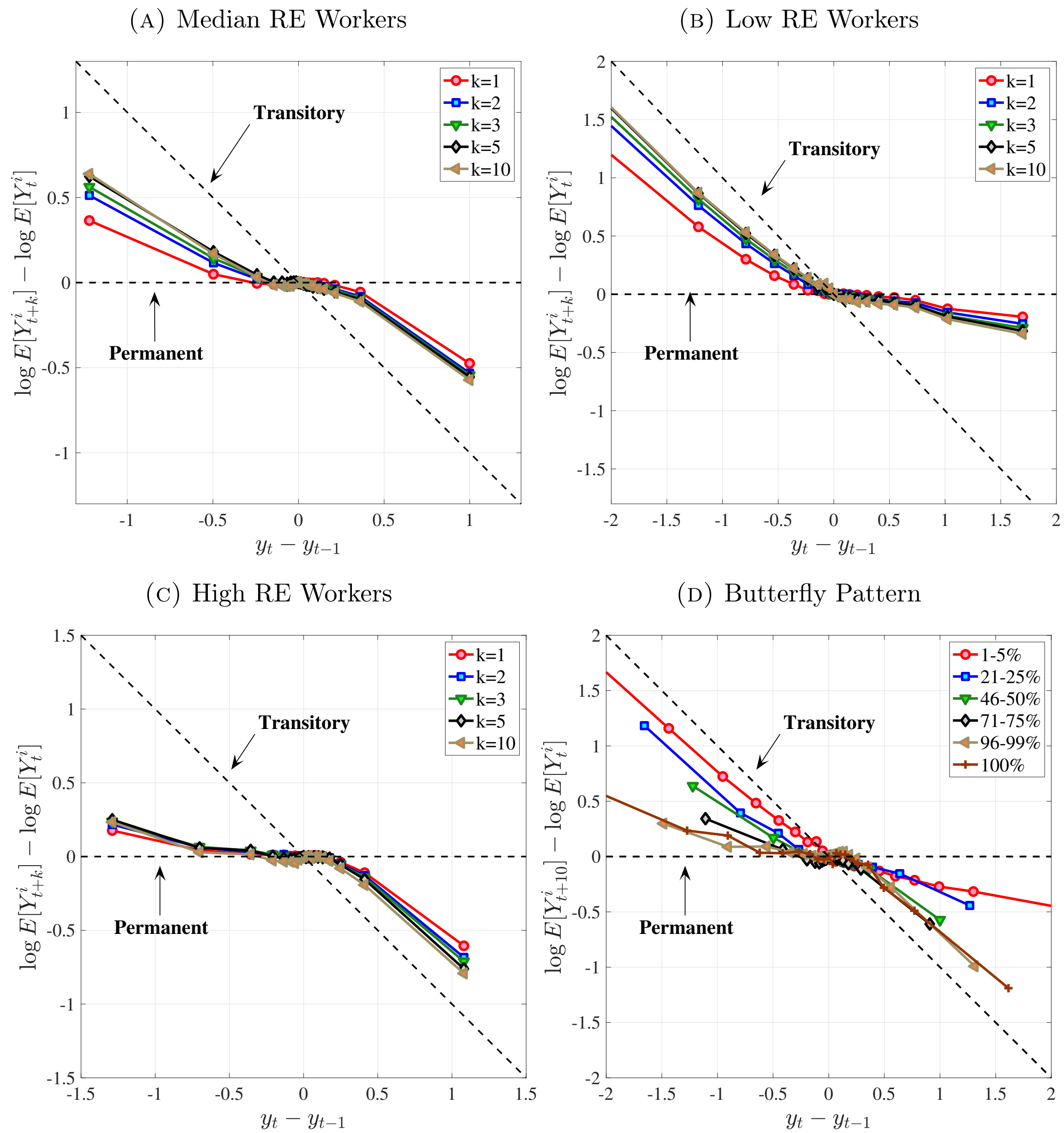
To sum up, in light of the vast micro literature that finds very small Frisch elasticities, large changes in earnings, especially declines, are much more likely to represent involuntary shocks beyond the worker's control such as health problems, reductions in hours imposed by the employers, or unemployment.

Related Work A growing literature uses detailed administrative data from various countries to study the determinants of nonnormalities in earnings growth. Kurmann and McEntarfer (2018) show that the distribution of hourly wage growth displays high excess kurtosis for job-stayers in the U.S., and wage changes constitute a substantial portion of the earnings changes, mostly for those experiencing increases. These findings are overall consistent with ours. They also argue that large declines in hours of job-stayers are involuntary and imposed by firms. Blass-Hoffmann and Malacrino (2017) use Italian data to argue that changes in weeks worked generate the tails of the one-year and five-year earnings growth distributions and account for their procyclical left skewness (first documented by Guvenen *et al.* (2014b) for the U.S.). In contrast, our analysis from the PSID also attributes an important role to wage growth. Moreover, we find that scarring effects are necessary to generate left skewness through extensive margin fluctuations. Finally, Halvorsen *et al.* (2018) use Norwegian data and find that hourly wage changes exhibit left-skewness and excess kurtosis, and both the magnitudes and their life-cycle and income variation are similar to what we find for earnings changes. Furthermore, they show that large earnings changes are mostly driven by wage changes for high-RE individuals, but the split between wages and hours is more equal for low-RE workers.

4 Dynamics of Earnings

Having studied the distribution of earnings changes, we now turn to their persistence. Typically, earnings dynamics are modeled as an AR(1) or a low-order ARMA process, and the persistence parameter is pinned down by the rate of decline of autocovariances with the lag order. While this linear approach might be a good first-order approximation, it imposes strong restrictions, such as the uniformity of mean reversion for positive and negative or large and small changes as well as for workers with different earnings levels.

FIGURE 10 – Impulse Responses, Prime-Age Workers



Notes: Median-, low-, and high-RE in panels A, B, and C refer to workers with $\overline{Y}_{t-1}$ in $(P46 - P55)$, $(P6 - P10)$, and $(P91 - P95)$, respectively. Prime age refers to age 35 to 50.

We exploit our large sample and employ a nonparametric strategy to characterize the nonlinear mean reversion. We do so by documenting the impulse response functions of earnings changes of different sizes and signs for workers with different recent earnings. In particular, we group workers by their earnings growth between $t-1$ and t, their recent earnings $\overline{Y}^i_{t-1}$, and age, and then follow their earnings over the next 10 years.

To reduce the number of graphs to a manageable level, we combine the first two age groups (ages 25 to 34) into "young workers" and the next three groups (ages 35 to 50) into "prime-age workers." Within each age group, we rank and group individuals by $\overline{Y}^i_{t-1}$ into the following 21 RE percentiles: 1–5, ... , 91–95, 96–99, and 100. Next, within each age and RE group, we sort workers by the size of their log earnings change between $t-1$ and t ($y^i_t - y^i_{t-1}$) into 20 equally-sized quantiles. Hence, all individuals within a group have similar age and average earnings up to $t-1$, and experience a similar change from $t-1$ to t. For each such group of individuals, we then compute the *log change of their average earnings* from t to $t+k$, $\log \mathbb{E}\left[Y^i_{t+k}\right] - \log \mathbb{E}\left[Y^i_t\right]$, where Y^i_t is the income level net of age and time effects. Rather than taking the *average of log earnings change*, this approach allows us to include workers with earnings below the minimum threshold, thereby keeping the composition of workers constant for each k. The results for the alternative approach are qualitatively very similar and are available upon request.

4.1 Impulse Response Functions Conditional on Recent Earnings

In Figure 10, we show the mean reversion of different sizes of earnings changes $y^i_t - y^i_{t-1}$ for prime-age workers over a 10-year period. Specifically, we plot $\log \mathbb{E}\left[Y^i_{t+k}\right] - \log \mathbb{E}\left[Y^i_t\right]$ of each $y^i_t - y^i_{t-1}$ quantile on the y-axis against its average on the x-axis.[15] This graphical construct contains the same information as a standard impulse response function but allows us to see the heterogeneous mean reversion patterns more clearly.

We start with the median-RE group ($\overline{Y}_{t-1} \in P46 - P55$) in Figure 10a. Even at the 10-year horizon, a nonnegligible fraction of the earnings change is still present for this group of workers, indicating a very persistent component in earnings growth. Also, negative changes tend to recover more gradually than positive ones for them. For example, workers whose earnings rise by 100 log points between $t-1$ and t lose about 50% of this increase in the following 10 years. Almost all of this mean reversion happens after one year. Workers whose earnings fall by 100 log points recover 25% of that decline in the first year and around 50% of the total within 10 years. Finally, the degree of mean

[15]The average mean reversion varies across the RE groups because of different earnings histories. Therefore, we normalize earnings changes on both the x- and y-axes such that their values at the median quantile of $y^i_t - y^i_{t-1}$ cross at zero.

reversion varies with the magnitude of earnings changes, with stronger mean reversion for large changes: Small innovations (i.e., those less than 10 log points in absolute value) look very persistent, whereas larger earnings changes exhibit substantial mean reversion. A univariate autoregressive process with a single persistence parameter will fail to capture this behavior. In Section 6, we will show how to modify the simple income process to accommodate this variation in persistence by the size and sign of the earnings shock.

The analogous impulse response functions for low-income ($\overline{Y}_{t-1} \in P6 - P10$) and high-income ($\overline{Y}_{t-1} \in P91 - P95$) workers (Figures 10b and 10c) show that for low-income individuals, negative changes are more short-lived, whereas positive ones are more persistent, and that for high-income individuals the opposite is true.

Extending the results to the entire distribution of recent earnings, we focus on a fixed horizon and plot the cumulative mean reversion from t to $t + 10$ for the 6 RE groups in Figure 10d. Starting from the lowest RE group (the bottom 5%), notice that negative changes are transitory, with an almost 75% mean reversion rate at the 10-year horizon. But positive changes are quite persistent, with only about a 25% mean reversion at the same horizon. As we move up the RE distribution, the positive and negative branches of each graph start rotating in *opposite* directions, so that for the highest RE group (top 1%), we have the opposite pattern: only 20 to 25% of earnings declines revert to the mean at the 10-year horizon, whereas around 80% of the increases do so at the same horizon. We refer to this shape as the "butterfly pattern."

This butterfly pattern broadly resonates the earnings dynamics in job ladder models. For high-RE workers—who are at the higher rungs of the ladder—a job loss leads to a more persistent earnings decline relative to low-RE workers because of search frictions. Similarly, for low-RE workers, large increases are likely due to unemployment-to-employment or job-to-job transitions, which have long-lasting effects on earnings.[16]

5 Earnings Growth and Employment: The Long View

In this section, we turn to two questions that complete the picture of earnings dynamics over the life cycle. Both questions pertain to long-term outcomes—covering the entire working life. The first one is about average earnings growth—complementing the

[16] Lise (2012) shows that a job ladder model with precautionary savings motive captures wealth and consumption dispersions better than the incomplete markets model with a linear-Gaussian income process.

second to fourth moments analyzed in Section 3. In particular, how much cumulative earnings growth do individuals experience over their working life, and how does that vary across individuals with different lifetime incomes?

The second question investigates the lifetime nonemployment rate—defined as the fraction of working life an individual spends as full-year nonemployed. Although incidence of *long-term* nonemployment is of great interest for many questions in economics, documenting it requires long panel data with no sample attrition, a phenomenon most common among long-term nonemployed. The administrative nature of the MEF data set and its long panel dimension provide an ideal opportunity to study this question.

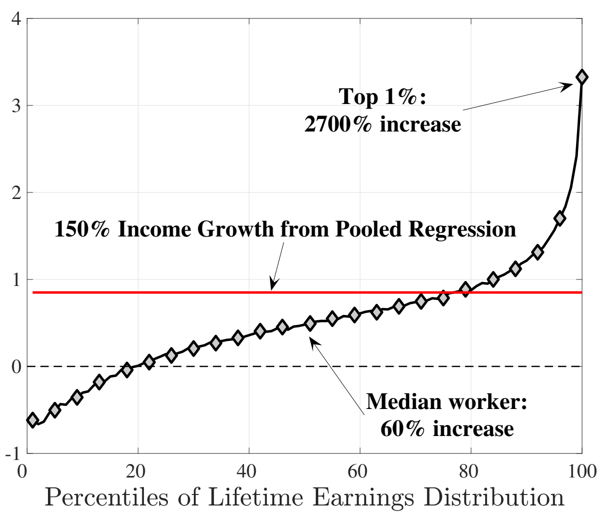
5.1 Lifecycle Earnings Growth And Its Distribution

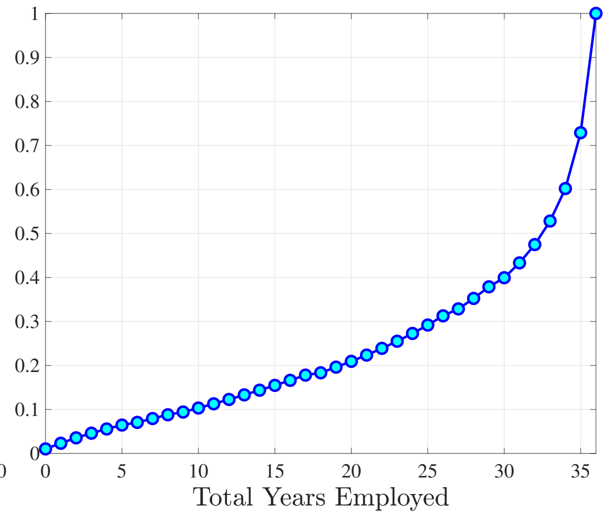
For the analysis in this section, we use the full length of the MEF panel, covering 1978 to 2013. We select individuals who (i) were born between 1951 and 1957 (hence for whom we have 33-years of data between ages 25 and 60), and (ii) had annual earnings above $Y_{\min,t}$ in at least 15 years, thereby excluding workers with very weak labor market attachment. We take a closer look at this latter group in the next subsection. We sort individuals into 100 percentiles by their lifetime earnings (LE), computed by averaging their earnings from age 25 through 60. For each LE percentile bin, denoted LEj, $j = 1, 2, ..., 99, 100$, we compute the growth rate between ages h_1 and h_2 by differencing the average earnings across all workers (including those with zero earnings) in those LE and age cells; i.e., $\log(\overline{Y}_{h_2,j}) - \log(\overline{Y}_{h_1,j})$, where $\overline{Y}_{h,j} \equiv \mathbb{E}(\tilde{Y}_t^i | i \in \text{LE}j, h(i,t) = h)$.

The results in Figure 11a show that between ages 25 and 55 the median individual (by LE) experiences a smaller earnings growth—about 60%—than a 150% mean growth estimated from a Deaton-Paxson pooled regression (see Appendix C.6). More importantly, higher-LE workers experience a much higher earnings growth over the lifecycle compared with the rest of the distribution. While an upward slope per se is not surprising (as it is partly mechanical—faster growth will deliver higher LE, everything else held constant), the variation at the top end is so large, and the curvature is so steep, that it turns out difficult to capture using simple earnings processes, as we discuss in the next section. For example, average earnings grow by 1.5-fold (91 log pts) over 31 years at $LE80$, by 4.8-fold (157 log pts) at $LE95$, and by 27.9-fold (333 log pts) in the top 1%.[17]

[17] Earnings decline from age 45 to 55 for 80% of the population, and those above LE80 experience only small increases (Figure C.21b in Appendix C.6). The decline in earnings later in the life could be due to a decline in hours because of partial retirement (see Aaronson and French (2004)).

FIGURE 11 – Earnings Growth and Employment: The Long View





One question is whether this extremely high growth rate at the top is driven by higher rates of school enrollment in these groups at age 25 (and thereby low earnings). While the lack of education data does not allow us to answer this question directly, several pieces of evidence are informative. First, about 21.7% of individuals in the $LE100$ has earnings below the $Y_{\min}$ threshold at age 25, which is higher than the rate for half of the sample, suggesting schooling could be playing some role (see Figure C.36a). However, this rate drops quickly to 5.95% by age 30, which is one of the lowest in our sample. At the same time, earnings growth for this group between ages 25 and 30 is only slightly higher compared to that between ages 30 and 35 (2.9-fold vs 2.6-fold), when schooling is unlikely to matter much. Similarly, looking at growth from 35 to 55, we still find a steep profile of earnings growth with respect to LE (see Figure C.21). These observations suggest that low labor supply at age 25 is not the major driver of these patterns.

Turning to the lower end, individuals below $LE20$ see their earnings *decline* from age 25 to 55. How important is disability for this decline? Adding SSDI to labor earnings has virtually no effect above the 40th LE group or so (Figure C.35). But it matters at the lower end, mitigating the decline by 18% for LE10 and by 24% for LE5.

5.2 Lifetime Employment Rate and Its Distribution

Next, we investigate the lifetime nonemployment rates across individuals. Using the same criteria as before—working life defined as the period between ages 25 and 60, and full-year nonemployed defined as annual earnings below $Y_{\min}$—we examine the cumulative distribution of total lifetime years employed in Figure 11b. The results show that, first, a large fraction of individuals are very strongly attached to the labor market: 28% of individuals were never nonemployed during their working life, and almost half (48%) were nonemployed for less than three years. But second, there is a long left tail of the distribution, showing a surprisingly large fraction of men who spend half of their working life or more without employment: 18.3% of men spend 18 years—or half of their working life—as full-year nonemployed, and 12.3% spend at least 24 years as nonemployed.

To understand these magnitudes, note that the employment-to-population ratio for prime-age men in cross-sectional data (such as the one the Bureau of Labor Statistics publishes monthly) has averaged around 86% during this time period, implying a monthly nonemployment rate around 14%. Getting nonemployment at annual frequency for 24 years for 12% of the male population requires an extremely high persistence of the long term nonemployment state. As we shall see in the next section, this statistic turns out to be very hard to match with a simple earnings process with standard parameter values.

6 Econometric Models for Earnings Dynamics

The empirical facts documented in the last three sections can be viewed as snapshots of an earnings process taken from different angles. They allowed us to identify key patterns by (partially) isolating other features. That said, for many purposes, it is essential to combine these different snapshots to get a fuller picture of the underlying earnings process. Therefore, our goal in this section is to search for earnings processes that can reproduce the key empirical facts documented above.

In light of the nonlinearities and nonnormalities revealed by the descriptive analysis, it is clear that we need to move beyond the linear-Gaussian framework, which has been the workhorse for modeling earnings dynamics (with a few exceptions noted earlier). The approach we follow here is to start from a simple and widely used linear-Gaussian model and extend it incrementally until we arrive at a more general model that is consistent with the key features documented. This approach has two advantages over starting from an entirely new nonlinear modeling framework. First, by adding each piece incrementally to a well-understood benchmark, we can learn what each new component brings. Second, the components we include are already familiar to economists from previous work, which should make it easier to be implemented in future quantitative models.

We have conducted an extensive search for a suitable model specification, which involves estimating more than a hundred different specifications. We have conducted a battery of diagnostic tests on each set of moments to understand which components have the potential to capture them. As always, there is a trade-off between the complexity of the specification and the set of moments it can reproduce. A key guiding principle was to keep the process as parsimonious as possible, both in terms of the number of parameters, but even more importantly, the number *state variables* the process will require when embedded in a dynamic programming problem. Overall, our benchmark process described below achieves both goals: It requires only one state variable—the same as a standard persistent-plus-transitory model—while offering a good fit to the data. It does however have more parameters, as we discuss next.

6.1 A Flexible Stochastic Process

The econometric models we estimate are special cases (with a few exceptions) of the following benchmark specification, which includes (i) an AR(1) process (z_t^i) with innovations drawn from a mixture of normals; (ii) a nonemployment shock whose incidence probability $p_\nu^i(t, z_t)$ can vary with age or income or both, and whose duration (ν_t^i) is exponentially distributed ; (iii) a heterogeneous income profiles component (HIP); and (iv) an i.i.d. normal mixture transitory shock (ε_t^i):

$$\text{Level of earnings:} \quad \tilde{Y}_t^i = (1-\nu_t^i)e^{\left(g(t)+\alpha^i+\beta^i t+z_t^i+\varepsilon_t^i\right)} \tag{2}$$

$$\text{Persistent component:} \quad z_t^i = \rho z_{t-1}^i + \eta_t^i, \tag{3}$$

$$\text{Innovations to AR(1):} \quad \eta_t^i \sim \begin{cases} \mathcal{N}(\mu_{\eta,1}, \sigma_{\eta,1}) & \text{with prob. } p_z \\ \mathcal{N}(\mu_{\eta,2}, \sigma_{\eta,2}) & \text{with prob. } 1-p_z \end{cases} \tag{4}$$

$$\text{Initial condition of } z_t^i: \quad z_0^i \sim \mathcal{N}(0, \sigma_{z,0}) \tag{5}$$

$$\text{Transitory shock:} \quad \varepsilon_t^i \sim \begin{cases} \mathcal{N}(\mu_{\varepsilon,1}, \sigma_{\varepsilon,1}) & \text{with prob. } p_\varepsilon \\ \mathcal{N}(\mu_{\varepsilon,2}, \sigma_{\varepsilon,2}) & \text{with prob. } 1-p_\varepsilon \end{cases} \tag{6}$$

$$\text{Nonemployment duration:} \quad \nu_t^i \sim \begin{cases} 0 & \text{with prob. } 1-p_\nu(t, z_t^i) \\ \min\{1, exp(\lambda)\} & \text{with prob. } p_\nu(t, z_t^i) \end{cases} \tag{7}$$

$$\text{Prob of Nonemp. shock:} \quad p_\nu^i(t, z_t) = \frac{e^{\xi_t^i}}{1+e^{\xi_t^i}}, \text{ where } \xi_t^i \equiv a + bt + cz_t^i + dz_t^i t. \tag{8}$$

In eq. (2), $t = (age - 24)/10$ denotes normalized age and $g(t)$ is a quadratic polynomial of age that captures the lifecycle profile of earnings common to all individuals. The random vector (α^i, β^i) determines ex ante heterogeneity in the level and growth rate of earnings and is drawn from a multivariate normal distribution with zero mean and a covariance matrix to be estimated.[18] The innovations η_t^i to the AR(1) component are drawn from a mixture of two normals. An individual draws a shock from $\mathcal{N}(\mu_{\eta,1}, \sigma_{\eta,1})$ with probability p_z; and otherwise from $\mathcal{N}(\mu_{\eta,2}, \sigma_{\eta,2})$. Without loss of generality, we normalize η to have zero mean (i.e., $\mu_{\eta,1} p_z + \mu_{\eta,2}(1 - p_z) = 0$) and assume $\mu_{\eta,1} < 0$ for identification. Heterogeneity in initial conditions of the persistent process is captured by z_0^i. Transitory shocks, ε_t^i, are also drawn from a mixture of two normals (eq. 6), with similar identifying assumptions (zero mean and $\mu_{\varepsilon,1} < 0$).

Our decision to use of normal mixtures is motivated by two considerations. First, they provide a flexible way to model shock distributions with non-Gaussian properties. In fact, by increasing the number of normals that are mixed one can approximate almost any distribution (see, e.g., McLachlan and Peel (2000)). Second, incorporating normal mixture innovations into a dynamic programming problem requires minimal adjustments to the computational methods commonly used with Gaussian shocks. This is a very appealing feature given our stated objectives.

The final component of the earnings process—and as it turns out, a critical one—is the non-employment shock (eq. 7), which is realized (i.e., $\nu_t > 0$) with probability $p_{\nu t}$ in each period. The duration, ν_t, follows an exponential distribution with mean $1/\lambda$ and is truncated at 1, corresponding to full-year nonemployment (zero annual income). An important difference of ν_t from z_t and ε_t is that it scales the *level* of income rather than its logarithm (as z_t and ε_t do), which allows the process to capture the sizable fraction of workers who transition into full-year nonemployment every year.[19]

None of the stochastic components introduced so far depend explicitly on age or recent earnings, even though the empirical facts revealed substantial variation along these dimensions. One promising way we found to capture such variation is by making p_ν depend on age t and z_t through a logistic function as shown in equation 8.[20] Notice

[18] One possible source of heterogeneity in the growth rate, β^i, could be human capital accumulation in the presence of ability heterogeneity. See, e.g., Huggett *et al.* (2011) and Guvenen *et al.* (2014a).

[19] It takes a –350 log point shock to z_t or ε_t for a worker earning \$50,000 a year to drop below $Y_{\min}$. Forcing z_t or ε_t to generate such large shocks with sufficiently high frequency to match worker exit rates makes it challenging for them to simultaneously match the frequency of smaller shocks, which make up the bulk of the shock distribution. We return to this point later.

[20] Another option would be to make p_z instead of p_ν depend on age and recent earnings, which we explore later. We have also considered a number of other options, among them modifying ξ_t^i by allowing quadratic terms in t or z_t or dependence on α^i and β^i, and introducing age or income dependence in the innovation variances or their mixture probabilities, among others. Results are available upon request.

that the dependence of p_ν on z_t induces persistence in nonemployment from one year to the next, generating what looks like scarring effects of nonemployment even though ν_t itself is independent over time.

This completes the description of the benchmark processes. We also considered a 2-state process, which, as expected, fits the data better but increases the computational burden by adding an extra state variable in a dynamic programming problem. Further details and results from this specification are in Appendix D.3.

Estimation Procedure

We employ the method of simulated moments (MSM) by targeting five sets of moments. The first four broadly correspond to the moments documented in Sections 3 to 5: (i) the standard deviation, skewness, and kurtosis of one- and five-year earnings growth; (ii) moments of the impulse response functions for all RE groups; (iii) average earnings of each LE group over the life cycle (essentially a more detailed version of Figure 11a); and (iv) the cumulative distribution function of nonemployment (Figure 11b).[21] In addition, the age profile of the within-cohort variance of log earnings is a key dimension of the data extensively studied in previous research. For both completeness and consistency with earlier work, we include these variances as a fifth set of moments (Figure D.3). The full list of moments and how they are constructed are reported in Appendix D.1.[22]

In the MSM estimation, we minimize the weighted sum of squared percentage deviations from targeted moments.[23] We chose a weighting matrix that reflects our subjective beliefs on the importance of each set of moments.[24] We employ a global search algorithm,

[21] The targeted moments differ from those in the descriptive analysis in two ways. First, some statistics that were omitted from the descriptive analysis to save space—such as the higher moments of *1-year* changes, levels of average earnings by LE percentiles—are targeted in estimation. Second, as noted in Section 3, using the log earnings change measure requires dropping individuals with very low earnings and in turn losing valuable information in the extensive margin. Therefore, we target the arc-percent analogs of the moments (shown in figures in Sections 3 and 4 of Appendix C.2).

[22] Our approach is in the spirit of Browning *et al.* (2010), Altonji *et al.* (2013), and Guvenen and Smith (2014), among others, who target moments beyond the variance-covariance matrix.

[23] As explained in the appendix, we employ a small adjustment term to ensure that the moments with values close to zero do not have a disproportionate effect on the objective function. The overall effect of this adjustment term is small for the purposes of the point made here.

[24] In particular, our weighting matrix first assigns 15% relative weight to the employment CDF moments. The rest of the moments share the remaining 85% weight according to the following scheme: the cross-sectional moments (standard deviation, skewness, kurtosis) collectively receive a relative weight of 35%, the life-cycle earnings growth moments and impulse response moments each receive a weight of 25%, and the variance of log earnings by age receives a weight of 15%. We have also experimented with a few other weighting matrices that also seemed natural (e.g., giving equal weight to each of the five

that often requires days on large parallel clusters to estimate a specification. Appendix D presents further details of estimation and our numerical method.

6.2 Results: Estimates of Stochastic Processes

In this section, we present the results for eight specifications (Table IV). We start from a simple linear-Gaussian model and add new features step by step until we attain our benchmark process. We discuss along the way which aspect of the data each feature helps capture. Each moment we target answers some concrete and well-defined questions about earnings dynamics (e.g., how skewed are earnings changes, what fraction of individuals work less than 70% of their prime ages, etc). Therefore, we believe, this approach allows researchers to better judge the trade-offs between the importance of matching a particular moment for the questions at hand and the additional complexity it brings.

In **Model (1)** we start with the linear-Gaussian model—the sum of an individual fixed effect, an AR(1) process, and an i.i.d. transitory shock, all drawn from Gaussian distributions (i.e, $\sigma_\beta = 0$, $p_z = 1$, $p_\nu = 0$, and $p_\varepsilon = 1$ in equations (2) to (8)). The estimates of key parameters are unusually large: the standard deviations of the fixed effect and the transitory shock ($\sigma_\alpha = 1.28$ and $\sigma_\varepsilon = 0.76$) are several times larger than what has been found in the previous literature, and the persistence parameter ($\rho = 1.02$) implies a nonstationary process, wherein the effects of shocks are amplified over time (c.f., Storesletten *et al.* (2004) and Heathcote *et al.* (2010b)).

The large magnitudes of the parameters should perhaps not be surprising, as the bulk of the moments targeted here have not been targeted in previous analyses. However, it turns out that one set of moments is responsible for most of these differences—the CDF of lifetime employment rates, which shows that nonemployment is an extremely persistent state for a nonnegligible fraction of men.[25] To match the high fraction of persistently nonemployed individuals, the estimation chooses a wide dispersion of fixed effects, placing many individuals closer to the minimum threshold. Combined with the large transitory shocks and nonstationary persistent shocks, this choice allows the model to match the distribution of lifetime (non)employment in the data very well (Figure 12e).

---

sets of moments or to each *individual* moment) and found substantively similar estimates (results are available upon request). Clearly, with sufficiently large changes in the weighting matrix, one can always get very different estimates.

[25]Reestimating the same process without targeting the employment CDF yields $\sigma_\alpha = 0.51$, $\rho = 1.0$, $\sigma_\eta = 0.156$; and $\sigma_\varepsilon = 0.49$ (see Table D.1 in Appendix D.4). These estimates are similar to those in previous studies cited above, with the exception of σ_ε, which is on the high side.

FIGURE 12 – Estimated Model vs. Data: Key Moments

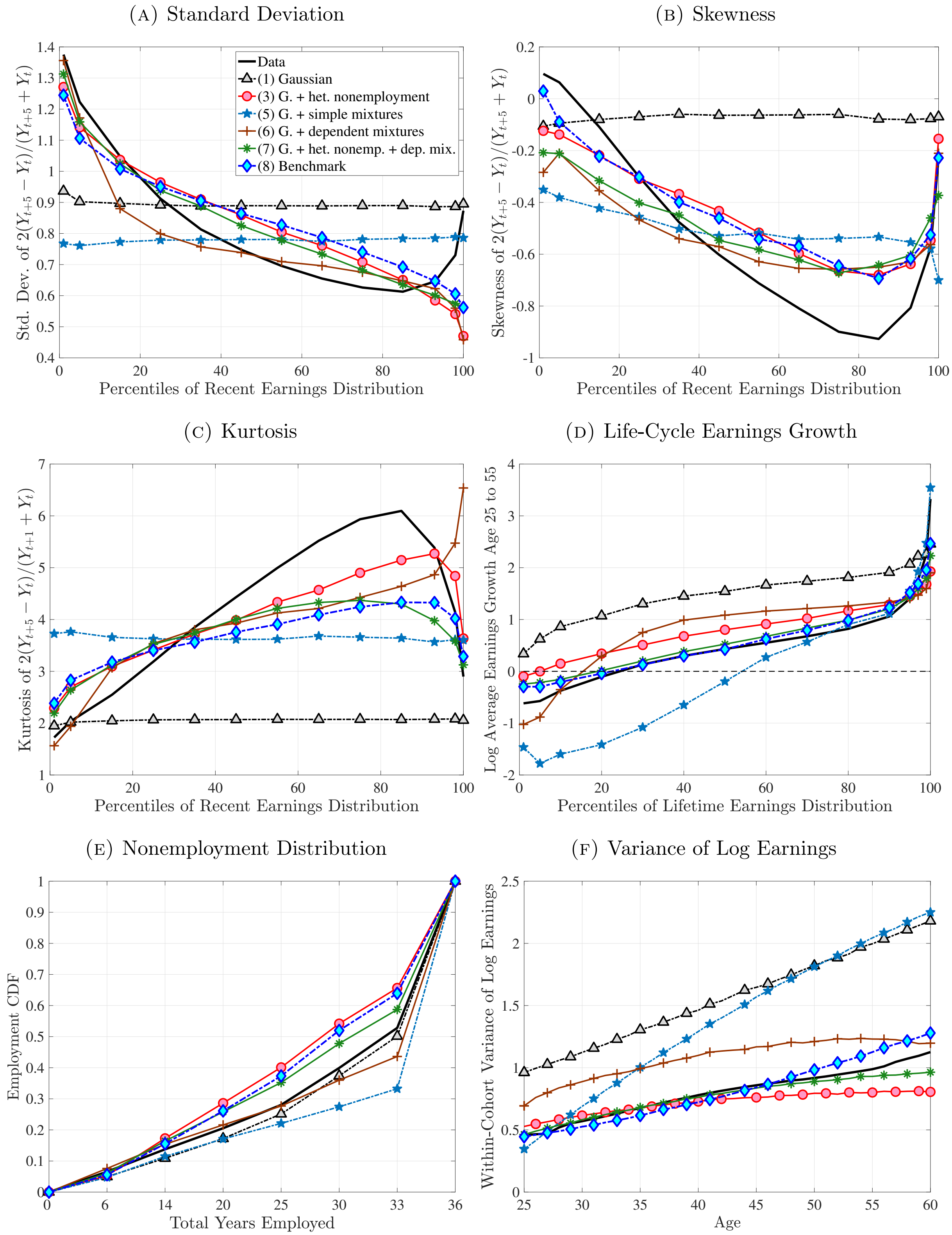


Table IV – Estimates of Stochastic Process Parameters

Model:	(1) Gaussian process	(2)	(3)	(4)	(5)	(6)	(7)	(8) Benchmark
AR(1) Component	G	G	G	mix	mix	mix	mix	mix
$\hookrightarrow$*Probability age/inc.*	—	—	—	no/no	no/no	yes/yes	no/no	no/no
Nonemployment shocks	no	yes	yes	no	no	no	yes	yes
$\hookrightarrow$*Probability age/inc.*	—	no/no	yes/yes	—	—	—	yes/yes	yes/yes
Transitory Shocks	G	G	G	G	mix	mix	mix	mix
HIP	no	no	no	no	no	no	no	yes
Parameters								
ρ	1.017	0.998	0.971	0.998	1.010	0.995	0.983	0.958
p_z				6.3%	5.7%	—†	26.7%	21.9%
$\mu_{\eta,1}$				–1.0*	–1.0*	–1.0*	–0.194	–0.147
$\sigma_{\eta,1}$	0.102	0.158	0.210	1.487	1.275	1.098	0.444	0.457
$\sigma_{\eta,2}$				0.031	0.020	0.108	0.076	0.139
$\sigma_{z_1,0}$	0.128	0.102	0.390	0.360	0.101	0.395	0.495	0.667
λ		0.690	0.070				0.044	0.001
p_ε					12.4%	6.8%	9.2%	12.6%
$\mu_{\varepsilon,1}$					–0.559	0.387	0.352	0.236
$\sigma_{\varepsilon,1}$	0.758	0.351	0.135	0.358	1.433	0.831	0.294	0.343
$\sigma_{\varepsilon,2}$					0.021	0.048	0.065	0.063
σ_α	1.284	1.288	0.594	1.066	0.334	0.493	0.467	0.298
$\sigma_\beta \times 10$								0.185
$\text{corr}_{\alpha\beta}$								0.976
Objective value	76.70	74.95	32.76	58.52	54.60	37.67	26.25	22.29
Decomposition:								
(i) Standard deviation	10.51	9.36	6.80	7.79	7.81	6.45	5.66	5.48
(ii) Skewness	44.01	39.92	15.16	21.60	22.54	15.63	14.47	10.34
(iii) Kurtosis	27.60	14.65	5.85	17.97	13.08	8.83	5.03	5.97
(iv) Impulse response	28.29	34.31	22.33	36.87	33.06	26.98	16.77	13.32
(v) Lifetime inc. growth	41.00	46.33	11.04	26.79	25.55	12.85	9.75	8.29
(vi) Within cohort ineq.	23.25	19.60	5.44	18.75	18.65	11.96	1.95	3.00
(vii) Nonemployment CDF	7.18	3.70	10.61	12.06	12.13	4.34	6.51	8.30

Notes: The top panel provides a summary of the features of each specification, the middle panel shows the estimated values of key parameters that are discussed in the main text (the rest are reported in Table D.2), and the bottom panel reports the weighted percentage deviation between the data and simulated moments for each set of moments (the total objective value is the square root of sum of squares of objective values of each component). The standard errors for the benchmark process are reported in Table D.1. The *'s indicate that in columns 4, 5, and 6 the value of $\mu_{\eta,1}$ is constrained by the lower bound we impose in the estimation. † : p_z is not a number but a function in this specification.

However, the model fails in most of the other dimensions. First, it generates nearly zero skewness and no excess kurtosis in income changes, which is not surprising given the Gaussian structure (Figure 12b and 12c). Second, it vastly overstates lifecycle income growth rates for all LE groups—for example, implying a 4.6-fold rise for the median worker compared with only a 60% rise in the data (Fig. 12d). Finally, it overshoots both the level of income inequality at age 25 as well as its rise over the life cycle (Fig. 12f).[26] Clearly, this process does not offer a good fit to these key features of the data.

Introducing Nonemployment Shocks

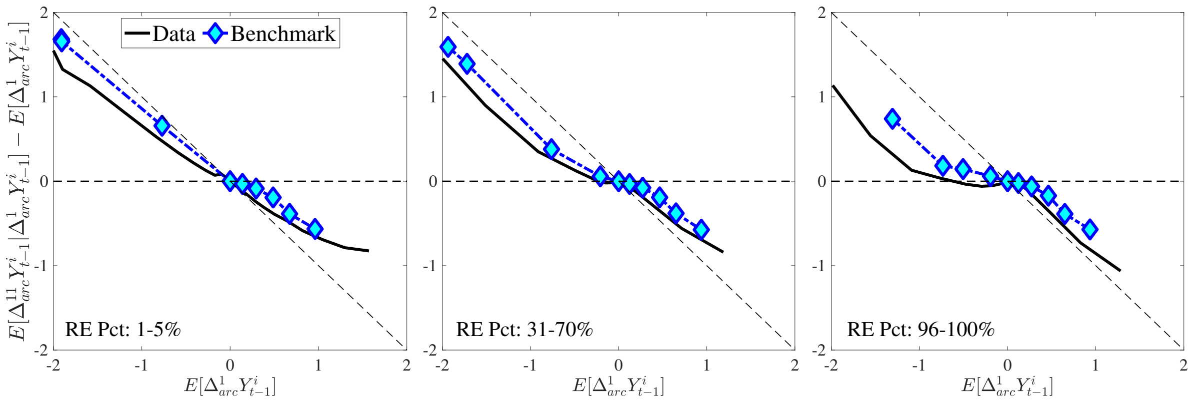
We first consider introducing nonemployment shocks (ν_t) to improve the fit by allowing the model to match the employment CDF more easily and leaving more flexibility for matching other moments. Another motivation is to investigate a common conjecture that the negative skewness of earnings changes could be entirely due to unemployment (disaster) shocks. Finally, unemployment shocks are a common feature in quantitative models, so it is instructive to understand what they add to the fit relative to the simple Gaussian model. We consider two versions. First, in Model (2), we restrict the nonemployment risk to be uniform across workers (i.e., $b, c, d \equiv 0$ in eq. 8). Second, in Model (3), we remove the restriction on heterogeneity and estimate a, b, c, and d.

The estimates from the first specification (**Model (2)**) imply that 2.3% of workers are hit with a nonemployment shock each year ($p_\nu = 0.023$, not reported in the table) and about half of those experience full-year nonemployment (implied by $\lambda = 0.69$). The other estimates remain similar to the linear-Gaussian model, except for a large decline in the size of transitory shocks ($\sigma_{\varepsilon,1}$ falling from 0.76 in Model (1) to 0.35 here) and a modest decline in the persistence of the AR(1) process. Overall, the fit improves marginally (75.0 vs 76.7 before). Further, as the breakdown at the bottom of the table shows, the fit to the skewness moment improves only slightly, with larger gains coming from the kurtosis and nonemployment CDF moments.

A key factor driving these results is the uniform nonemployment risk, under which nonemployment spells are basically large but *fully* transitory shocks: every worker, whose income goes down due to nonemployment in the current period, bounces back to his previous income level (on average) in the next period. Consequently, they stretch both tails of the income change distribution, leaving its symmetry unaffected. The longer tails generate some excess kurtosis in annual earnings growth (less so over longer horizons) and a longer left tail of the lifetime nonemployment CDF, which improves the fit in

[26] Notice that the variance of log earnings at age 25 (100 log points) is less than half what would be predicted from the estimated parameters, the sum of σ_α^2 and σ_ε^2, or 225 log points. This is because the variance in simulated data is computed excluding workers below the minimum threshold; and with the large variance of shocks there are many such individuals.

FIGURE 13 – Estimated Model vs. Data: Selected Impulse Response Moments



Notes: In the estimation we target arc-percent earnings growth between $t+k$ and $t-1$, which allows us to keep the composition of workers constant for each k (for details see Appendix D.1). To keep concepts analogous to what is shown in Section 4, where our focus is on income change between t and $t+k$, we plot $\mathbb{E}[\Delta^{k+1}_{arc} Y^i_{t-1} \mid \Delta^1_{arc} Y^i_{t-1}] - \mathbb{E}[\Delta^1_{arc} Y^i_{t-1}])$ for $k = 10$.

these dimensions. Nevertheless, this specification still misses out completely on income or age variation in the cross-sectional moments, and implies an unusually large initial heterogeneity ($\sigma_\alpha = 1.29$). These results show that transitory shocks—regardless of whether they are modeled as a shock to the level or log of earnings (ν_t or ε_t)—cannot account for the left-skewness in earnings growth, as they miss the key features of the left tail shocks in the data, which are that they are *recurring* and have *scarring* effects (see Jacobson *et al.* (1993); Von Wachter *et al.* (2009); Guvenen *et al.* (2017)).

Motivated by this observation, in **Model (3)**, we allow nonemployment shocks to vary by age and z_t. This change improves the performance of the model dramatically: the objective value falls from 75.0 to 32.8, with improved fit in all sets of moments except for the nonemployment CDF. The improvements are apparent in Figure 12, which also shows that the model captures the variation in cross-sectional moments by RE levels fairly well.[27] Perhaps somewhat surprisingly, the model also better fits the lifecycle earnings growth moments. Furthermore, this model now understates the rise in within-cohort inequality rather than overstating it substantially. This is especially surprising given that the estimated model also matches the large standard deviation of earnings changes. It has been a challenge to jointly match moments in earnings levels and differences (see Heathcote *et al.* (2010a) and Daly *et al.* (2016)).

Model (3) yields these improvements by not pushing workers into nonemployment

[27]Because the age variation is much smaller than the RE variation in the data, we omit the age patterns here to save space, but the fit improves along this dimension as well (see Figures D.5 and D.6).

through very large shocks, as under the previous specifications, but instead by generating systematic nonemployment that is concentrated among certain groups of workers. Consequently, it is able to fit the data with a smaller dispersion of fixed effects ($\sigma_\alpha = 0.594$) and of transitory shocks ($\sigma_\varepsilon = 0.135$) and with a lower persistence ($\rho = 0.97$). We conclude that heterogeneous nonemployment risk is a promising ingredient for an earnings process that aims to reproduce a broad set of moments of earnings levels and changes.[28]

Introducing Normal Mixtures

In the next step, we investigate the potential of modeling z_t and ε_t as normal mixtures. To see their effects more clearly, we shut down nonemployment risk for the time being ($p_\nu \equiv 0$). **Model (4)** allows a mixture in the persistent component, z_t, whereas **Model (5)** allows it in both components. As before, we begin by restricting the mixture probabilities to be the same for all individuals.[29]

The fit improves significantly relative to Model (1), and most of it comes from adding the mixture component in z (comparing Model (4) to Model (5)), which improves the fit on skewness, kurtosis, and lifetime income growth moments.[30] In both models, one of the shocks to the persistent component looks like a rare disaster shock: it has a low probability of around 6%, with a large negative mean of –1 (the lower bound imposed) and a large standard deviation bigger than 1.25.[31] Similarly, the mixture in transitory shocks also displays rare large shocks, with a probability of 12%, a mean of –0.55, and a standard deviation of 1.43. Thus, allowing for a non-Gaussian ε_t has very little impact on the estimated nonnormality of η_t and its importance for the overall estimation.

Not surprisingly, with constant mixture probabilities (p_z and p_ε) these models do not capture the age and income variation in higher-order moments. To improve the fit,

[28] Krusell *et al.* (2011) also show that persistent idiosyncratic productivity shocks play a key role in matching the persistence of the employment and out-of-the-labor-market states found in individual labor market histories in the CPS data. Two recent papers attribute an important role to heterogeneity in unemployment risk. Jarosch (2017) finds job security to be important in accounting for scarring effects of unemployment. Karahan *et al.* (2018) find differences in unemployment risk to be important in explaining lifetime earnings differences among the bottom half of the distribution.

[29] Nonemployment shocks can be mimicked by the mixture of normals if one of them has a very large negative mean and a small variance. To keep the distinction between two modeling tools clear, we impose a lower bound on the smaller mean of both mixture distributions ($\mu_{\eta,1}$ and $\mu_{\varepsilon,1}$). The lower bound is set to –1 (or –63% mean shock) but we also estimated with a bound of –5 (or –99.3%). The objective value for the latter is 47.9, down from 54.6.

[30] We have estimated a specification that adds uniform nonemployment risk to Model (5) and found its probability to be almost zero (0.01%).

[31] As you will see, in all models discussed below (and in virtually all of our explorations not included in the paper) we find very strong left skewness and excess kurtosis in persistent innovations.

Model (6) models p_z with the same functional form used before for p_ν (eq. 8) but with ξ_t^i is replaced with ξ_{t-1}^i. Now, the model matches better the age and income variation in higher-order moments, except for kurtosis at the top of the RE distribution (Figure 12).

Next, in **Model (7),** we combine the two most promising features we found so far: the normal mixture specifications of z_t and ε_t of Model (5) with the heterogeneous nonemployment risk of Model (3).[32] The resulting income process lacks only the HIP from the benchmark specification defined by equations (2) to (7). Relative to Model (6), this combined model improves the fit across the board, but especially so for the impulse response moments. In other words, while the cross-sectional moments can be generated to a large extent with either heterogeneous nonemployment risk or mixture of normals, the nonlinear earnings persistence revealed by impulse responses requires both components together. In particular, the extreme earnings changes followed by moderate mean reversion (as was shown in Figure 10) cannot be explained by fully transitory or permanent shocks, whereas nonemployment shocks are better able to generate this pattern with their long-lasting but nonpermanent effects.

Introducing Heterogeneous Income Profiles

Finally, we turn to the benchmark model, which adds a HIP component to Model (7). This leads to a nonnegligible decline in the objective value driven by a better fit to the skewness and impulse response moments. The improvement mainly arises from a lower persistence for z_t ($\rho = 0.958$ vs. 0.983), with the half-life of η_t shocks declining to 16 years from 40 in Model 7. This allows the process to better capture the persistence patterns of large earnings changes just described (Figure 13).[33] This process also offers a decent fit to the age variation in cross-sectional moments (Figures D.7 and D.8).

Overall, we believe the benchmark process offers a reasonable trade-off between a good fit to the data and the need for a parsimonious process that can be implemented in models without increasing the computational complexity.[34] Nonetheless, one could make

[32] We also explored an alternative case in which nonemployment risk is uniform but the mixture probability of persistent shocks is allowed to vary with age and z. Model (7) outperforms this specification.

[33] The fact that adding HIP to a process lowers the estimated ρ is well understood. Basically, the age profile of the within-cohort variance of earnings is close to linear, which is generated by an AR(1) process only if $\rho = 1$. Without HIP, this force pushes up the estimate of ρ relative to what is implied by the mean reversion patterns in the data. HIP provides extra flexibility by generating the linear age profile, allowing ρ to fall to values more consistent with other moments (Guvenen (2009)).

[34] A feature that has found empirical support among other researchers is heterogeneity in innovation variances. Although we have not explicitly modeled that feature, the variation in mixture probabilities by income and age naturally gives rise to heterogeneity in these variances (similar to introducing ARCH

an argument in favor of Model (7) as well. An even simpler process with a reasonable fit is Model (3), which does not have the normal mixture in z_t. But the computational burden introduced by a normal mixture is often very small, so in our subjective opinion, the trade-off present in Model (3) is not likely to be worth it in many applications.

6.3 Parameter Estimates of the Benchmark Process

We now turn to the parameter estimates from the benchmark process and its fit to some untargeted moments. Starting with the AR(1) process, the persistent shock is drawn about every five years ($p_z = 21.9\%$) from an "unfavorable" distribution—with a negative mean and large standard deviation ($\mu_{\eta,1} = -0.15$ and $\sigma_{\eta,1} = 0.46$)—and in the other years from a "favorable" one—with a positive mean and small standard deviation ($\mu_{\eta,2} = 0.041$, $\sigma_{\eta,2} = 0.14$). This mixture of normals implies that innovations to the persistent component are both strongly left skewed (skewness of -1.10) and leptokurtic (kurtosis around 8). In contrast, transitory shocks (ε_t) are typically smaller: in most years (with 87% probability), they are drawn from a tight distribution, $\mathcal{N}(-0.034, 0.063^2)$; and every eight years or so ($p_\varepsilon = 12.6\%$) from a distribution with large positive mean and dispersion, $\mathcal{N}(0.236, 0.343^2)$. Consequently, transitory shocks feature a skewness of 2.75 and a kurtosis of 15.4. However, η_t and ε_t are not the only sources of higher-order moments in our model; workers also face nonemployment risk, which is another key source of (age- and income-varying) skewness and kurtosis in *persistent* earnings changes. Thus, as expected, Models (4)-(6) (without nonemployment risk) find persistent shocks to be more strongly leptokurtic and left skewed than Model (8).[35] Thus, we conclude that persistent innovations are key drivers of non-Gaussian features in the data.

Regarding initial heterogeneity, it is captured by (i) the permanent fixed effect σ_α (since $\beta^i \times 0 = 0$), and (ii) σ_{z_0}, the dispersion in the initial conditions of the z_t process, whose effect declines at rate ρ. Since the dispersion in z_0^i is about twice that of α^i ($\sigma_{z_0} = 0.67$ versus $\sigma_\alpha = 0.30$), a large part of the wage dispersion observed at age 25 reflects persistent but not permanent differences.

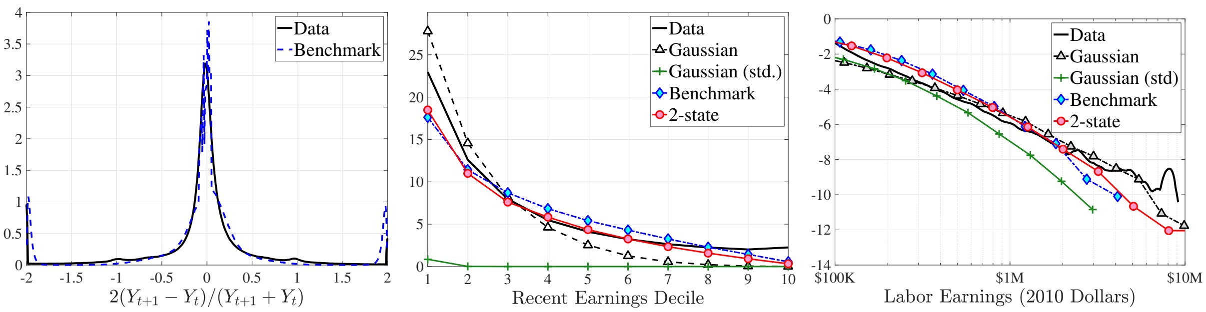
Finally, our estimate $\sigma_\beta \simeq 0.02$ is close to earlier estimates from the PSID despite different data sets and targets. However, the estimated correlation here is 0.98, rather

effects as in Meghir and Pistaferri (2004)). We found that allowing for additional ex ante heterogeneity as in Browning *et al.* (2010) does not improve the fit on the targeted moments.

[35] In fact, in Model (5), in which the only sources of nonnormalities are the persistent and transitory innovations, we find persistent innovations to be more left-skewed and leptokurtic than transitory ones.

Figure 14 – Model Fit: Non-targeted Statistics

(A) Density of Arc-Percent Change (B) Fraction Nonemployed in t, % (C) Log Density of Earnings



Notes: The parameters of the Gaussian process are presented in column (1) of Table IV, whereas Gaussian (std.) is the same process being estimated without targeting the employment CDF (see Table D.1 for its parameter estimates). The data series on Panel (B) is conditional on past 2 year income (in $t-1$ and $t-2$).

than close to zero in those studies (c.f. Haider (2001) and Guvenen and Smith (2014)).

Persistent Nonemployment Shocks Virtually all workers hit by nonemployment shocks experience *full-year* nonemployment ($\lambda = 0.001$). These shocks, together with the non-Gaussian η_t and ε_t shocks, allow the model to generate a leptokurtic density for arc-percent changes with spikes at both ends (figure 14a). To the extent that these shocks are realized with a non-negligible probability, they can have important effects on a worker's earnings dynamics. This brings up the question of how often these shocks are realized, to which we now turn.

The coefficients of the probability function in (8) are hard to interpret on their own (see Figure D.11 for its 3D graph), so instead, we investigate the probability of drawing a nonemployment shock for various age and RE percentile groups for workers who satisfy the conditions of the base RE sample. Starting with the age variation, nonemployment risk declines modestly over a working life, from 7.1% over 25–34 to 6.1% in the next 10 years, and to 5.5% over 45–54. There is less variation in the data with the corresponding figures being 6.8%, 6.2%, and 6.6%, respectively.

Differences in nonemployment risk between income groups are much more pronounced (Figure 14b). In the bottom RE decile, workers experience full-year nonemployment almost once every six years (with 17.8% probability). This probability declines sharply over the RE distribution to 4.9% for the median quintile and to 0.6% for the top decile. The fit to the data is fairly good, though the variation with respect to RE is slightly less pronounced compared with the data.[36] Perhaps surprisingly, the linear-Gaussian

[36] These patterns are also consistent with the evidence from the Survey of Income and Program

Table V – Persistence of Nonemployment Risk, 1-State Benchmark Process

	25-35		**36–45**		**46–55**	
Nonemp. at $t \rightarrow$	$t+1$	$t+5$	$t+1$	$t+5$	$t+1$	$t+5$
RE Groups						
1–10	0.534	0.476	0.570	0.490	0.608	0.513
41–60	0.460	0.429	0.548	0.477	0.609	0.504
$91-100$	0.385	0.385	0.530	0.467	0.596	0.496

income process also matches this feature of the data reasonably well. As noted in the earlier discussion, however, this specification's ability to capture the nonemployment CDF comes at the cost of an implausibly steep inequality profile (Figure 12f). Under a more plausible parameterization, less than 1% of individuals ever experience a full-year of nonemployment over their working life, a result severely at odds with the data.

An important feature of the benchmark process is the dependence of nonemployment probability on z_t. A negative innovation to the persistent component increases not only the nonemployment risk in the current period, but also its future incidence. Thus, the autocorrelation of shocks is not precisely captured by ρ, and nonemployment shocks are autocorrelated even though v_t is drawn in an i.i.d. manner. This feature, along with the normal mixture innovations to persistent and transitory components, contributes to the asymmetric mean reversion documented in the impulse response analysis (Figure 13).

To illustrate the persistence of nonemployment, we examine nonemployment rates in $t+1$ and $t+5$ for workers who experience nonemployment in t. As usual, we further condition workers on their RE in $t-1$. Nonemployment is fairly persistent overall and more so for low-income workers: In the first 10 years of the working life, around 48% of the workers in the bottom decile of the RE distribution experience another nonemployment spell after five years of the initial nonemployment (Table V). This number declines monotonically over the RE distribution to 39% for the top decile. Furthermore, nonemployment risk becomes more persistent over a working life, particularly for higher-income workers. For example, in the top decile, the conditional probability of nonemployment in $t+5$ increases from 39% in the first 10 years to 50% in the last 10.

An important feature of the data not targeted in the estimation is income concentration at the top. We consider the log density of earnings (levels, not changes) in the data for 45–49-year-olds, along with the counterparts from the benchmark specification and

Participation on income and age variation in unemployment risk (see Karahan *et al.* (2018)).

the Gaussian process (Figure 14c). While the benchmark process provides an accurate fit to the earnings distribution under the \$1 million mark, it fails to capture the density of higher levels of income in the data. On this front, the Gaussian process surprisingly generates a longer Pareto tail, which is due to its extreme parameter values, as discussed in Section 6.2. In contrast, the standard parameterization of the same process vastly understates the number of individuals with very high earnings.

A 2-State Specification. Our benchmark process contains a single AR(1) and can be easily implemented in quantitative models. We have investigated how much the fit can be improved by a process with two AR(1) components by extending our benchmark specification to allow for an additional AR(1) component, and letting the mixture probabilities vary with ξ_{t-1}^i. The resulting process provides a significantly better fit to the targeted moments and matches top income inequality as well as the income variation in nonemployment risk (Figures 14b and 14c). See Appendix D.3 and D.4 for more details.

We find that the two AR(1) components are quite different from each other, especially in terms of their persistence with $\rho_2 = 0.98$ vs. $\rho_1 = 0.82$. The composition of large negative shocks changes from (hard-to-insure) more persistent innovations to the less persistent ones over the life cycle. The probability of receiving at least one large shock to one of the two AR(1) components or a nonemployment shock in a given year is declining in recent earnings, ranging from 53% at the low end to 16% for individuals above the 90th percentile. Finally, the age and income variation of nonemployment risk in the 2-state specification is similar to that in the benchmark process.

7 Concluding Thoughts

Using earnings histories of millions of U.S. workers, we have studied non-Gaussian and nonlinear earnings dynamics and reached the following conclusions: First, the distribution of earnings growth is not symmetric but left skewed; it is leptokurtic in that most individuals experience very small changes, while changes for a small but nonnegligible number are extremely large. Critically, these features vary substantially over the life cycle and across the earnings distribution: Higher-income and older workers on average face earnings growth that is more left-skewed and leptokurtic. Finally, earnings changes display asymmetric persistence: Increases for high-earning individuals are quite transitory, whereas declines are very persistent; the opposite is true for low earners.

After establishing these facts nonparametrically, we estimated an earnings process that is broadly consistent with these features of the data.[37] This specification allows for normal mixture innovations to the persistent and transitory components and a long term nonemployment shock with a realization probability that varies with age and earnings. We found that this feature generates systematic recurring nonemployment with long-term scarring effects, which is important to match the data.

Our empirical findings are broadly consistent with job ladder models. In these models, workers in most years see little change in their earnings but once in a while experience a large change due to an unemployment shock, a job switch, or an outside offer, which in turn leads to leptokurtic changes. Furthermore, the scarring effects of falling off the job ladder can generate left skewness. However, the variation in earnings dynamics over the life cycle and by recent earnings are so large that it is an open question whether existing models can be quantitatively consistent, and, if not, how they should be modified. Recently, Hubmer (2018) and Karahan *et al.* (2018) explore the ability of various frameworks to generate some of these patterns.

A natural question is what non-Gaussian features imply for the consumption-savings behavior of households. As a first approximation, consider the following well-known thought experiment: A household with constant relative risk aversion preferences (with a coefficient of 10) pays a risk premium π to avoid a gamble that changes consumption by a random proportion $(1+\tilde{\delta})$. We compare two scenarios. First, $\tilde{\delta}^A$ is drawn from a Gaussian distribution with a mean of zero and a standard deviation of 0.10, for which the risk premium is 4.9%. Second, $\tilde{\delta}^B$ has the same first two moments as $\tilde{\delta}^A$, but it has a skewness of -2 and a kurtosis of 30 (roughly corresponding to the one-year earnings change of a 45-year-old at the 90th RE percentile), which amplifies π to 22.2%. Interestingly, excess kurtosis plays a more important role in this amplification. A leptokurtic but symmetric risk increases π to 18.8%, compared with 6.3% for a left skewed but mesokurtic risk.

Of course, our calculations are only suggestive rather than conclusive. Incorporating higher-order moments of earnings dynamics into quantitative models is still in its infancy. Recently, De Nardi *et al.* (forthcoming) examine the consumption-savings behavior in

[37] The nonparametric empirical facts documented in Sections 3 to 5 (along with those reported in the appendix) add up to more than 10,000 empirical moments of individual earnings data. Adding analogous moments for women, as mentioned in footnote 1, doubles this number. The richness of this information is far beyond what we are able to fully utilize in the estimation exercise in this paper. Furthermore, for different questions, it would make sense to focus on a subset of these moments that are different from what we have aimed for in this paper. With these considerations in mind, we make these detailed moments available online for download as an Excel file.

the presence of non-Gaussian income risk. Guvenen *et al.* (2018) analyzes a life-cycle incomplete market model with the stochastic process we estimate in Section 6. Building on the evidence in Guvenen *et al.* (2014b), Constantinides and Ghosh (2016) and Schmidt (2016) show that an asset pricing model with incomplete markets and procyclical skewness generates plausible asset pricing implications. And McKay (2014) studies cyclical consumption dynamics of these skewness fluctuations. Targeting the moments documented in this paper, Golosov *et al.* (2016) show that a process with negative skewness and excess kurtosis implies a substantially higher optimal top marginal tax rate on earnings compared with a traditional Gaussian calibration. Similarly, Kaplan *et al.* (2016) introduce leptokurtic idiosyncratic risk to generate a realistic asset portfolio distribution in a New Keynesian model. We hope our findings will further feed back into economic research and policy analyses.

References

Aaronson, D. and French, E. (2004). The effect of part-time work on wages: Evidence from the social security rules. *Journal of Labor Economics*, **22** (2), 329–252.

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

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

Bagger, J., Fontaine, F., Postel-Vinay, F. and Robin, J.-M. (2014). Tenure, experience, human capital and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, **104** (6), 1551–1596.

Baker, M. and Solon, G. (2003). Earnings dynamics and inequality among Canadian men, 1976–1992: Evidence from longitudinal income tax records. *Journal of Labor Economics*, **21** (3), 289–321.

Blass-Hoffmann, E. and Malacrino, D. (2017). *Employment Time and the Cyclicality of Earnings Growth*. Tech. rep.

Bonhomme, S. and Robin, J. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies*, **77** (2), 491–533.

— and Robin, J.-M. (2009). Assessing the equalizing force of mobility using short panels: France, 1990 a 2000. *Review of Economic Studies*, **76** (1), 63–92.

Bound, J., Brown, C. and Mathiowetz, N. (2001). Measurement error in survey data. In J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, *59*, Elsevier, pp. 3705–3843.

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

Constantinides, G. M. and Ghosh, A. (2016). Asset pricing with countercyclical household consumption risk. *Journal of Finance*.

Crow, E. L. and Siddiqui, M. (1967). Robust estimation of location. *Journal of the American Statistical Association*, **62** (318), 353–389.

Cunha, F., Heckman, J. and Navarro, S. (2005). Separating uncertainty from heterogeneity in life cycle earnings. *Oxford Economic Papers*, **57** (2), 191–261.

Daly, M., Hryshko, D. and Manovskii, I. (2016). *Improving the Measurement of Earnings Dynamics*. Tech. rep., National Bureau of Economic Research.

Davis, S. J., Haltiwanger, J. C. and Schuh, S. (1996). *Job Creation and Destruction*. Cambridge, MA: MIT University Press.

De Nardi, M., Fella, G. and Pardo, G. P. (forthcoming). *Nonlinear Household Earnings Dynamics, Self-Insurance, and Welfare*. Tech. rep.

Deaton, A. and Paxson, C. (1994). Intertemporal choice and inequality. *Journal of Political Economy*, **102** (3), 437–67.

Dobkin, C., Finkelstein, A., Kluender, R. and Notowidigdo, M. J. (2018). The economic consequences of hospital admissions. *American Economic Review*, **108** (2), 308–52.

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

Golosov, M., Troshkin, M. and Tsyvinski, A. (2016). Redistribution and social insurance. *American Economic Review*, **106** (2), 359–386.

Guvenen, F. (2009). An Empirical Investigation of Labor Income Processes. *Review of Economic Dynamics*, **12** (1), 58–79.

—, Karahan, F. and Ozkan, S. (2018). *Inequality over the Life Cycle in the Presence of Higher-Order Risk*. Working paper, University of Minnesota.

—, —, — and Song, J. (2017). Heterogeneous scarring effects of full-year nonemployment. *American Economic Review*, **107** (5), 369–73.

—, Kuruscu, B. and Ozkan, S. (2014a). Taxation of Human Capital and Wage Inequality: A Cross-Country Analysis. *Review of Economic Studies*, **81**, 818–850.

—, Ozkan, S. and Song, J. (2014b). The Nature of Countercyclical Income Risk. *Journal of Political Economy*, **122** (3), 621–660.

— and Smith, A. A. (2014). Inferring Labor Income Risk and Partial Insurance from Economic Choices. *Econometrica*, **82** (6), 2085–2129.

Haider, S. J. (2001). Earnings instability and earnings inequality of males in the United States: 1967-1991. *Journal of Labor Economics*, **19** (4), 799–836.

Halvorsen, E., Holter, H. A., Ozkan, S. and Storesletten, K. (2018). Dissecting idiosyncratic income risk.

Heathcote, J., Perri, F. and Violante, G. L. (2010a). Unequal We Stand: An Empirical Analysis of Economic Inequality in the United States: 1967-2006. *Review of Economic Dynamics*, **13** (1), 15–51.

—, Storesletten, K. and Violante, G. L. (2010b). The macroeconomic implications of rising wage inequality in the united states. *Journal of Political Economy*, **118** (4), 681–722.

Horowitz, J. J. and Markatou, M. (1996). Semiparametric estimation of regression models for panel data. *Review of Economic Studies*, **63** (1), 145–168.

Hubmer, J. (2018). The job ladder and its implications for earnings risk. *Review of Economic Dynamics*, **29**, 172–194.

Huggett, M., Ventura, G. and Yaron, A. (2011). Sources of lifetime inequality. *American Economic Review*, **101** (7), 2923–54.

Jacobson, L. S., LaLonde, R. J. and Sullivan, D. G. (1993). Earnings losses of displaced workers. *American Economic Review*, **83** (4), 685–709.

Jarosch, G. (2017). *Searching for Job Security and the Consequences of Job Loss*. Working paper.

Kaplan, G., Moll, B. and Violante, G. L. (2016). *Monetary Policy According to HANK*. Working paper, Princeton University.

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

—, — and Song, J. (2018). *Sources of Inequality in Earnings Growth Over the Life Cycle*. Working paper.

Kelley, T. L. (1947). *Fundamentals of Statistics*. Harvard University Press.

Krusell, P., Mukoyama, T., Rogerson, R. and Şahin, A. (2011). A three state model of worker flows in general equilibrium. *Journal of Economic Theory*, **146** (3), 1107–1133.

Kurmann, A. and McEntarfer, E. (2018). *Downward Wage Rigidity in the United States: New Evidence from Administrative Data*. Working paper.

Lillard, L. A. and Weiss, Y. (1979). Components of variation in panel earnings data: American scientists 1960-70. *Econometrica*, **47** (2), 437–454.

— and Willis, R. J. (1978). Dynamic aspects of earnings mobility. *Econometrica*, **46** (5), 985–1012.

Lise, J. (2012). On-the-job search and precautionary savings. *Review of Economic Studies*, **80** (3), 1086–1113.

Low, H., Meghir, C. and Pistaferri, L. (2010). Wage risk and employment risk over the life cycle. *American Economic Review*, **100** (4), 1432–1467.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics*, **18** (1), 83–114.

McKay, A. (2014). *Time-Varying Idiosyncratic Risk and Aggregate Consumption Dynamics*. Working paper, Boston University.

McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley and Sons.

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

Moors, J. J. A. (1986). The meaning of kurtosis: Darlington reexamined. *The American Statistician*, **40**, 283–284.

— (1988). A quantile alternative for kurtosis. *Journal of the American Statistical Association. Series D (The Statistician)*, **37** (1), 25–32.

Olsen, A. and Hudson, R. (2009). Social security administration's master earnings file: Background information. *Social Security Bulletin*, **69** (3), 29–46.

Panis, C., Euller, R., Grant, C., Bradley, M., Peterson, C. E., Hirscher, R. and Stinberg, P. (2000). *SSA Program Data User's Manual*. Baltimore, MD: Social Security Administration.

Pareto, V. (1897). *Cours d'Economie Politique*. Geneva: Droz.

Pistaferri, L. (2003). Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply. *Journal of Labor Economics*, **21** (3).

Schmidt, L. D. W. (2016). *Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk*. Working paper, University of Chicago.

Storesletten, K., Telmer, C. I. and Yaron, A. (2004). Consumption and risk sharing over the life cycle. *Journal of Monetary Economics*, **51** (3), 609–633.

Topel, R. H. and Ward, M. P. (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics*, **107** (2), 439–79.

Von Wachter, T., Song, J. and Manchester, J. (2009). *Long-Term Earnings Losses due to Mass Layoffs During the 1982 Recession: An Analysis Using U.S. Administrative Data from 1974 to 2004*. Tech. rep., Columbia University.

Zhang, H., Conn, A. R. and Scheinberg, K. (2010). A derivative-free algorithm for least-squares minimization. *SIAM Journal on Optimization*, **20** (6), 3555–3576.

Supplemental Online Appendix

NOT FOR PUBLICATION

A Data Appendix

Constructing a nationally representative panel of males from the MEF is relatively straightforward. The last four digits of the SSN are randomly assigned, which allows us to pick a number for the last digit and select all individuals in 1978 whose SSN ends with that number.[38] This process yields a 10% random sample of all SSNs issued in the United States in or before 1978. Using SSA death records, we drop individuals who are deceased in or before 1978 and further restrict the sample to those between ages 25 and 60. In 1979, we continue with this process of selecting the same last digit of the SSN. Individuals who survived from 1978 and who did not turn 61 continue to be present in the sample, whereas 10% of new individuals who just turn 25 are automatically added (because they will have the last digit we preselected), and those who died in or before 1979 are again dropped. Continuing with this process yields a 10% representative sample of U.S. males in every year from 1978 to 2013.

The measure of wage earnings in the MEF includes all wages and salaries, tips, restricted stock grants, exercised stock options, severance payments, and many other types of income considered remuneration for labor services by the IRS as reported on the W-2 form (Box 1). This measure does not include any pre-tax payments to IRAs, retirement annuities, independent child care expense accounts, or other deferred compensation. We apportion 2/3 of the self-employment income as labor income. Given the lack of direct data on this, the 2/3 allocation has been the convention adopted by the literature as well as the PSID. In a previous version we ignored self-employment income altogether and found similar results, leading us to believe that the exact allocation matters very little.

Finally, the MEF has a small number of extremely high earnings observations. For privacy and confidentiality reasons, we cap (winsorize) observations above the 99.999th percentile of the year-specific income distribution. For background information and detailed documentation of the MEF, see Panis *et al.* (2000) and Olsen and Hudson (2009).

Recent Earnings. Average income of a worker i between years $t-1$ and $t-5$ is given by $\hat{Y}^i_{t-1} = \frac{1}{5}\sum_{j=1}^{5} \max\left\{\tilde{Y}^i_{t-j}, Y_{\min,t}\right\}$, where $\tilde{Y}^i_t$ denotes his earnings in year t. We then control for age and year effects by regressing $\hat{Y}^i_{t-1}$ on age dummies separately for each year, and define the residuals as **recent earnings** (RE), $\bar{Y}^i_{t-1}$. In Sections 3 and 4, we grouped individuals by age and by $\bar{Y}^i_{t-1}$ to investigate how the dynamics of income vary over the life cycle and by income levels.

Table A.1 shows some sample size statistics regarding the sample used in the cross-sectional moments. Recall that we compute these statistics for each age-year-RE percentile and aggregate them across years. Therefore, sample sizes refer to the sum across all years of a given age by percentile group. Each row reports the median, minimum, maximum and total number of observations used to compute the cross-sectional moments for a given age group. Note that even the smallest cell has a sample size of more than 75,000 on which the computation of higher-order moments is based.

[38] In reality, each individual is assigned a transformation of their SSN number for privacy reasons, but the same method applies.

Table A.1 – Sample Size Statistics for Cross-Sectional Moments of Five-Year Earnings Growth

	# Observations in Each RE Percentile Group			
Age group	Median	Min	Max	Total ('000s)
28-34	141,914	75,417	147,867	13,593
35-44	202,203	103,688	210,169	19,193
45-54	171,043	91,058	180,318	16,312

Note: Each row reports several statistics for the number of observations used to compute the cross-sectional moments of five-year earnings changes for a given age group. Since cross-sectional moments are computed for each age-year-RE percentile cell and then averaged over all years, sample sizes refer to the sum across all years of a given age by percentile group. The last column ("Total") reports the sum of observations across all 100 RE percentile groups for the age group indicated.

A.1 Imputation of Self Employment Income Above SSA Taxable Limit

We restrict our main sample for cross-sectional and impulse response moments to years between 1994 and 2013 during which neither self employment income nor wage/salary income is capped. However, this sample period–covering only 20 years–is too short to construct reliable measures of lifetime incomes of individuals. For this purpose, lifetime income moments in section 5 are computed using the whole sample that covers 36 years between 1978 and 2013. But self employment income is capped by the SSA maximum taxable earnings limit before 1994. In this section we introduce a methodology to impute self-employment income above the top code for years before 1994, and show that imputing self employment income has a negligible effect on our results.

Let y_t^{max} be the official SSA maximum taxable earnings limit in year t. Our goal is to impute the uncapped (unobservable) self employment income measure, $\tilde{y}_{i,t}^{SE}$ for individuals who have self employment income around the maximum taxable earnings limit reported in the MEF data specified by threshold χy_t^{max} (i.e., $y_{it}^{SE} \geq \chi y_t^{max}$), where $\chi < 1$.[39] For this purpose we take the uncapped self employment income measure in 1996, $y_{i,1996}^{SE}$, and regress it on observables that can also be constructed for the period before 1994.[40] In particular, we first group workers into three bins based on their age in year 1996: 28–29, 30–34, and 35–40.[41] Next, within each age group h, we estimate quantile regressions of uncapped self employment income in 1996 for 75 equally-spaced quantiles τ. Thus, in total we estimate the following specification 3×75

[39] We assume $\chi = 0.95 < 1$ because the MEF data have several observations above the SSA taxable limit implying measurement error around the limit.

[40] The first year with uncapped self employment income is 1994 but we rather use 1996 self employment income in the regression due to measurement issues in 1992 self employment income data.

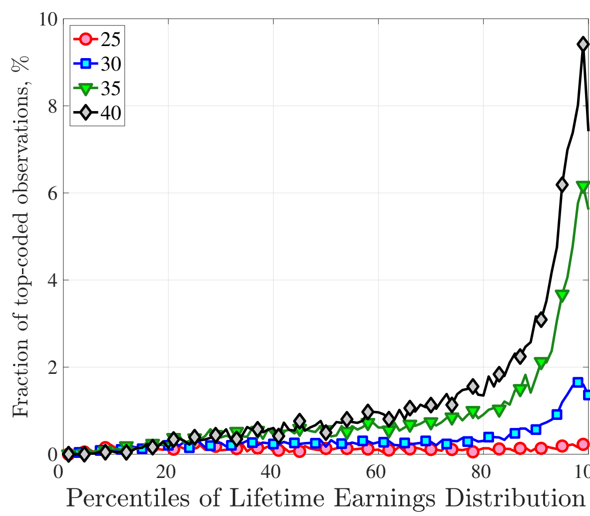
[41] Our imputed lifetime income sample employs a balanced panel that selects all individuals who are between ages 25 and 28 in 1981 (who were born between 1954 and 1957). This condition ensures that we have 33-year earnings histories between ages 25 and 60 for each individual (which might include years with zero earnings). The same condition also implies that, in this sample, only workers younger than 40 are affected by the top coding until 1993 and we impute their capped self employment income.

times—one for each age group and quantile:

$$\log y^{SE}_{i,1996} =$$
$$\sum_{k=0}^{3} \alpha^{h,\tau}_{1,k} \mathbb{I}\left\{y^{W}_{i,1996-k} < Y_{\min,1996-k}\right\} + \sum_{k=0}^{3} \alpha^{h,\tau}_{2,k} \mathbb{I}\left\{y^{W}_{i,1996-k} \geq Y_{\min,1996-k}\right\} \log y^{W}_{i,1996-k}$$
$$+ \sum_{k=1}^{3} \alpha^{h,\tau}_{3,k} \mathbb{I}\{y^{SE}_{i,1996-k} > \chi y^{max}_{1996-k}\} + \sum_{k=1}^{3} \alpha^{h,\tau}_{4,k} \mathbb{I}\left\{y^{SE}_{i,1996-k} < Y_{\min,1996-k}\right\}$$
$$+ \sum_{k=1}^{3} \alpha^{h,\tau}_{5,k} \mathbb{I}\left\{y^{SE}_{i,1996-k} \geq Y_{\min,1996-k}\right\} \min\left(\log y^{SE}_{i,1996-k}, \log \chi y^{max}_{1996-k}\right) + \varepsilon_{it}, \tag{9}$$

where $y^W_{i,t}$ is the wage and salary income of individual i in year t, $\mathbb{I}$ is the indicator function, and ε_{it} is the residual term. The right-hand side variables are as follows: (i) a dummy variable of whether the worker's wage earnings $y^W_{i,t}$ is less than the minimum income threshold $Y_{min,t}$; (ii) if it is higher then $Y_{min,t}$, the log of wage earnings $\log y^W_{i,t}$; (iii) a dummy variable of whether the self employment income $y^{SE}_{i,t}$ is above the maximum cap χy^{max}_t; (iv) a dummy variable of whether $y^{SE}_{i,t}$ is less than the minimum threshold $Y_{min,t}$; (v) if it is higher than $Y_{min,t}$, the log self employment income capped at the maximum threshold $\log\left(\min(y^{SE}_{i,t}, \chi y^{max}_t)\right)$. We also include 3 lags of these as independent variables. Then, $\alpha^{h,\tau}_{i,k}$ denotes the regression coefficient of variable i with lag k for age group h, quantile τ.

FIGURE A.1 – Fraction of Top-Coded Self Employment Income Observations



We then use these regression coefficients to impute the uncapped self employment income before 1994 for individuals who have SE income above the limit χy^{max}_t reported in the MEF data. For this purpose, we randomly assign individuals to quantiles $\tau = 1....75$ in our lifetime income sample. Then, the imputed self employment income for an individual in age group h with quantile τ who has recorded self employment income above the limit χy^{max}_t in year

$t = 1981, 1981, ...1993$ is given by the following equation:[42]

$$\begin{aligned} &\log \tilde{y}_{i,t}^{SE} = \\ &\textstyle\sum_{k=0}^{3} \alpha_{1,k}^{h,\tau} \mathbb{I}\left\{ y_{i,t-k}^{W} < Y_{\min,t-k} \right\} + \sum_{k=0}^{3} \alpha_{2,k}^{h,\tau} \mathbb{I}\left\{ y_{i,t-k}^{W} \geq Y_{\min,t-k} \right\} \log y_{i,t-k}^{W} \\ &+ \textstyle\sum_{k=1}^{3} \alpha_{3,k}^{h,\tau} \mathbb{I}\{ y_{i,t-k}^{SE} > \chi y_{t-k}^{max} \} + \sum_{k=1}^{3} \alpha_{4,k}^{h,\tau} \mathbb{I}\left\{ y_{i,t-k}^{SE} < Y_{\min,t-k} \right\} \\ &+ \textstyle\sum_{k=1}^{3} \alpha_{5,k}^{h,\tau} \mathbb{I}\left\{ y_{i,t-k}^{SE} \geq Y_{\min,t-k} \right\} \min\left(\log y_{i,t-k}^{SE}, \log \chi y_{t-k}^{max} \right). \end{aligned} \tag{10}$$

FIGURE A.2 – Income Growth for Imputed and Non-Imputed Data

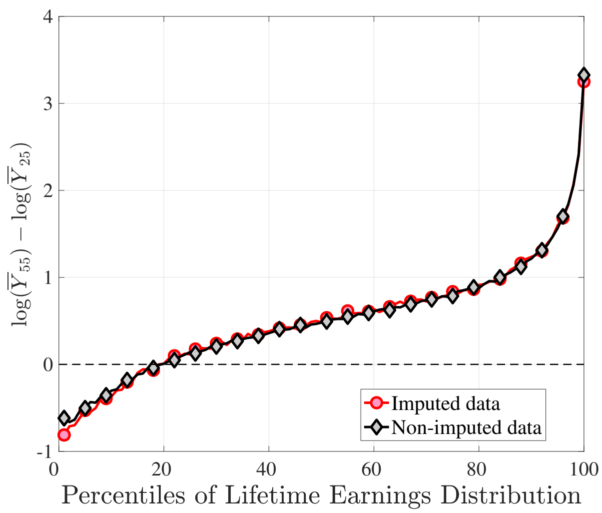


Figure A.2 plots the fraction of top coded self employment income observations against the percentiles of lifetime earnings distribution at ages 25, 30, 35, and 40 in our imputed lifetime income sample.[43] Almost no observations are top coded for individuals below the 20th percentile of the lifetime earnings distribution, in particular at young ages. As expected, the fraction of top-coded observations increases with age and with lifetime earnings and is highest for workers in the 99th percentile when they are 40 years old.

Furthermore, Figure A.2 plots the lifetime income growth between 25 and 55 against lifetime earnings percentiles using imputed and non-imputed data, which is already shown on Figure 11a. The two series are almost indistinguishable, indicating that top-coding has very little effect on lifetime income growth. This is because only a very small number of workers are affected by the top coding; those who had very high self-employment income before 1994 or when the cohort was younger than age 41.

B Transitory vs. Persistent Income Changes

Let's suppose that the earnings dynamics are given by the commonly used random-walk permanent/transitory model in which i.i.d. permanent (η_t^i) and transitory (ε_t^i) innovations are

[42] The imputed lifetime income sample starts with year 1981 because, to impute self employment income, we need to observe wage and self employment income in the previous three years between 1978 and 1980.

[43] Recall that in this sample only workers younger than 40 are affected by the top coding.

drawn from some general distributions F_η and F_ε, respectively. Then, the $k-$year log growth of earnings is given by:

$$\Delta^k_{\log} y^i_t = y^i_{t+k} - y^i_t = \sum_{j=t+1}^{t+k} \eta^i_j + \varepsilon^i_{t+k} - \varepsilon^i_t.$$

Let's denote the variance, skewness and excess kurtosis of distribution $F_x,\ x \in \{\eta, \varepsilon\}$ by σ^2_x, $\mathcal{S}_x$, and $\mathcal{K}_x$, respectively. Then the variance is given by:

$$\sigma^2(\Delta^k_{\log} y^i_t) = k\sigma^2_\eta + 2\sigma^2_\varepsilon.$$

In order to derive the skewness of $\Delta^k_{\log} y^i_t$ we use the following properties:

$$\begin{aligned}
\mathcal{S}(kx) &= \mathcal{S}_x,\ \text{for any } k > 0, \\
\mathcal{S}(x+y) &= \left(\frac{\sigma_x}{\sigma_{x+y}}\right)^3 \times \mathcal{S}_x + \left(\frac{\sigma_y}{\sigma_{x+y}}\right)^3 \times \mathcal{S}_y, \\
\mathcal{S}(x-y) &= \left(\frac{\sigma_x}{\sigma_{x+y}}\right)^3 \times \mathcal{S}_x - \left(\frac{\sigma_x}{\sigma_{x+y}}\right)^3 \times \mathcal{S}_y.
\end{aligned}$$

Then:

$$\begin{aligned}
\mathcal{S}(\Delta^k_{\log} y^i_t) &= \sum_{j=t+1}^{t+k} \left(\frac{\sigma_\eta}{\sigma^2(\Delta^k_{\log} y^i_t)}\right)^3 \times \mathcal{S}_\eta \\
&+ \left(\frac{\sigma_\varepsilon}{\sigma^2(\Delta^k_{\log} y^i_t)}\right)^3 \times \mathcal{S}_\varepsilon - \left(\frac{\sigma_\varepsilon}{\sigma^2(\Delta^k_{\log} y^i_t)}\right)^3 \times \mathcal{S}_\varepsilon \\
&= \frac{k\sigma^3_\eta S_\eta}{\sigma^3(\Delta^k_{\log} y^i_t)}
\end{aligned}$$

In order to derive the kurtosis of $\Delta^k_{\log} y^i_t$ we use the following properties:

$$\begin{aligned}
\mathcal{K}(kx) &= \mathcal{K}_x,\ \text{for any } k > 0, \\
\mathcal{K}(\sum_{j=1}^{k} x_j) &= \sum_{j=1}^{k} \left[\left(\frac{\sigma_{x_j}}{\sigma(\sum_j x_j)}\right)^4 \cdot \mathcal{K}_{x_j}\right].
\end{aligned}$$

We obtain:

$$\mathcal{K}(\Delta^k_{\log} y^i_t) = \frac{k \times \sigma^4_\eta}{\sigma^4(\Delta^k_{\log} y^i_t)}\mathcal{K}_\eta + \frac{2 \times \sigma^4_\varepsilon}{\sigma^4(\Delta^k_{\log} y^i_t)}\mathcal{K}_\varepsilon.$$

C Appendix: Robustness and Additional Figures

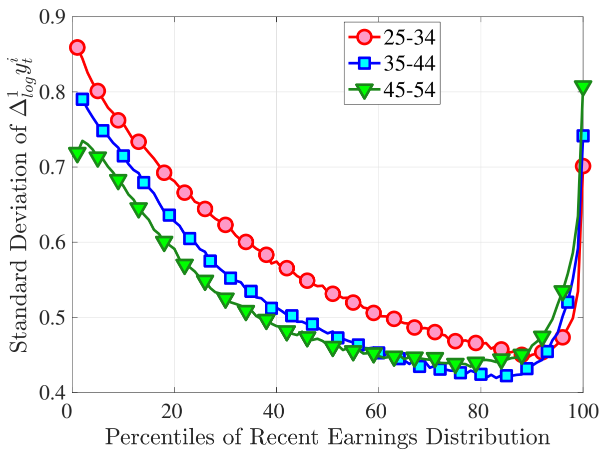
This section reports additional results from the data. Section C.1 reports the cross-sectional moments of one-year earnings growth. Section C.2 shows the cross-sectional moments of one-year and five-year arc-percent changes of earnings. Finally, Section C.3 shows several features of the data that are mentioned in the paper but are relegated to the appendix.

C.1 Cross-Sectional Moments of One-Year Earnings Growth

Throughout the main text, we showed the cross-sectional moments of five-year (log) earnings growth. This section shows analogous features of the data for one-year earnings growth.

Figure C.1 – Dispersion of One-Year Log Earnings Growth

(a) Second Standardized Moment



(b) P90-10

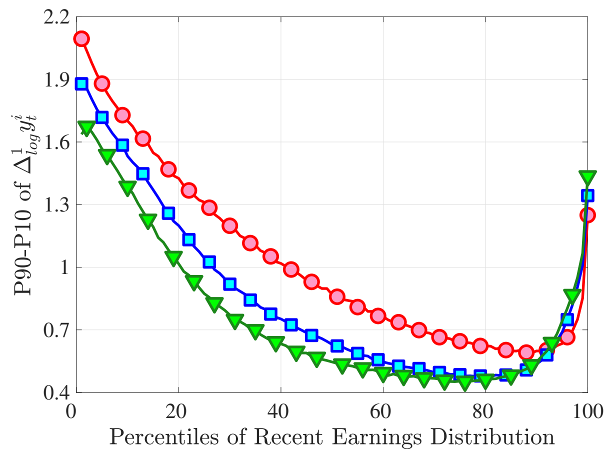
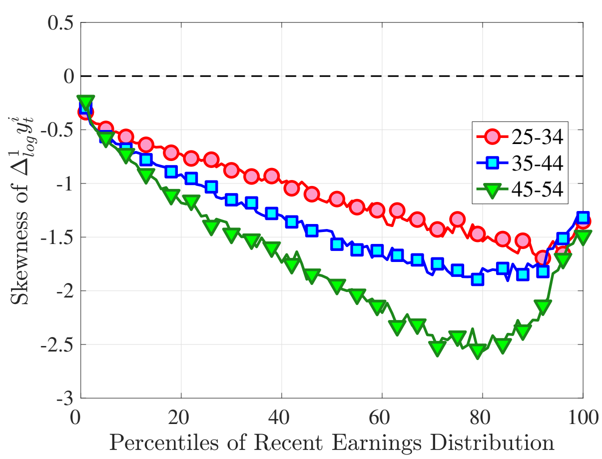


Figure C.2 – Skewness of One-Year Log Earnings Growth

(a) Third Standardized Moment



(b) Kelley Measure

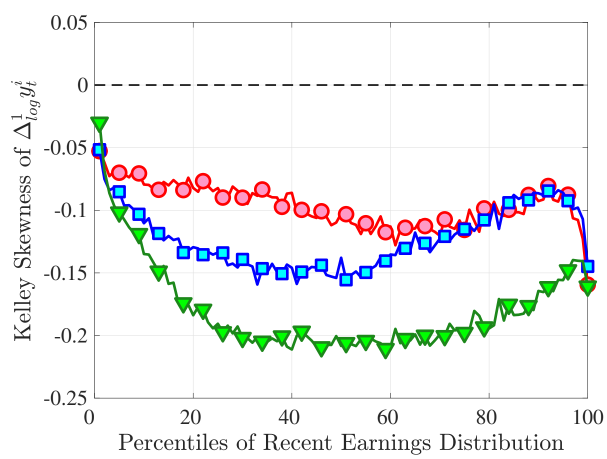


FIGURE C.3 – Kelley's Skewness Decomposed: Change in P90-P50 and P50-P10 Relative to Age 25–29 (Log Growth)

(A) P90-P50 (B) P50-P10

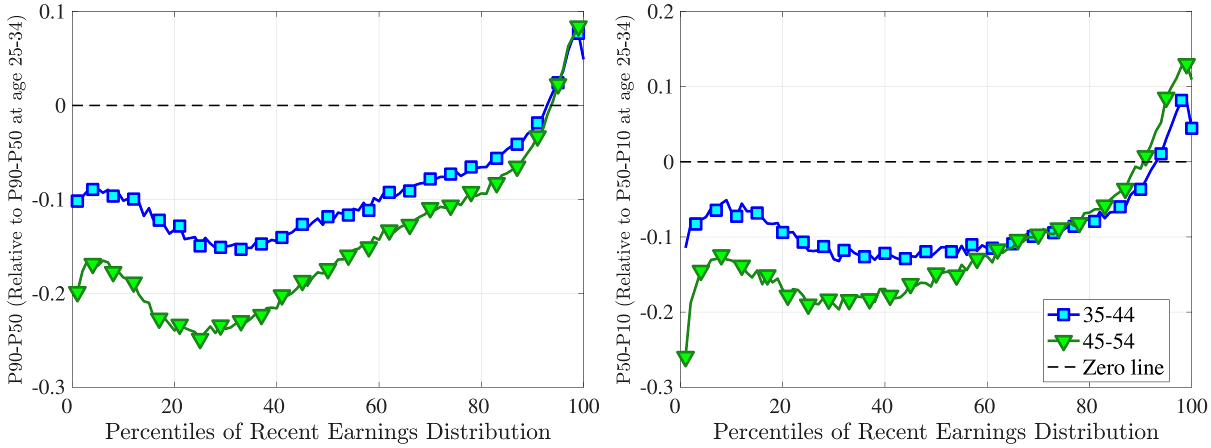
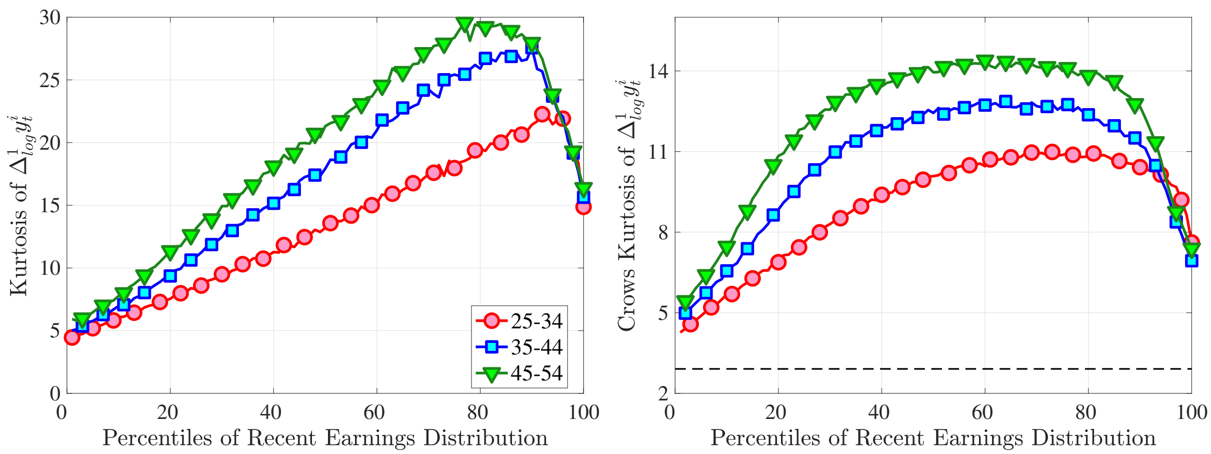


FIGURE C.4 – Kurtosis of One-Year Log Earnings Growth

(A) Fourth Standardized Moment (B) Crow-Siddiqui Measure



C.2 Arc-Percent Moments

In the main text, we documented moments of log earnings changes. In doing so, we are forced to drop observations close to zero to obtain sensible statistics. However, as we discuss in Section 2, such observations contain potentially valuable information, as they inform us about very large changes in earnings caused by events such as long-term nonemployment. To complement our analysis, this section reports the cross-sectional moments of arc-percent changes defined in Section 2, which we reproduce here for convenience:

$$\text{arc percent change:} \qquad \Delta_{\text{arc}} Y^i_{t,k} = \frac{Y^i_{t+k} - Y^i_t}{\left(Y^i_{t+k} + Y^i_t\right)/2}.$$

This measure allows computation of earnings growth even when the individual has zero income in one of the two years t and $t+k$. Section C.2.1 shows the moments of one-year arc percent change, and C.2.2 shows the moments of five-year change.

C.2.1 Moments of Annual Arc-Percent Changes

FIGURE C.5 – Dispersion of Annual Arc-percent Earnings Change

(A) Second Standardized Moment (B) P90-10

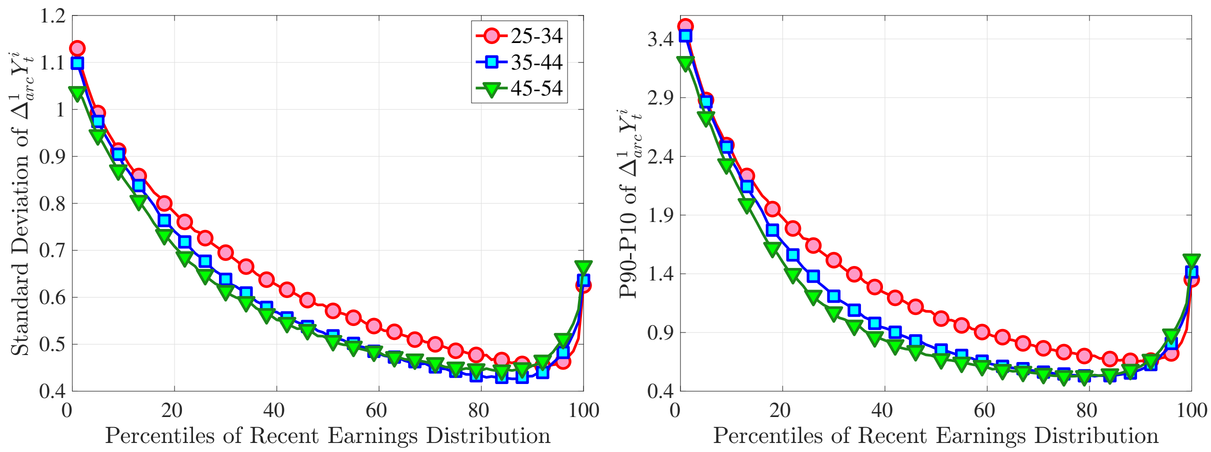


FIGURE C.6 – Skewness of Annual Arc-percent Earnings Change

(A) Third Standardized Moment (B) Kelley Measure

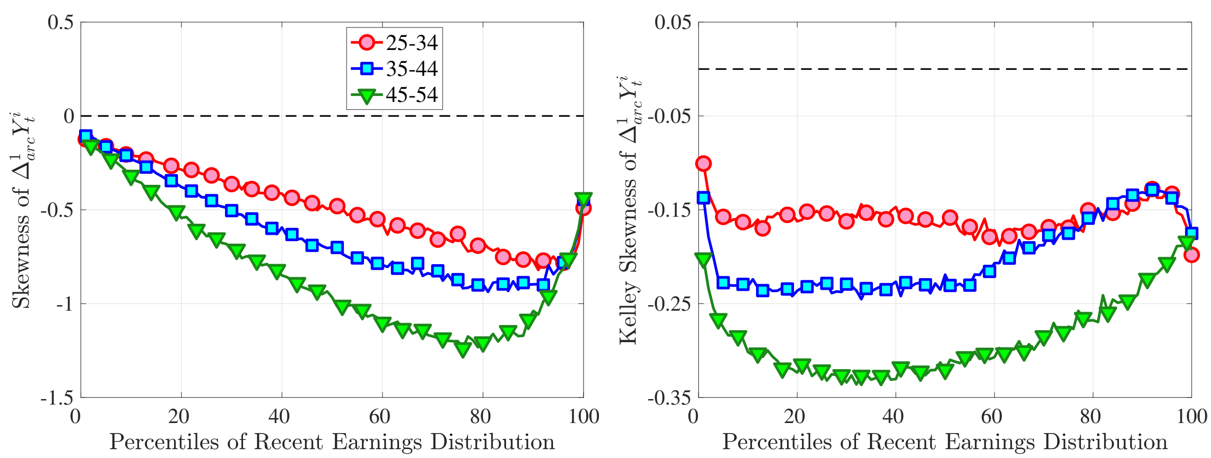


Figure C.7 – Kelley's Skewness Decomposed (Annual Arc-percent Growth): Change in P90-P50 and P50-P10 Relative to Age 25–34

(a) P90-P50 (b) P50-P10

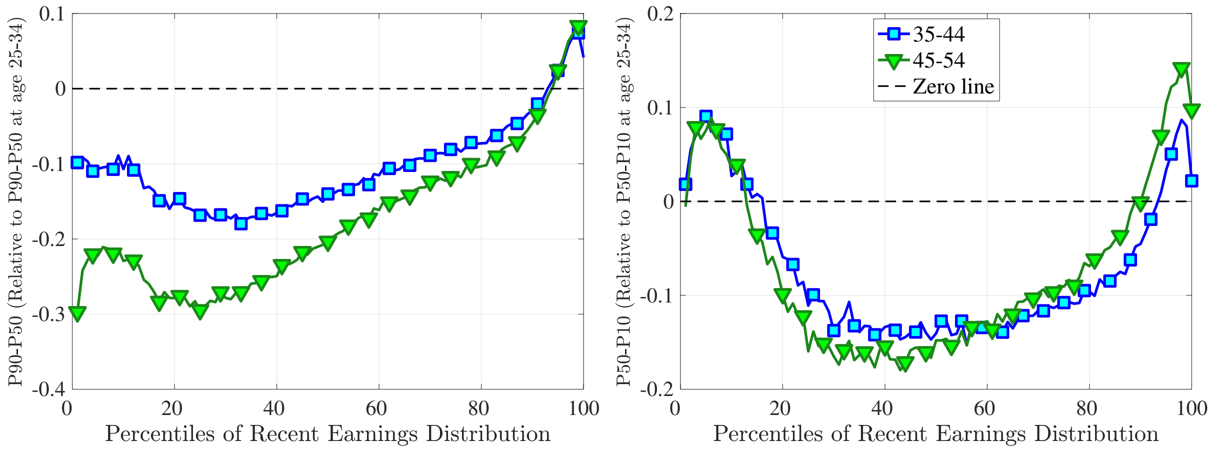
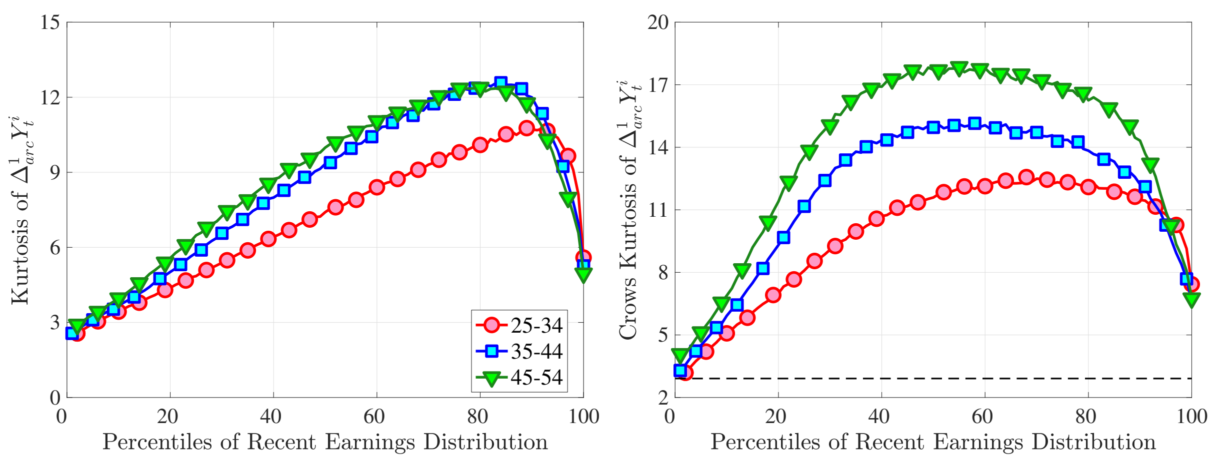


Figure C.8 – Kurtosis of Annual Arc-percent Earnings Change

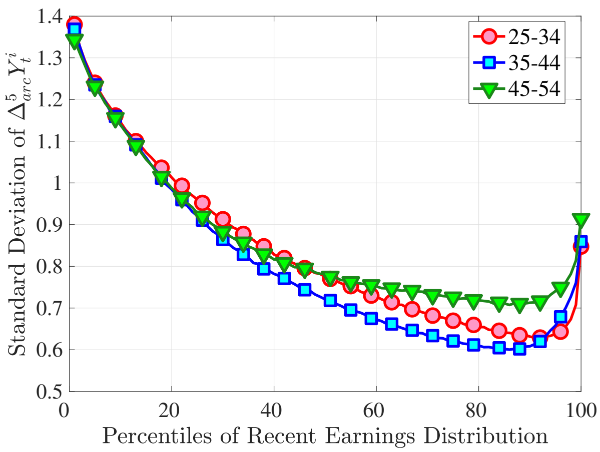
(a) Fourth Standardized Moment (b) Crow-Siddiqui Measure



C.2.2 Moments of Five-Year Arc-Percent Changes

Figure C.9 – Dispersion of Five-Year Arc-percent Earnings Change

(a) Second Standardized Moment



(b) P90-10

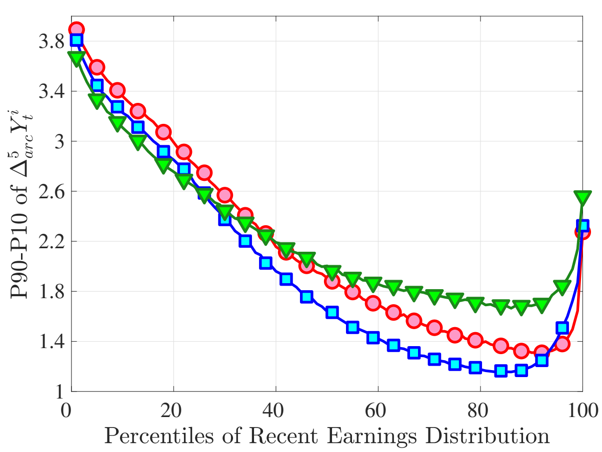
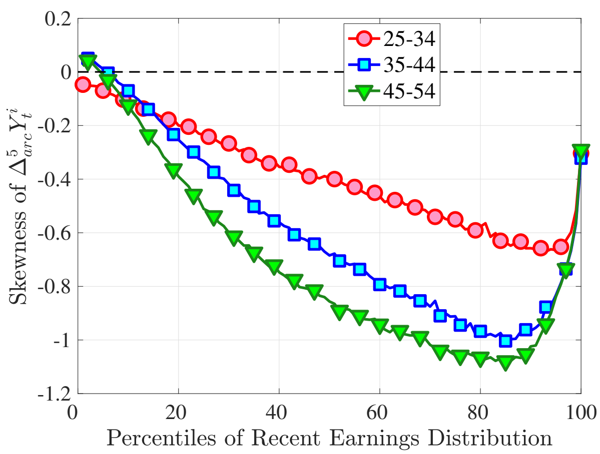


Figure C.10 – Skewness of Five-Year Arc-percent Earnings Change

(a) Third Standardized Moment



(b) Kelley Measure

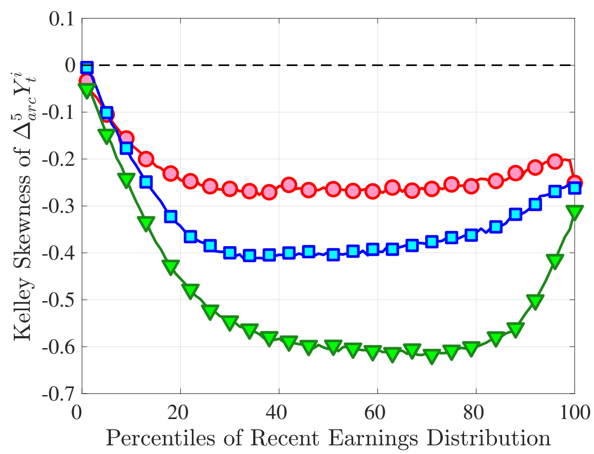


FIGURE C.11 – Kelley's Skewness Decomposed (5-Year Arc-percent Growth): Change in P90-P50 and P50-P10 Relative to Age 25–34

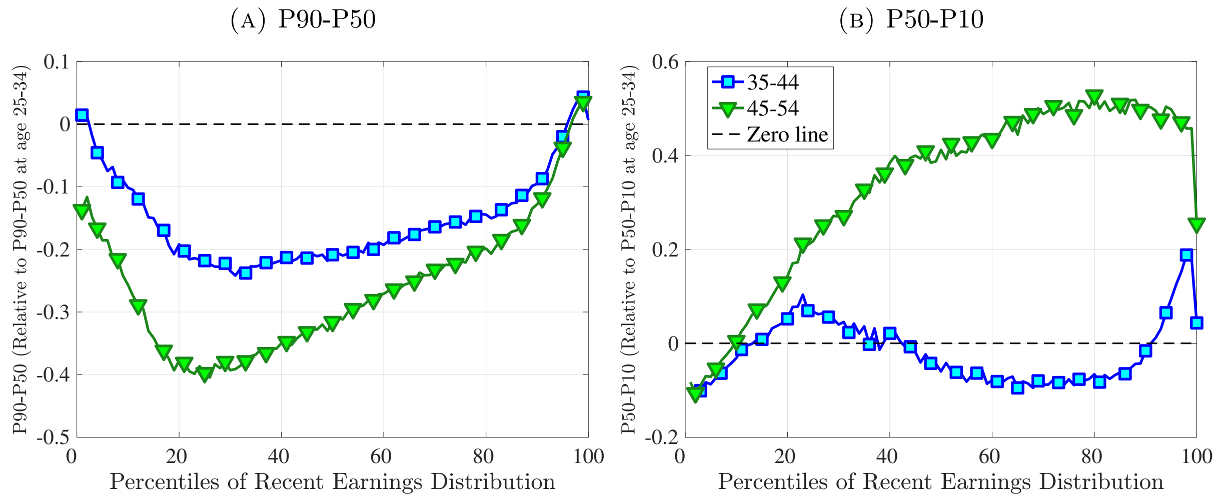
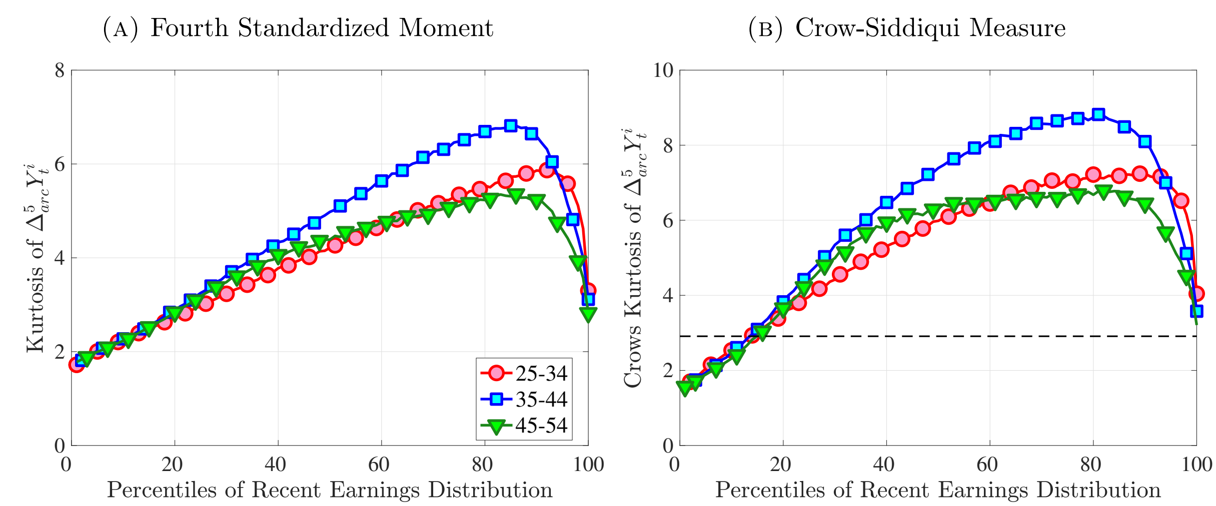


FIGURE C.12 – Kurtosis of Five-Year Arc-percent Earnings Change

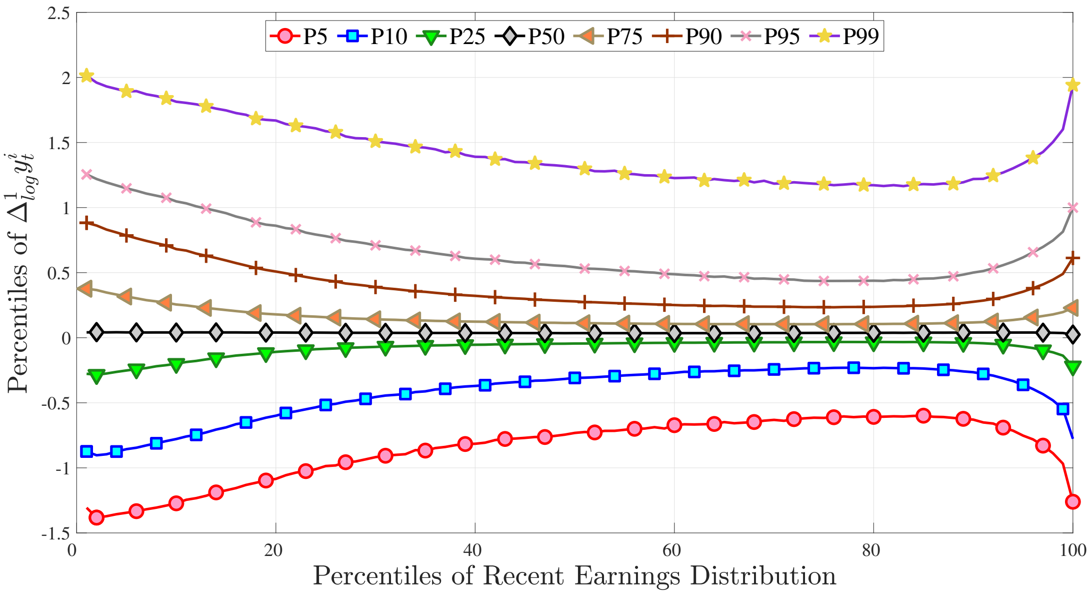


C.3 Further Figures

In this section, we report some additional figures of interest that are omitted from the main text due to space constraints. First, Figure C.13 plots selected percentiles of the annual and five-year log earnings change distribution for every RE percentile.

Figure C.13 – Selected Percentiles of Log Earnings Changes

(A) One-Year Changes



(B) Five-Year Changes

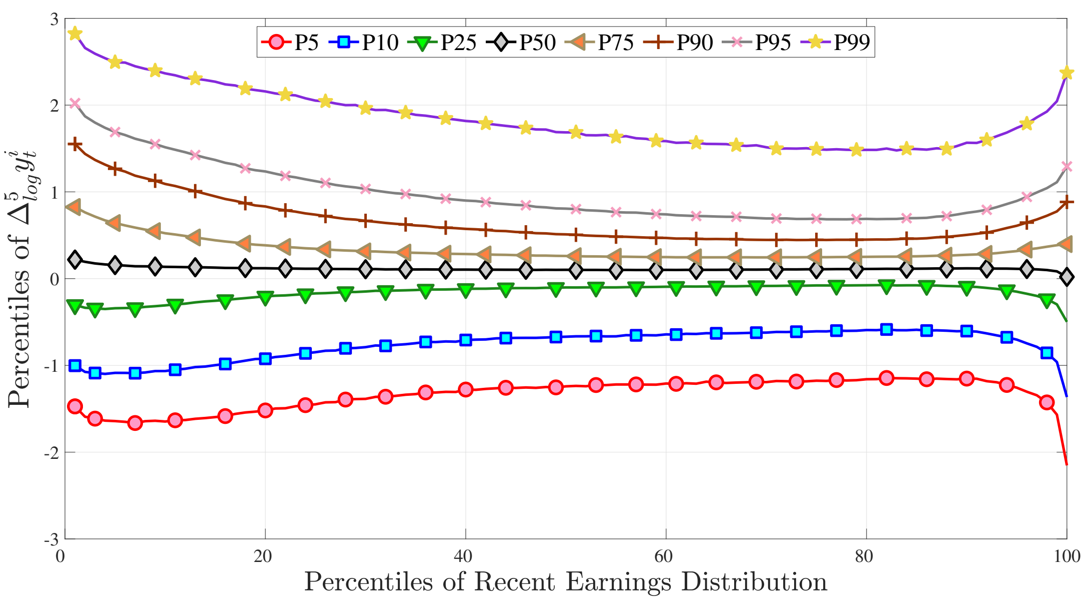
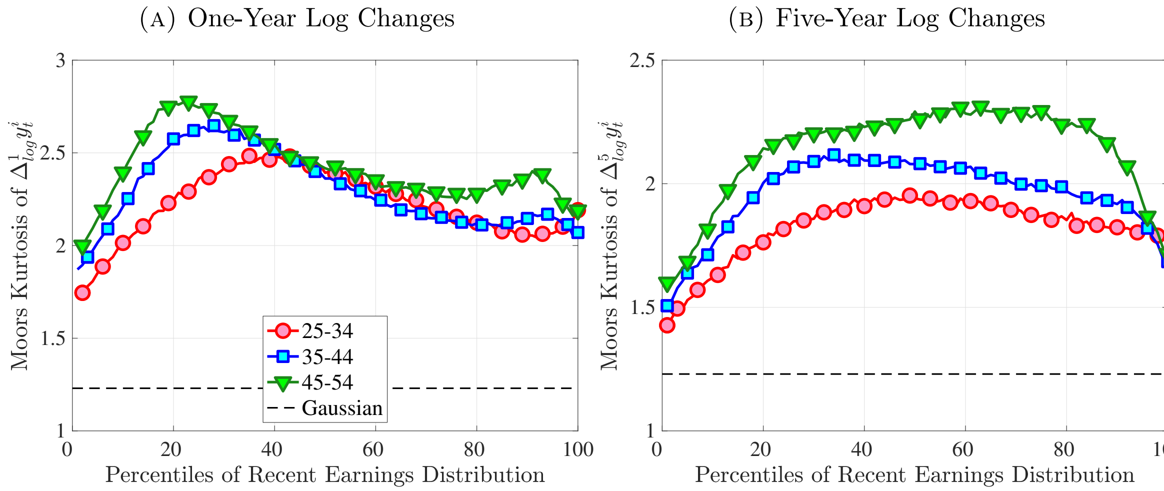


FIGURE C.14 – Moors Kurtosis of Log Earnings Changes



Second, Figure C.14 shows an additional measure of kurtosis proposed by Moors (1988) for one- and five-year earnings changes. Similar to the measure proposed by Crow and Siddiqui (1967), this measure is robust to outliers in the tails. Moors' kurtosis, κ_M is defined as

$$\kappa_M = \frac{(P87.5 - P62.5) + (P37.5 - P12.5)}{P75 - P25}.$$

For a Gaussian distribution, Moors' kurtosis is 1.23 (shown on dashed lines).

C.4 Cross-Sectional Moments for Job-Stayers vs. Job-Switchers

In the main text, we analyzed the properties of earnings growth separately for job stayers and switchers by showing percentile-based moments of earnings growth. Here, we complement our analysis by showing several features of the data that were omitted in the main text to save space. First, Figure C.15 shows our measure of the fraction of job stayers as a function of recent earnings and age. The probability of staying with the same employer increases with recent earnings and age. For the youngest age group (25-34), the probability of staying in the same job is around 20% at the bottom of the recent earnings distribution. This fraction increases with recent earnings and reaches a peak around 60% at the 95th percentile of the RE distribution. This pattern reverses itself slightly at the top of the RE distribution. As workers age, the probability of staying with the same employer increases across the RE distribution.

Second, Figure C.16 shows the age profile of higher-order moments shown in Section 3.5. The age patterns are broadly the same across switchers and stayers: P90-10 declines slightly, skewness becomes more negative, and kurtosis increases for both job-stayers and job-switchers over the life cycle.

Next, we complement our analysis of job stayers and switchers by investigating the centralized moments of one- and five-year earnings changes (as opposed to percentile-based moments analyzed in the main text). We plot these moments in Figure C.18.

Figure C.15 – Fraction Staying Jobs Between t and $t+1$

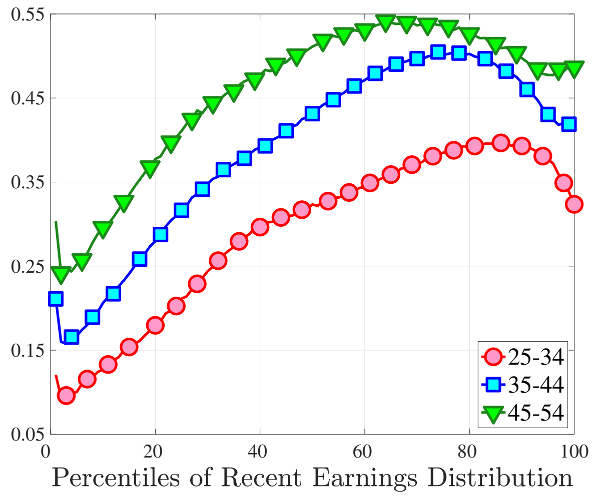
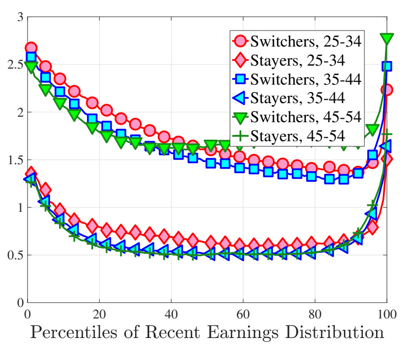
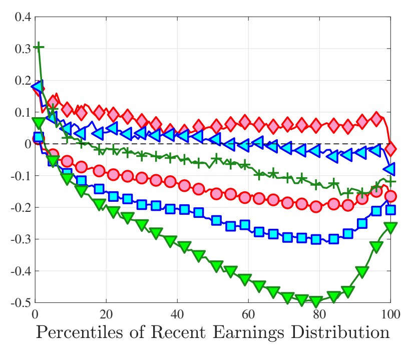


Figure C.16 – Percentile-Based Moments of Five-Year Earnings Growth: Stayers vs Switchers

(a) P90-P10



(b) Kelley's Skewness



(c) Crow-Siddiqui Kurtosis

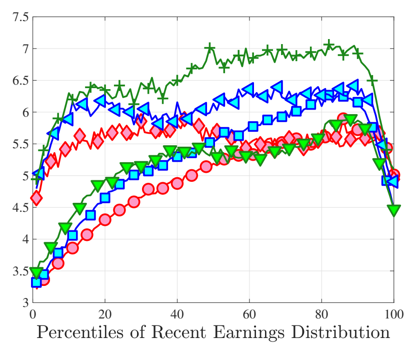
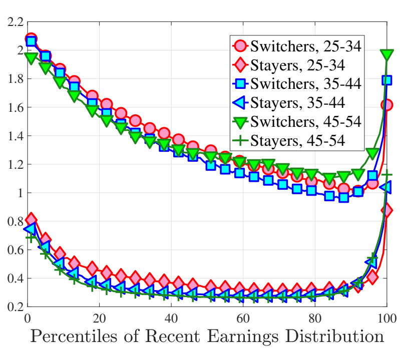
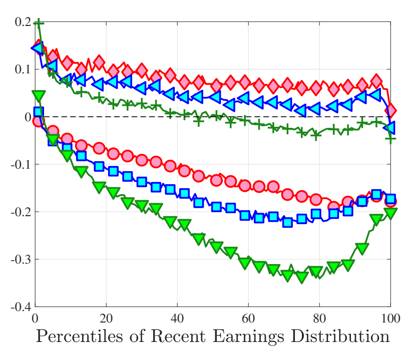


Figure C.17 – Percentile-Based Moments of One-Year Earnings Growth: Stayers vs Switchers

(a) P90-P10



(b) Kelley's Skewness



(c) Crow-Siddiqui Kurtosis

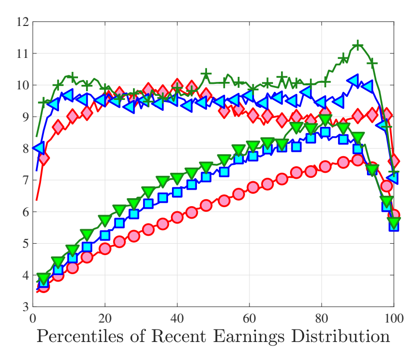
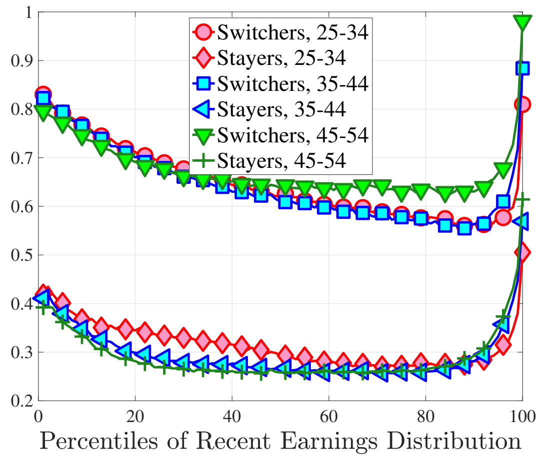
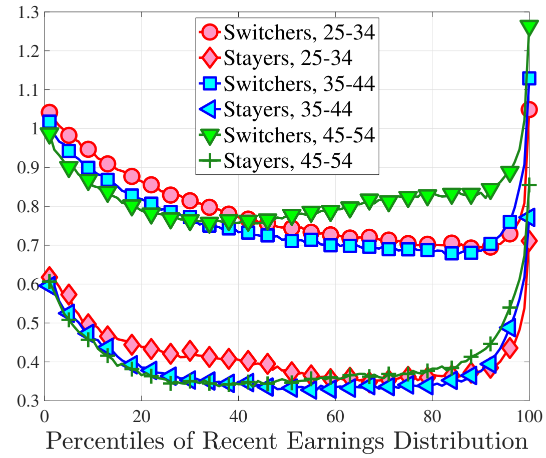


Figure C.18 – Centralized Moments of Earnings Growth: Stayers vs Switchers

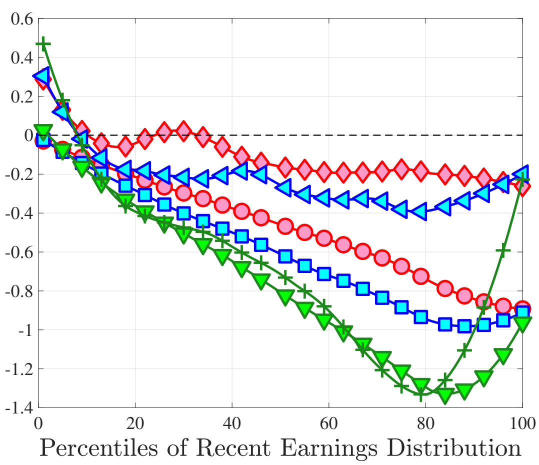
(a) Std. Dev., One-year



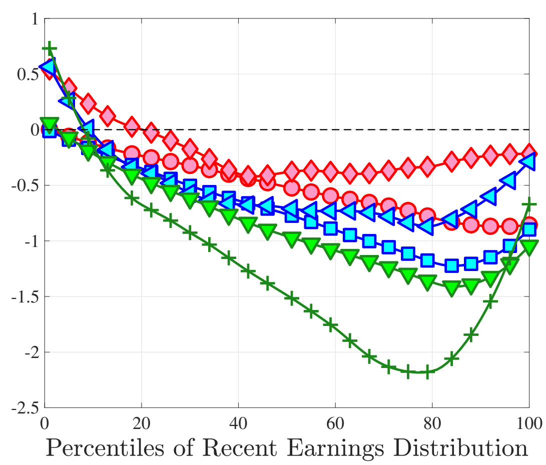
(b) Std. Dev., Five-Year



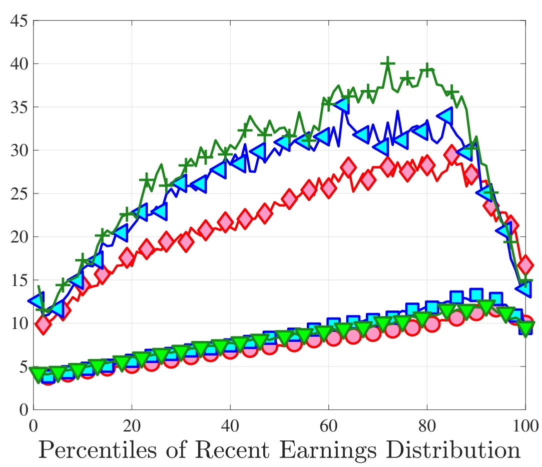
(c) Skewness, One-Year



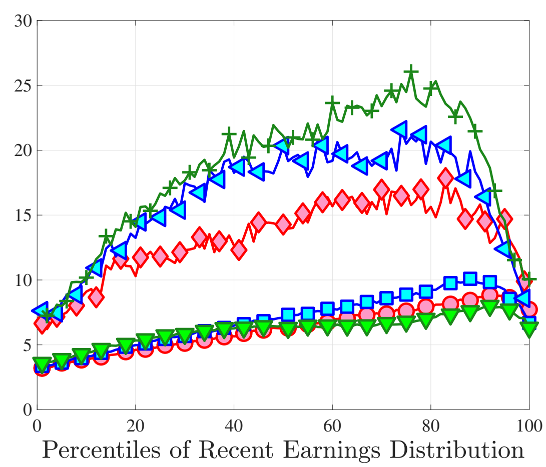
(d) Skewness, Five-Year



(e) Kurtosis, One-Year



(f) Kurtosis, Five-Year



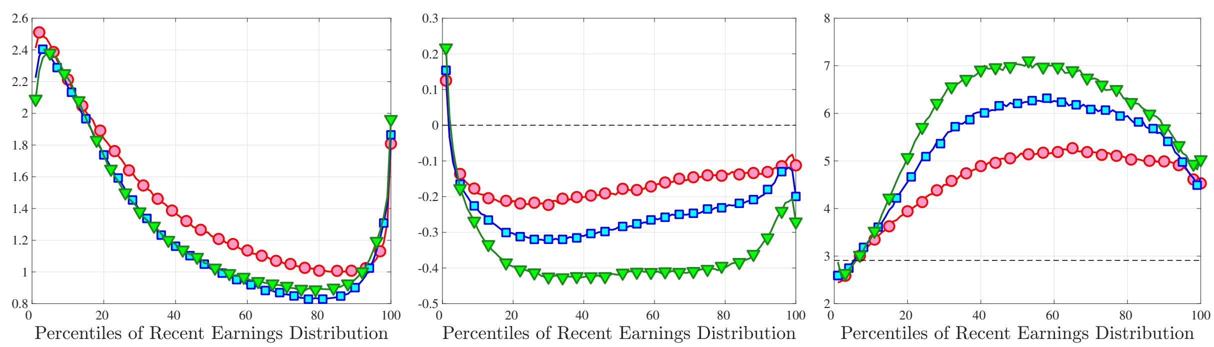
The results are consistent with what one might expect. Job-stayers face earnings changes that (i) have half the dispersion of job-changers, (ii) are less negatively skewed as opposed to job-switchers, who face very negatively skewed changes, and (iii) have a much higher kurtosis than job-switchers. In fact, kurtosis is as high as 40 for annual changes and 25 for five-year changes for job-stayers, but is less than 10 for job-switchers at both horizons.

C.5 An Alternative Measure for Persistent Earnings Changes

In section 3, we studied the distribution of five-year earnings changes, and explained that the five-year changes reflect more of the distribution of the persistent innovations rather than transitory innovations. We also considered an alternative measure ($\overline{\Delta}^5_{\log}(\bar{y}_t{}^i) \equiv \bar{y}^i_{t+4} - \bar{y}^i_{t-1}$, where $\bar{y}^i_{t+4} \equiv \log(\bar{Y}^i_{t+4})$ and $\bar{y}^i_{t-1} \equiv \log(\bar{Y}^i_{t-1})$) to deal with the caveat that our baseline measure is contaminated by transitory changes in years t and $t+k$. The main text showed the centralized moments of this alternative measure; we now show the percentile-based moments.

FIGURE C.19 – Percentile-based Moments of Persistent Earnings Changes, $\overline{\Delta}^5_{\log}(\bar{y}_t{}^i)$

(A) P90-P10 (B) Kelley Measure (C) Crow-Siddiqui Measure



C.6 Additional Figures on Lifecycle Patterns of Earnings

To provide a benchmark for the analysis in Section 5, we estimate the average lifecycle profile of log earnings using a standard pooled regression of log individual earnings on a full set of age and (year-of-birth) cohort dummies using the admissible observations (as defined in Section 2) between 1994-2013.[44] The estimated age dummies are plotted as circles in Figure C.20 and represent the average life-cycle profile of log earnings. It has the usual hump-shaped pattern that peaks around age 50. These age dummies turn out to be indistinguishable from a fourth-order polynomial in age:

$$y_h = -0.0240 + 0.2013 \times h - 0.6799 \times h^2 + 1.2222 \times h^3 + 9.4895 \times h^4,$$

where $h = (\text{age} - 24)/10$. Figure C.21 contains two panels on the distribution of lifecycle growth rates that complement the analysis in Section 5.

[44] This procedure is standard in the literature; see, e.g., Deaton and Paxson (1994) and Storesletten *et al.* (2004).

FIGURE C.20 – Life-Cycle Profile of Average Log Earnings

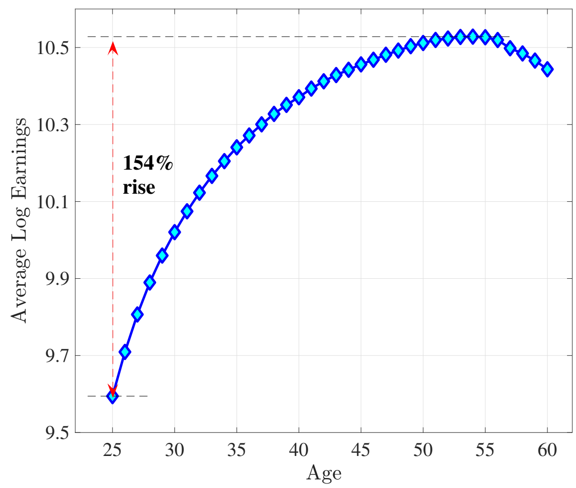
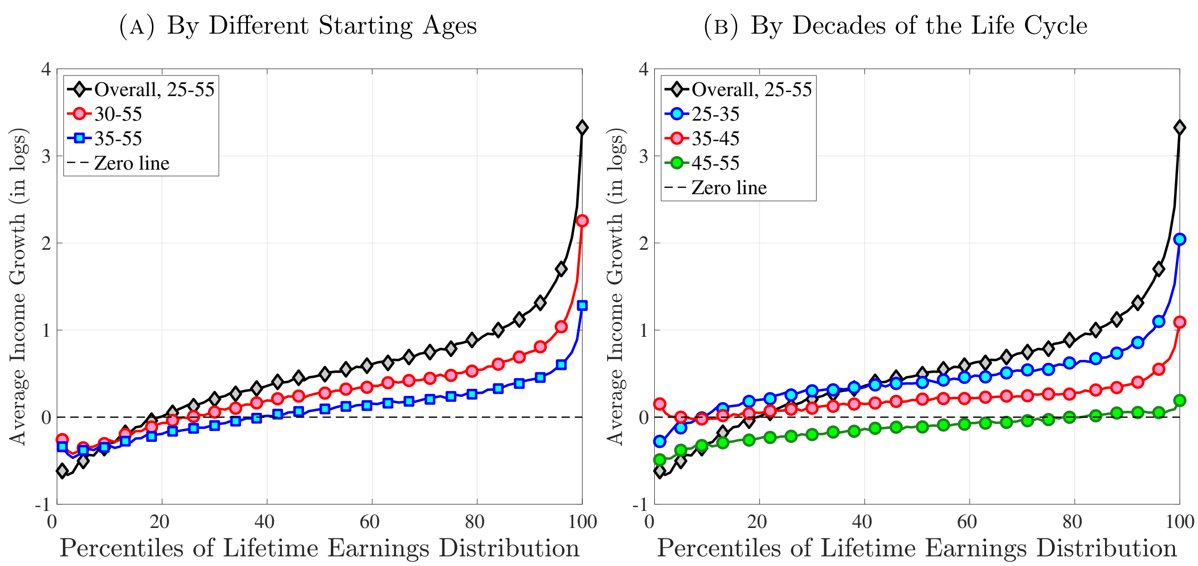


FIGURE C.21 – Log Earnings Growth over Subperiods of Life Cycle



C.7 Survey Data and Higher-Order Moments

Panel Study of Income Dynamics (PSID)

Panel Study of Income Dynamics has a smaller sample size compared to the CPS, but has the advantage of following the same household over a much longer period of time. The PSID started collecting data annually in 1968 on a sample of around 5,000 households. 3,000 households were representative and the remaining were low-income families (the Census Bureau's Survey of Economic Opportunities sample, SEO). We restrict our study to households in the core sample and do not use households in the SEO or the Latino subsamples. The questions on income are retrospective, meaning that respondents in a given year are asked about the previous calendar year. We analyze data for the period 1999–2013. During this period, the survey was biennial.

Our measure of labor income (variable ER16463 in 1999) is the sum of wages and salaries, bonuses, pay for overtime, tips, commissions, professional practice or trade, market gardening, miscellaneous labor income and extra job income. To remain consistent with the rest of the paper, we focus on male head of households between ages 25 and 55. We deflate annual earnings by the price level with the base year 2010.[45] We drop observations that report earnings less than $1,500. We residualize earnings, wages and hours by regressing them on a full set of age dummies, controlling for 3 race, 3 education and 8 region dummies. The three education levels are: less than 11 years (less than high school), 12 years (high school), and more (college dropout, BA degree or more). We take the maximum grade achieved as the relevant education level of an individual throughout the sample. Race dummies correspond to white, black and the remaining race and ethnicities. We clean the age variable so that it increases by two for each individual across two surveys. We run these regressions separately by year, obtain the residuals and analyze the change in the residuals between consecutive interviews. We group observations into 7 bins, depending on the magnitude of this change.

We utilize different variables in the PSID to identify individuals that experience a change in health, start experiencing disability, experience time out of work or a job or occupation change. We now describe specifically which variables are used to construct the various measures in Table III.

Current Population Survey (CPS)

The CPS is a rotating panel based on addresses. Each address in the survey is interviewed for four consecutive months, then leaves the sample for eight months and then returns for another four consecutive months. Because of this rotating nature, it is possible to match at most three quarters of respondents across months. Since the survey is based on addresses and doesn't follow households, if households move, they leave the sample and may be replaced by others that move in to the same address. To have a reliable panel, we match individual records using rotation groups, household identifiers, individual line numbers, race, sex, and age.

The Annual Social and Economic Supplement (ASEC) of the CPS, a supplement to the CPS in March, asks respondents about their earnings, hours and weeks worked during the past calendar year (variables incwage, wkslyr and hrslyr, respectively.) We use data for the period 1968–2013 and focus on males between ages 25 and 55. Similar to the SSA sample, we impose a minimum earnings threshold that corresponds to working for 13 weeks for 40 hours a week at half the minimum wage. We focus on three measures: annual earnings, average weekly wages, and average hourly wages. We regress each measure on a full set of age dummies and a race dummy (white and nonwhite). We run these regressions separately for each year and education group (college and non-college), thereby allowing the coefficients on age and race dummies to depend on age and education. We then obtain the residuals from this regression and analyze the changes in the residuals. We use the CPS weights throughout this analysis.

Bad health. The head is asked the following question (ER15447 in 1999): "Would you (HEAD) say your health in general is excellent, very good, good, fair, or poor?" We classify

[45]We use the Consumer Price Index for All Urban Consumers (CPI-U) published by the Bureau of Labor Statistics.

someone in bad health if he reports a poor health condition (==5). Transitions into poor health are identified as someone who reported being in excellent, very good, good, or fair health in the previous survey and reports being in poor health in the current survey. This variable is available throughout our sample period.

Disability. The head is asked the following question on disability (ER15451 in 1999): "For work you can do, how much does it limit the amount of work you can do--a lot, somewhat, just a little, or not at all?" We classify someone as having some disability if he reports having an issue that affects his work a lot (==1), somewhat (==3) or just a little (==5). A new disability is coded as someone who did not have such an issue the last time and reports an issue in the current survey. This variable is available throughout our sample period.

Weeks unemployed. Some individuals report time spent unemployed in units of months (ER21322 in 2003), whereas some report it in units of weeks (ER21320 in 2003). We combine these two variables by taking the maximum reported unemployment duration in weeks. These variables are available starting in 2003.

Weeks out of labor force. The PSID asks about head's total weeks out of the labor force in the previous calendar year (ER24087 in 2003). This variable is available throughout our sample period.

Move in response to outside events (involuntary reasons) The PSID asks about geographic moves (whether the head changed residence) and the reasons of the move (variable ER13080 in 1999). We classify someone as having moved for involuntary reasons if they report having moved for being evicted, armed services, health reasons, divorce and health-related retirement. Other observations are classified as non movers. This variable is available throughout our sample period.

Group $\Delta y \in$	$(-\infty,-1)$	$[-1,-0.25)$	$[-0.25,0)$	$[0,0.25)$	$[0.25,1)$	$(1,\infty)$
Share %	3.8%	14.4%	31.2%	31.1%	16.5%	3.0%
Invol. move %	6.9%	4.5%	3.2%	2.6%	3.6%	4.3%

Occupation change. The PSID asks about the head's occupation in the main employer, labeled as job 1 (ER21145 in 2003). This variable is available starting in 2013 and uses the 3-digit occupation code from 2000 Census of Population and Housing. This variable is available every year since 2003. We code someones as having changed occupations if i) his occupation in the current year t is different than the previous survey $t-2$, ii) he reports having changed jobs (explained below), and iii) his occupation in the next survey $t+2$ is different from his occupation in year $t-2$. The last condition is used to deal with potential coding error of occupations prevalent in most survey data.

Job change. We use the start year of the current main job (job 1) to identify job changes (ER21130 in 2003). If the head reports having started the job in the same year as the survey or the year before, we code him as a job switcher. This variable is available every year since 2003.

Higher-Order Moments in CPS. In the main text, we reported higher-order moments of two-year changes in earnings and wages by age groups. Table C.1 provides similar results from the CPS. The growth measure in the CPS is annual and are therefore not easily comparable to the figures from the PSID. However, the findings are qualitatively similar

Table C.1 – Higher-Order Moments of Income Changes in PSID and CPS

		PSID					
		All		25–39		40-55	
	Gaussian	Earnings	Wages	Earnings	Wages	Earnings	Wages
Skewness	0.0	–0.26	–0.14	–0.17	–0.20	–0.34	–0.09
Kelley Skewness	0.0	–0.02	–0.02	0.03	0.016	–0.06	–0.04
Kurtosis	3.0	12.26	13.65	10.44	9.00	14.01	17.10
Crow Kurtosis	2.91	6.83	5.59	6.33	5.02	7.33	6.11
		CPS					
		All		25–39		40-55	
		Earnings	Wages	Earnings	Wages	Earnings	Wages
Skewness	0.0	–0.15	–0.09	–0.09	–0.023	–0.21	0.004
Kelley Skewness	0.0	–0.02	–0.01	0.002	–0.008	–0.03	–0.017
Kurtosis	3.0	9.29	11.2	9.12	10.60	9.53	12.1
Crow Kurtosis	2.91	7.15	5.93	6.97	5.72	7.29	6.05

Note: Wages are obtained by dividing annual earnings by annual hours in the PSID, and by the weekly wage variable in the CPS.

C.8 Cross-sectional moments by age without sample selection

When choosing the sample for cross-sectional moments, we required an individual to have an earnings level above the minimum threshold in $t-1$ and in at least two more years between $t-5$ and $t-2$. Figure C.22 shows that these conditions result in a substantial share of the initial sample to be dropped from the analysis. This large selectivity opens up the possibility that some of our results might be specific to our final sample. Here, we relax the selection criteria and include any person for whom earnings change can be computed. Figures C.23–C.28 show the centralized and percentile-based moments of one-year and five-year earnings changes. We find that our substantive conclusions are unchanged: Dispersion of earnings changes declines with age for most of the life cycle, earnings changes become more negatively skewed and more leptokurtic.

Figure C.22 – Fraction of Observations Dropped by Age

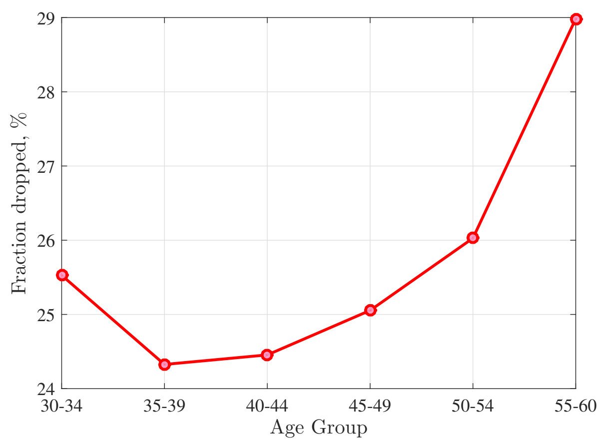
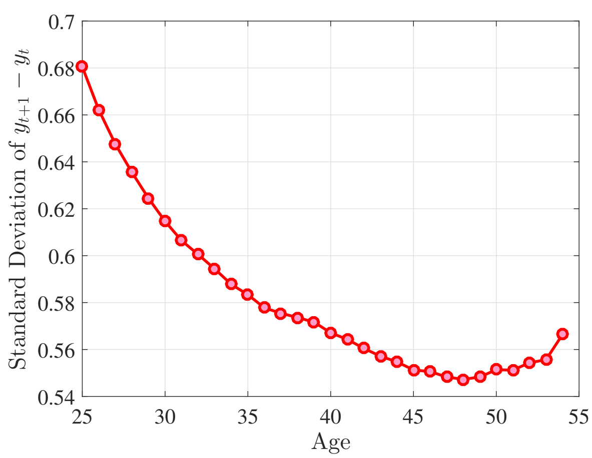


Figure C.23 – Dispersion of One-Year Log Earnings Growth by Age

(a) Second Standardized Moment



(b) P90-P10

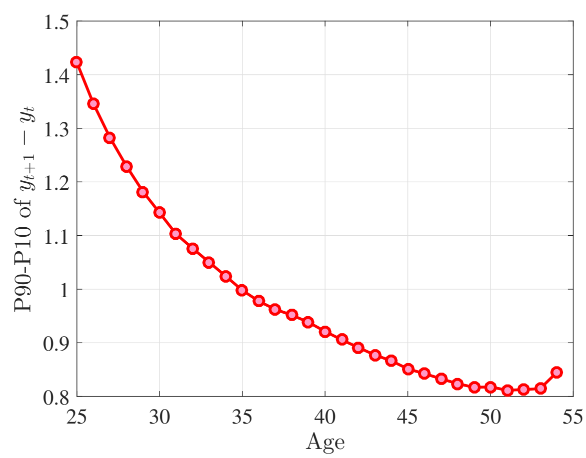


Figure C.24 – Skewness of One-Year Log Earnings Growth by Age

(a) Third Standardized Moment (b) Kelley Measure

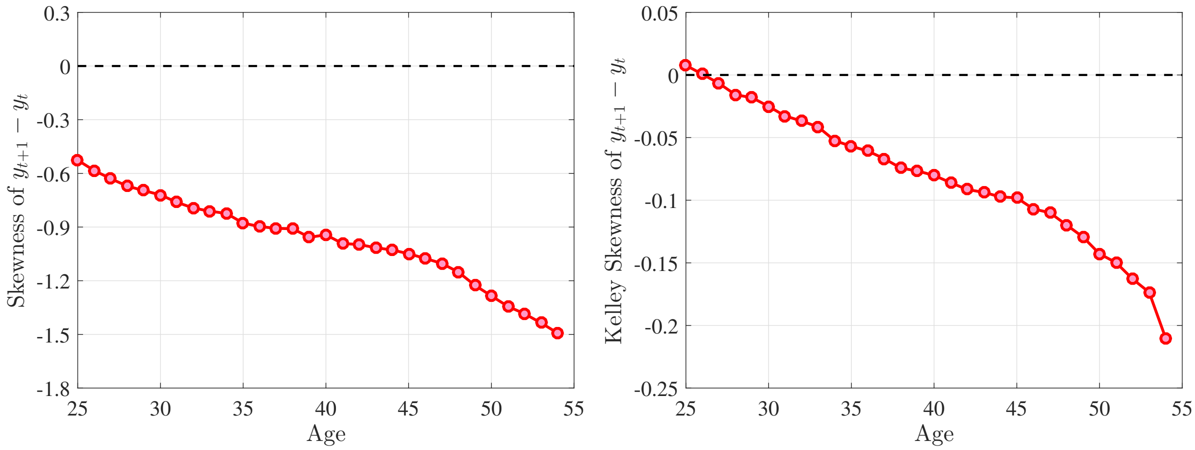


Figure C.25 – Kurtosis of One-Year Log Earnings Growth by Age

(a) Fourth Standardized (b) Crow-Siddiqui Measure

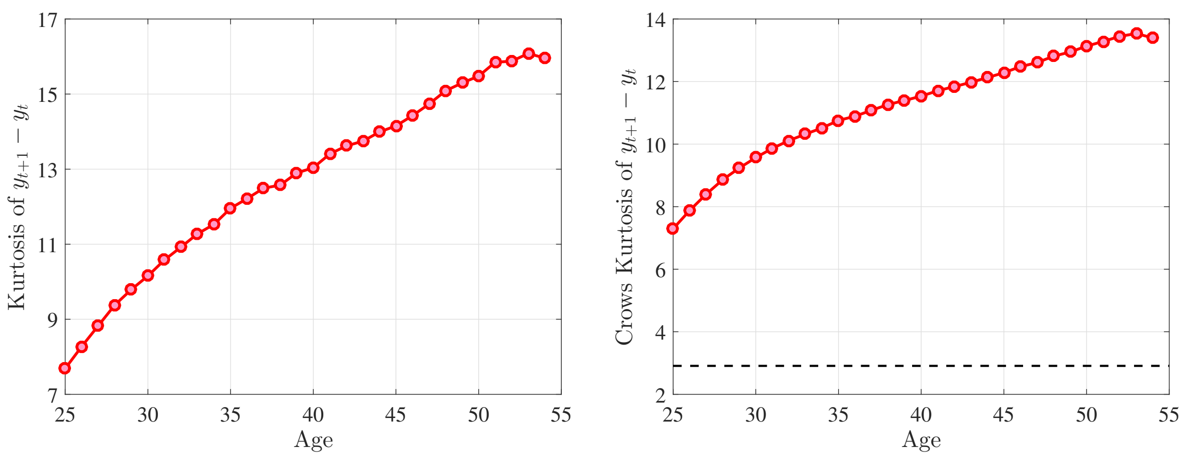


Figure C.26 – Dispersion of Five-Year Log Earnings Growth by Age

(a) Second Standardized Moment

(b) P90-P10

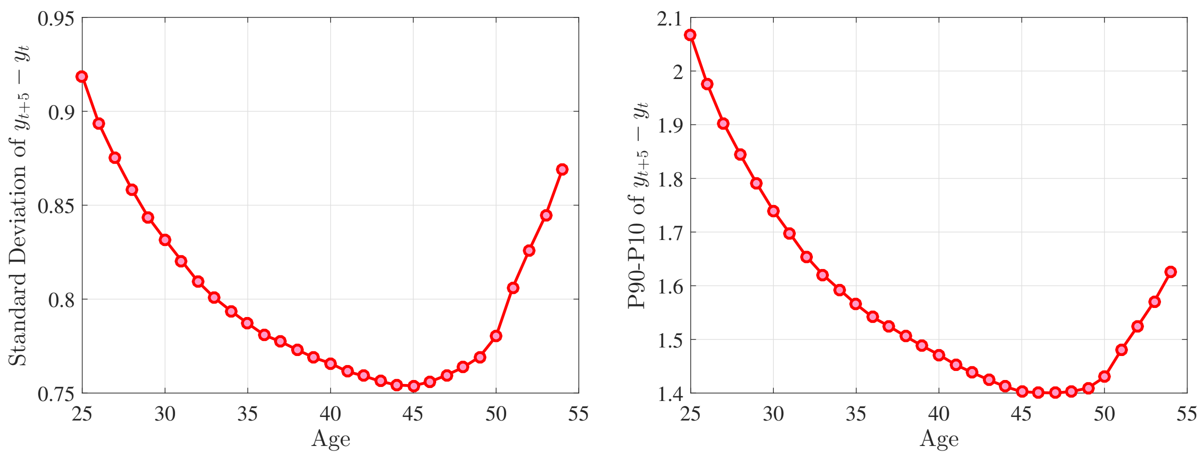


Figure C.27 – Skewness of Five-Year Log Earnings Growth by Age

(a) Third Standardized Moment

(b) Kelley Measure

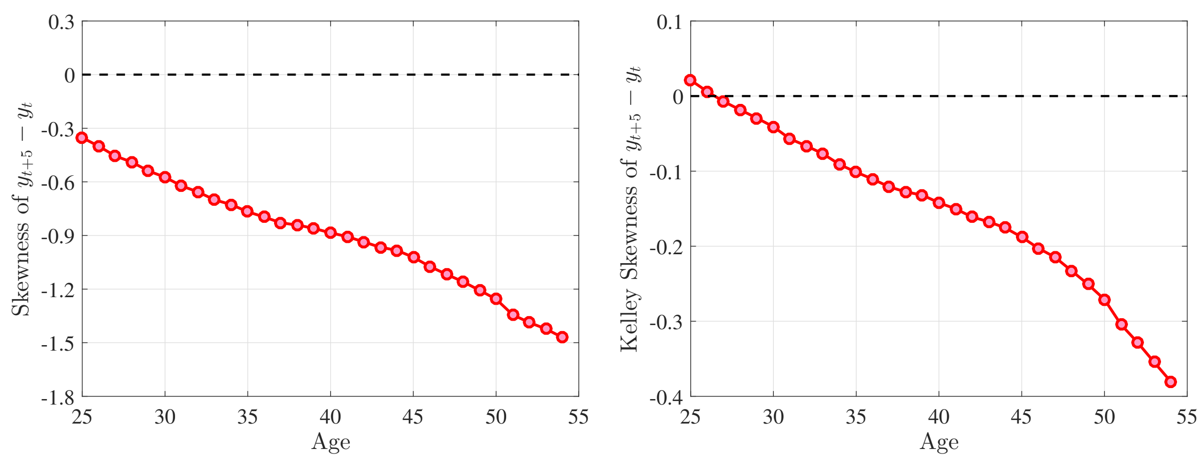
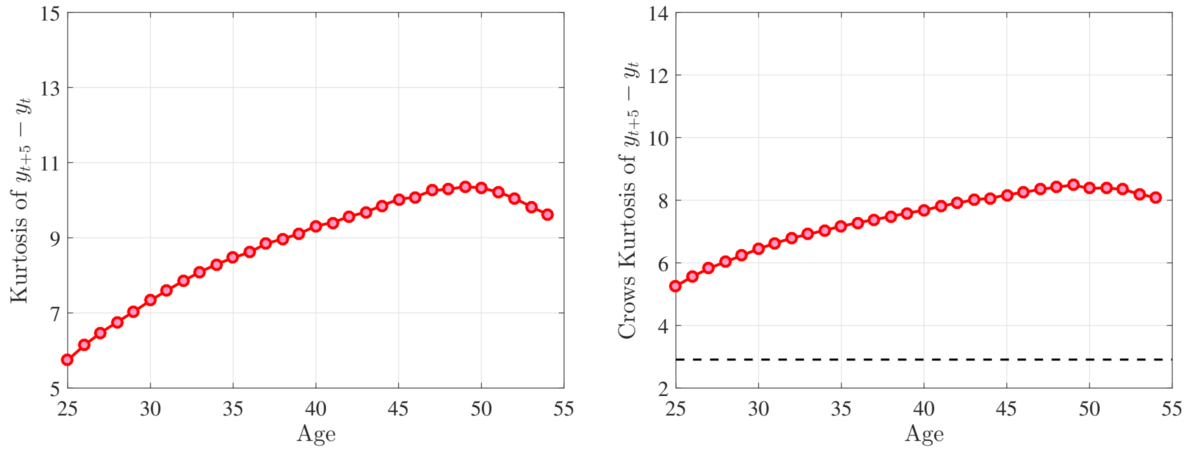


Figure C.28 – Kurtosis of Five-Year Log Earnings Growth by Age

(a) Fourth Standardized (b) Crow-Siddiqui Measure



C.9 The Role of Disability Income

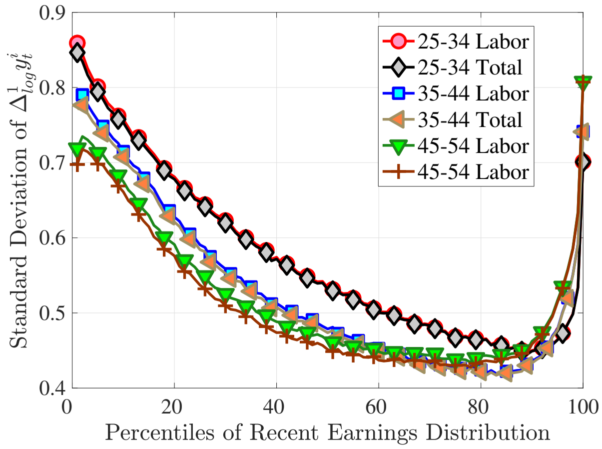
In this section, we investigate the robustness of our findings to the inclusion of income from Social Security disability benefits (SSDI). This is particularly relevant for thinking about the tails of the distribution of earnings changes. To this end, we link to our dataset information about disability benefits from the SSA records. We then define a measure of "total income" as the sum of labor income and annual disability income. Section C.9.1 compares the cross-sectional moments of earnings changes to our baseline and section C.9.2 does the same for lifetime income growth.

C.9.1 Cross-sectional Moments

Figures C.29–C.34 show several moments of one-year and five-year earnings changes. In each figure and for each age group, we show these moments for labor income and total income separately, where total income is labor income plus disability benefits. For each measure of income, recent earnings is re-calculated using that measure. Otherwise, these graphs are calculated in an analogous fashion as described in section 3. The main finding here is that the inclusion of disability income has little effect on cross-sectional moments, even at low levels of earnings.

Figure C.29 – Dispersion of One-Year Log Earnings Growth

(a) Second Standardized Moment



(b) P90-P10

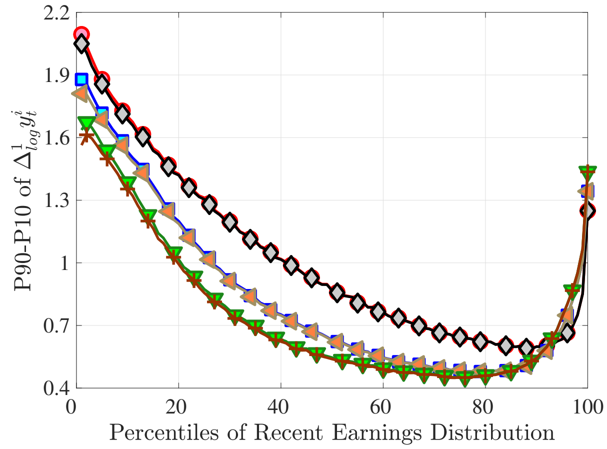
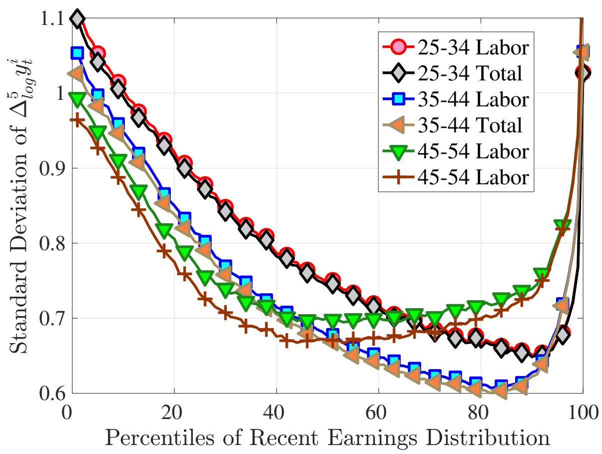


Figure C.30 – Dispersion of Five-Year Log Earnings Growth

(a) Second Standardized Moment



(b) P90-P10

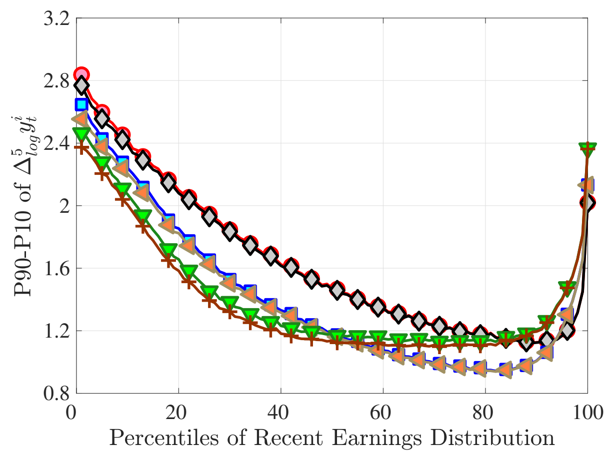


Figure C.31 – Skewness of One-Year Log Earnings Growth

(a) Third Standardized Moment (b) Kelley Measure

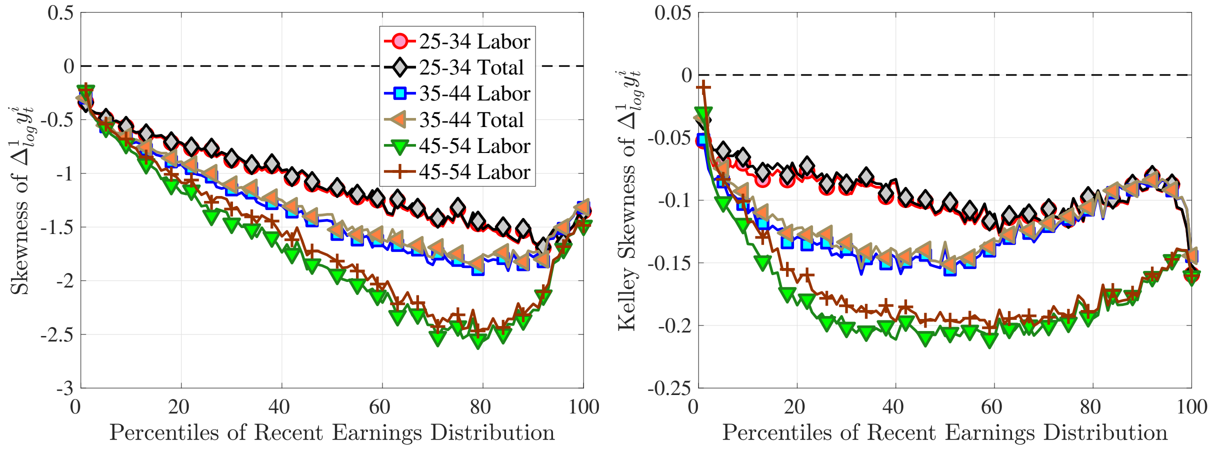


Figure C.32 – Skewness of Five-Year Log Earnings Growth

(a) Third Standardized Moment (b) Kelley Measure

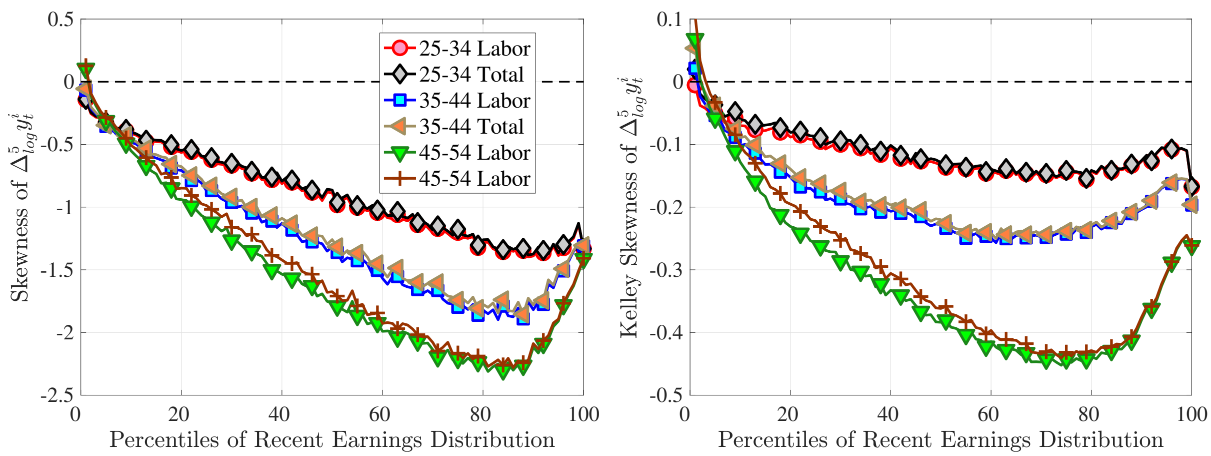


Figure C.33 – Kurtosis of One-Year Log Earnings Growth

(a) Fourth Standardized (b) Crow-Siddiqui Measure

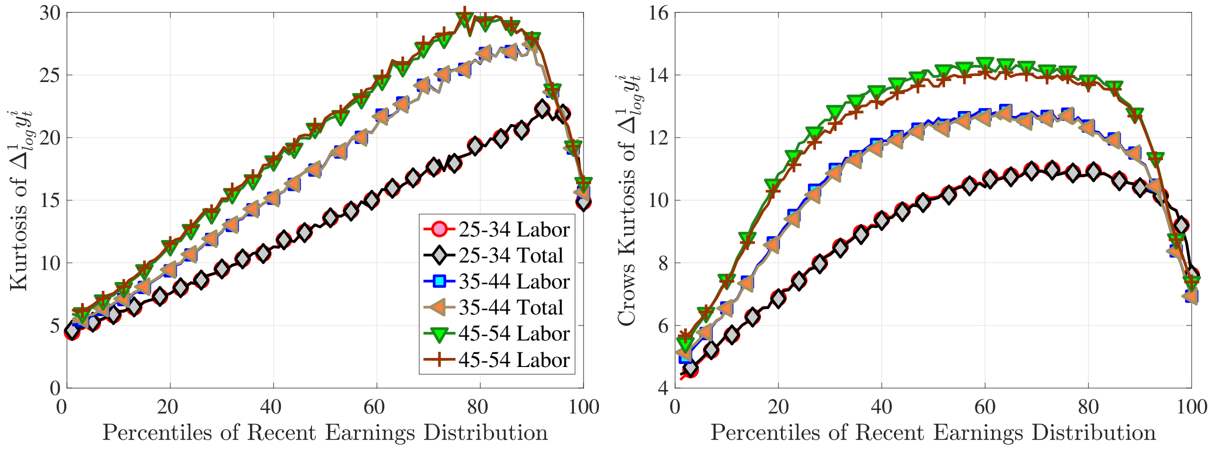
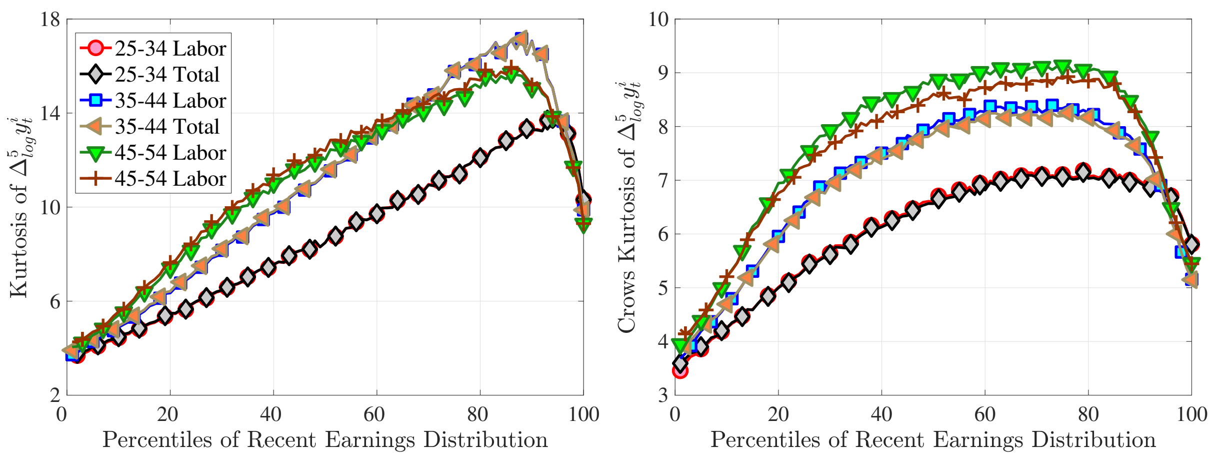


Figure C.34 – Kurtosis of Five-Year Log Earnings Growth

(a) Fourth Standardized (b) Crow-Siddiqui Measure

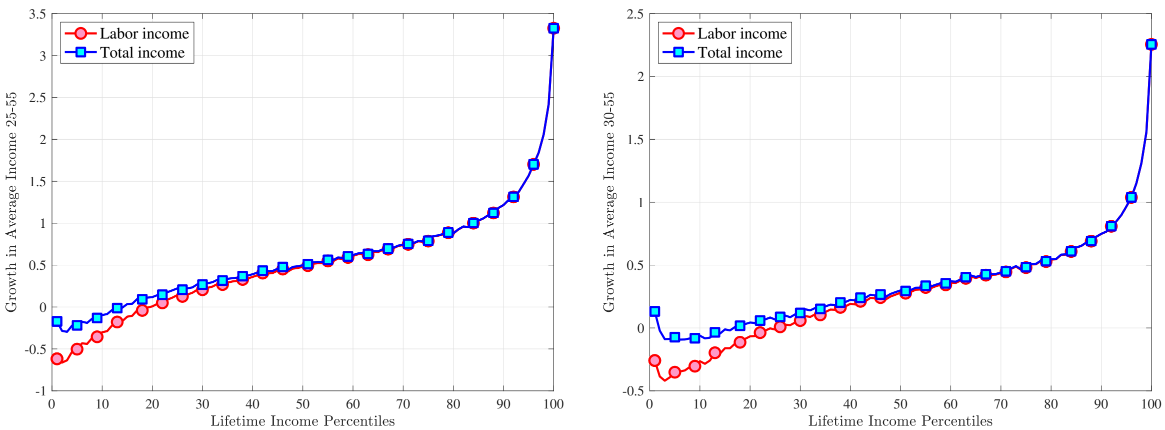


C.9.2 Lifetime income growth moments

Figure C.35 shows growth in average earnings over the life cycle. The left panel shows (log) growth in average earnings between ages 25 and 55, whereas the right panel does the same for 30 and 55. We consider two measures: Labor income and total income (including disability benefits). Figure C.35 plots income growth for the two measures against lifetime income. To allow comparability across the two measures, we use labor income to construct each individual's lifetime income and impose the sample selection criteria based on this measure. This allows us to keep the same overall sample as well as the same people in each lifetime income group. The differences in the two series are therefore only due to disability payments. We find that SSDI makes a difference for the income growth of the bottom LE individuals. For the bottom 1%, we find that SSDI can undo around 50% of earnings declines over the lifetime. This contribution declines gradually and vanishes by around 20th percentile of the LE distribution. This result is not very surprising since bottom LE workers are more likely to be claiming disability benefits.

FIGURE C.35 – Life-Cycle Earnings Growth Rates, by Lifetime Earnings Group

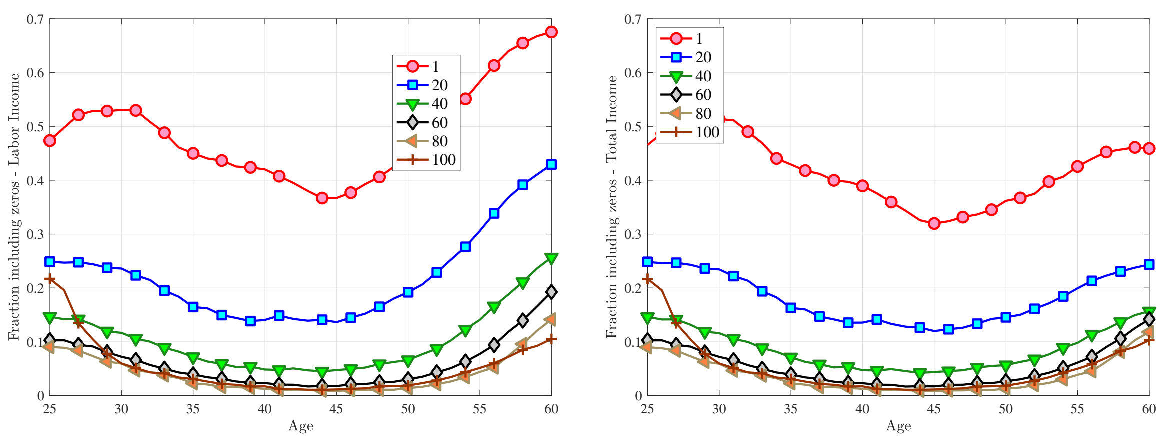
(A) 25–55 (B) 30–55



C.10 Concentration of nonemployment by lifetime earnings group

FIGURE C.36 – Nonemployment Concentration over the Life Cycle, by Lifetime Earnings Group

(A) Fraction Nonemployed wrt Labor Income (B) Fraction Nonemployed wrt Total Income



In this section, we investigate how concentrated (full-year) nonemployment is. We rank individuals by their lifetime earnings and group them into percentiles. For each lifetime earnings group, we compute what fraction of full-year nonemployment at a given age is accounted for by that group. These are shown in Figure C.36. For example, the bottom one percent of the lifetime earnings distribution accounts for around 50 percent of total nonemployment at ages 25–30.

D Estimation

In this section we describe the steps of our estimation procedure of method of simulated moments (MSM) in more detail.

D.1 Moment Selection and Aggregation

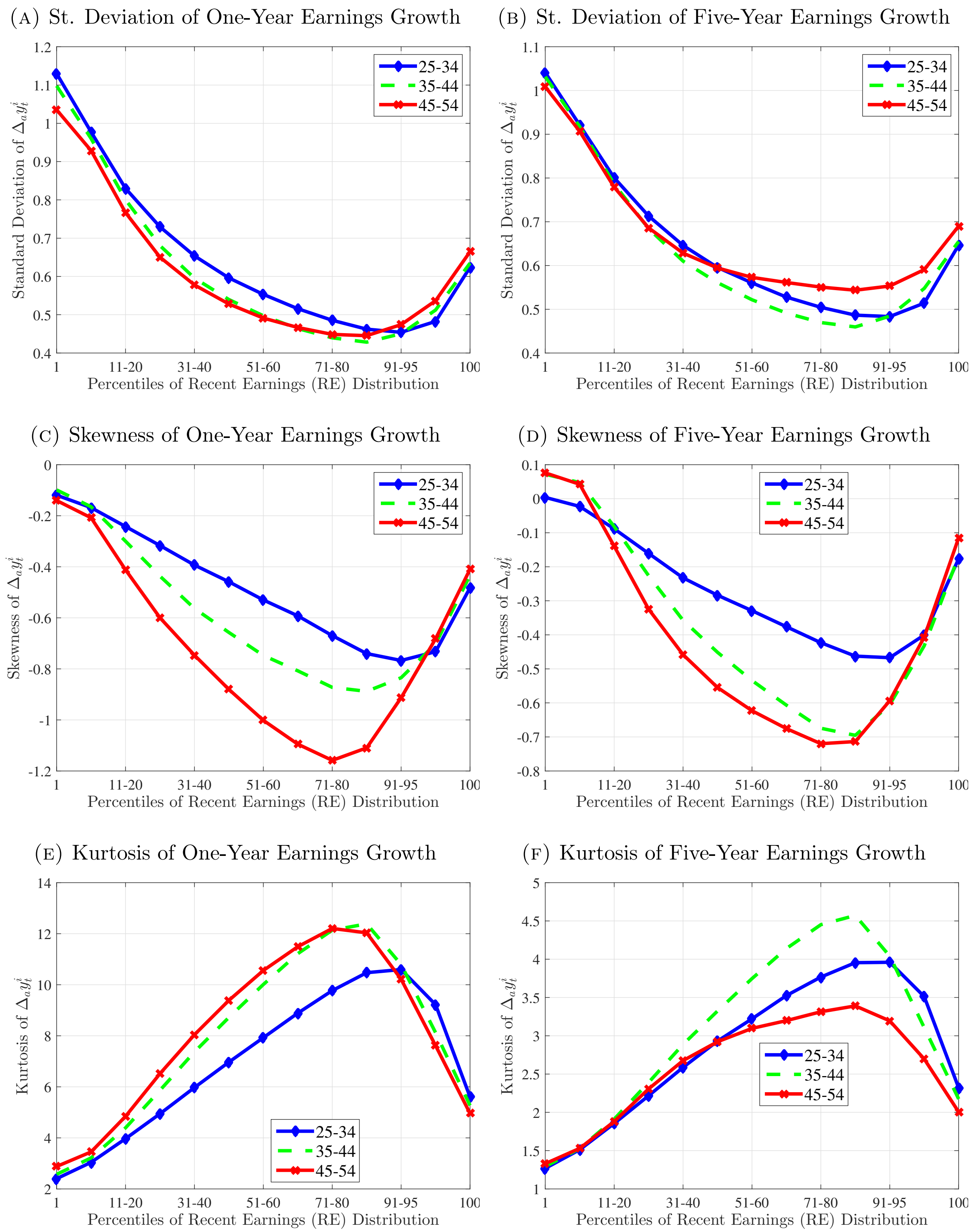
Accounting for Zeros. Recall that in order to construct the cross-sectional moments of log growth, we have dropped individuals who had very low earnings—below $\overline{Y}_{\min}$—in year t or $t+k$ so as to allow taking logarithms in a sensible manner. Although this approach made sense for documenting empirical facts that are easy to interpret, for the estimation exercise, we would like to also capture the patterns of these "zeros" (or very low earnings observations), given that they clearly contain valuable information. Targeting log growth moments also creates technical issues with the optimization due to (little) jumps in the objective function as workers cycle in and out of employment. To this end, instead of targeting moments of log earnings change, we target moments of arc percent change, as defined in Section 3. According to this measure, an income change from any positive level to 0 corresponds to an arc percent change of -2, whereas an income change from 0 to any positive level indicates an arc percent change of 2.

Aggregating Moments. If we were to match all data points for every RE percentile and every age group, it would yield more than 10,000 moments. Although this step is doable, not much is likely to be gained from such a level of detail, and it would make the diagnostics—that is, judging the performance of the estimation—quite difficult. To avoid this, we aggregate the 100 RE percentiles and the 6 age groups into fewer homogenous groups. We discuss the full details of this aggregation and the moments targeted in our estimation below:

1. Cross-sectional moments of earnings changes. In order to capture the variation in the cross-sectional moments of earnings changes along the age and recent earnings dimensions, we condition the distribution of earnings changes on these variables. For this purpose, we first group workers into 6 age bins (five-year age bins between 25 and 54) and within each age bin into 13 selected groups of RE percentiles in age $t-1$. The RE percentiles are grouped as follows: 1, 2–10, 11–0, 21–30, ..., 81–90, 91–95, 96–99, 100. Thus, we compute the three moments of the distribution of one- and five-year earnings changes for $6 \times 13 = 78$ different groups of workers. We aggregate these 6 age bins into 3 age groups, A^i_{t-1}. The first age group is defined as young workers between ages 25 through 34, the second is between ages 35-44, whereas the third age group is defined as workers between the ages of 35 and 54. Consequently, we target three centralized moments moments (i.e., standard deviation, skewness, and kurtosis) of one- and five-year arc percent change for three age and 13 recent earnings group, giving us $3 \times 2 \times 3 \times 13 = 234$ cross-sectional moments. These moments are shown in Figure D.1.

2. Mean of log earnings growth. The second set of moments captures the heterogeneity in log earnings growth over the working life across workers that are in different percentiles of the LE distribution. We target the average dollar earnings at 8 points over the life cycle: ages

Figure D.1 – Centralized Cross Sectional Moments of Arc-Percent Growth Targeted in the Estimation



25, 30, ..., and 60 for different LE groups. We combine LE percentiles into larger groups to keep the number of moments at a manageable number, yielding 15 groups consisting of percentiles of the LE distribution: 1, 2–5, 6–10, 11–20, 21–30,..., 81–90, 91–95, 96–97, 98–99, and 100. The total number of moments we target in this set is $8 \times 15 = 120$.

3. Impulse response functions. We target average arc percent change in earnings over the next k years for $k = 1, 2, 3, 5, 10$ conditional on groups formed by crossing age, recent earnings $\overline{Y}_{t-1}$, and earnings change between $t-1$ and t $\Delta^1_{arc} Y^i_{t-1}$: $\mathbb{E}[\Delta^{k+1}_{arc} Y^i_{t-1} \mid age, \overline{Y}_{t-1}, \Delta^1_{arc} Y^i_{t-1}]$.[46] In each year, we first group workers into two age bins, denoted by h: young workers (25-34) and prime-age workers (35-55). Then, within each age group individuals are ranked into the following RE percentiles, denoted by j: 1–5, 6–10, 11–30, 31–50, 51–70, 71–90, 91–95, 96–100.

Within each age h and RE group j, we then estimate our targets for the persistence of earnings growth. For each k-year expected future earnings growth we create a piecewise linear function of arc percent growth between t and $t-1$, i.e., $\mathbb{E}_{h,j}[\Delta^{k+1}_{arc} Y^i_{t-1} \mid \Delta^1_{arc} Y^i_{t-1}] = f^k_{h,j}(\Delta^1_{arc} Y^i_{t-1})$. For this purpose, we condition workers in the data into 23 groups with respect to their ranking in the $\Delta^1_{arc} Y^i_{t-1}$ distribution, defined by the following percentiles: 1–2, 3–5, 6–10, 11-15, 16-20, 21-25, ..., 91–95, 96–98, 99–100. Then the piecewise linear function $f^k_{h,j}(\Delta^1_{arc} Y^i_{t-1})$ for year-k is determined by the linear interpolation of 23 data points of average earnings growth between $t-1$ and t, $\mathbb{E}\left[\Delta^1_{arc} Y^i_{t-1}\right]$ and their corresponding k-year future expected growth, $\mathbb{E}[\Delta^k_{arc} Y^i_t]$.

For the model-simulated data, we group workers into 2×8 age and RE groups—similar to the data moments. But within each age h and RE group j we now rank workers into 10 $\Delta^1_{arc} Y^i_{t-1}$ groups defined by the following percentiles: 1–2, 3–5, 6–10, 11–30, 31–50, 51–70, 71–90, 91–95, 96–98, 99–100. In the estimation of income processes we minimize the distance between 10 different values of $\mathbb{E}^{model}_{h,j}[\Delta^{k+1}_{arc} Y^i_{t-1} \mid \Delta^1_{arc} Y^i_{t-1}]$ (for each age and RE group and k-year expectation) from the model and its corresponding data moment from the piecewise linear function, $f^k_{h,j}(\Delta^1_{arc} Y^i_{t-1})$. As a result, we have a total of $2 \times 8 \times 5 \times 10 = 800$ moments based on impulse response.

The impulse response functions targeted in the estimation are plotted in Figures D.2a–D.2d (to keep the figures similar to their counterparts in Section 4, we plot $\mathbb{E}[\Delta^{k+1}_{arc} Y^i_{t-1} \mid \Delta^1_{arc} Y^i_{t-1}] - \mathbb{E}[\Delta^1_{arc} Y^i_{t-1}]$). More specifically, Figure D.2a plots for prime age workers with median recent earnings, the mean reversion patterns at various horizons. Figures D.2b and D.2c do the same for workers at the 90th and 10th percentiles of the recent earnings distribution, respectively. Lastly, Figure D.2d shows the variation of these impulse response functions with recent earnings.

4. Variance of log earnings. Although the main focus of this section is on earnings *growth*, the life-cycle evolution of the dispersion of earnings *levels* has been at the center of the incomplete markets literature since the seminal paper of Deaton and Paxson (1994). For completeness, and comparability with earlier work, we have estimated the within-cohort variance of log earnings over the life cycle and report it in Figure D.3. We target the life cycle profile of the variance of log earnings for income observations larger than our minimum income threshold. As a result, we have a total of 36 moments based on the variance of log earnings.

[46] Notice that, different from the moments we have shown in Section 4, we target earnings growth between $t+k$ and $t-1$. This is because all workers have $\tilde{Y}^i_{t-1} \geq Y_{min,t-1}$ in $t-1$ by construction of

5. CDF of Employment over the Life Cycle. We target the distribution of total number of years employed ($\tilde{Y}^i_{t,h} \geq \overline{Y}_{\min,t}$) over life cycle. In particular, we target the fraction of workers who have worked at least 6, 14, 20, 25, 30, 33 over a 36-year working life (shown on Figure 11b). Thus, in total there are 6 such moments targeted in our estimation.

In sum, we target a total of $234 + 120 + 800 + 36 + 6 = 1,196$ moments in our estimation.[47]

D.2 Numerical Method for Estimation

Objective Function Let m_n for $n = 1, ..., N = 1,196$ denote a generic empirical moment, and let $d_n(\theta)$ be the corresponding model moment that is simulated for a given vector of earnings process parameters, θ. We simulate the entire earnings histories of 100,000 individuals who enter labor market at age 25 and work until age 60. When computing the model moments, we apply precisely the same sample selection criteria and employ the same methodology to the simulated data as we did with the actual data. To deal with potential issues that could arise from the large variation in the scales of the moments, we minimize the *scaled* arc percent deviation between each data target and the corresponding simulated model moment. For each moment n, define

$$F_n(\theta) = \frac{d_n(\theta) - m_n}{0.5\left(|d_n(\theta)| + |m_n|\right) + \psi_n}, \tag{11}$$

where $\psi_n > 0$ is an adjustment factor. When $\psi_n = 0$ and m_n is positive, F_n is simply the (arc) percentage deviation between data and model moments. This measure becomes problematic when the data moment is very close to zero, which is not unusual (e.g., impulse response of arc percent earnings changes close to zero). To account for this, we choose ψ_n to be equal to the 10th percentile of the distribution of the absolute value of the moments in a given set. The MSM estimator is

$$\hat{\theta} = \arg\min_{\theta} \boldsymbol{F}(\theta)' W \boldsymbol{F}(\theta), \tag{12}$$

where $\boldsymbol{F}(\theta)$ is a column vector in which all moment conditions are stacked, that is,

$$\boldsymbol{F}(\theta) = \left[F_1(\theta), ..., F_N(\theta)\right]^T.$$

The weighting matrix, W, is chosen such that we first assign 15% relative weight to the employment CDF moments. The rest of the moments share the remaining 85% weight according to this scheme: the life-cycle average earnings growth moments and impulse response moments are assigned a relative weight of 0.25 each, the cross-sectional moments of earnings growth receive a relative weight of 0.35, and the variance of log earnings is given a relative weight of 0.15.[48]

Numerical Method The global stage is a multi-start algorithm where candidate parameter vectors are uniform Sobol (quasi-random) points. We typically take about 250,000 initial Sobol

the RE sample. Thus, we can compute the arc-percent growth between $t + k$ and $t - 1$ for all workers, which keeps the composition of workers constant in each k.

[47] The full set of moments targeted in the estimation are reported (in Excel format) as part of an online appendix available from the authors' websites.

[48] More precisely, each employment CDF moment is weighed by 0.15/6, life-cycle growth moment is weighed $0.85 * 0.25/120$, each cross-sectional moment by $0.85 * 0.35/234$, each impulse response moment by $0.85 * 0.25/800$, and each variance moment by $0.85 * 0.15/36$.

points for pre-testing and select the best 1,000 (i.e., ranked by objective value) for the multiple restart procedure depending on the number of parameters to be estimated. For processes with a large number of parameters to be estimated (e.g. the benchmark process or the 2-state process), we also tried using 300,000 initial Sobol points and used the best 2,000 of them. We found this wider search for parameter values to be inconsequential for our estimates. The local minimization stage is performed with a mixture of Nelder-Mead's downhill simplex algorithm (which is slow but performs well on difficult objectives) and the DFNLS algorithm of Zhang *et al.* (2010), which is much faster but has a higher tendency to be stuck at local minima. We have found that the combination balances speed with reliability and provides good results.

D.3 2-State Process

We also estimate a more flexible income process which has *2* AR(1) components, denoted by z_1 and z_2, each subject to innovations from a mixture of two normals with age and income dependent shock probabilities. Furthermore, we allow the variance of each innovation to be individual-specific in the spirit of Browning *et al.* (2010). Here is the full specification:

$$Y_t^i = (1-\nu_t^i)\exp\left(g(t) + \alpha^i + \beta^i t + z_{1,t}^i + z_{2,t}^i + \varepsilon_t^i\right) \tag{13}$$

$$z_{1,t}^i = \rho_1 z_{1,t-1}^i + \eta_{1t}^i \tag{14}$$

$$z_{2,t}^i = \rho_2 z_{2,t-1}^i + \eta_{2t}^i, \tag{15}$$

where $t = (age - 24)/10$ denotes normalized age, and for $j = 1, 2$

$$\text{Innovations to AR(1):} \quad \eta_{j,t}^i \sim \begin{cases} \mathcal{N}(\mu_{z_j,1}, \sigma_{z,1}) & \text{with pr. } p_{z_j,t} \\ \mathcal{N}(\mu_{z_j,2}, \sigma_{z,2}) & \text{with pr. } 1 - p_{z_j,t} \end{cases} \tag{16}$$

$$\text{Initial value of AR(1) process:} \quad z_{j0}^i \sim \mathcal{N}(0, \sigma_{j0}). \tag{17}$$

$$\text{Nonemployment shocks: equation 7} \tag{18}$$

$$\text{Transitory shock: equation 6} \tag{19}$$

Each AR(1) component, z_1 and z_2, receives a shock drawn from a mixture of 2 Gaussian distributions as in our benchmark specification. We again normalize the mean of innovations to the persistent components to zero; i.e., $\mu_{z_j,1}p_{z_j} + \mu_{z_j,2}(1 - p_{z_j}) = 0$.[49] We also allow for heterogeneity in the initial conditions of the persistent processes, $z_{1,0}^i$ and $z_{2,0}^i$, given in equation (18). Since the specifications of z_1 and z_2 are the same so far, we need an identifying assumption to distinguish between the two, so, without loss of generality, we impose $\rho_1 < \rho_2$.

The age and income dependence of moments is captured by allowing the mixture probabil-

[49] We don't have to make the identification assumption of $\mu_{z_j,1} < 0$ as we did for the benchmark process, because the first Gaussian, $\mathcal{N}(\mu_{z_j,1}, \sigma_{z,1})$ is already different than the second one $\mathcal{N}(\mu_{z_j,2}, \sigma_{z,2})$ by having a mean $\mu_{z_j,1}$ constant over age and income, whereas $\mu_{z_j,2}$ varies by income and age. The latter is because p_{z_j} is a function of persistent components and age.

ities to depend on age and the sum of persistent components $(z_1 + z_2)$:[50]

$$p^i_{jt} = \frac{\exp\left(\xi^i_{j,t-1}\right)}{1+\exp\left(\xi^i_{j,t-1}\right)}, \tag{20}$$
$$\xi^i_{jt} = a_{z_j} + b_{z_j} \times t + c_{z_j} \times \left(z^i_{1,t} + z^i_{2,t}\right) + d_{z_j} \times \left(z^i_{1,t} + z^i_{2,t}\right) \times t,$$

for $j = 1, 2$. The equation for $p_{\nu t}$ is the same as (20) but $\xi^i_{j,t-1}$ is replaced with ξ^i_{jt}. This completes the description of the 2-state process.

D.4 Additional Estimation Results

This section contains estimation results not reported in the main text. We first report the estimates of additional specifications. Then, for all of the estimated income processes, we report some of the parameters that were not reported in Table IV. A comprehensive set of parameters for all income processes are available online for download as an Excel file on authors' websites.

Model Fit: Additional Figures

Deviations of estimated models from targeted values In the main text, we compared the model counterparts of targeted moments to the data. In this section, we show how the fit looks through the lens of our objective function in (11). Figure D.9 shows these for several key moments. More specifically, we plot equation 11 for each set of moments, with the exception of income growth moments. Recall that in our estimation we target the *levels* of income at various ages of different LE percentiles and not the life cycle growth rates.

Results for Additional Specifications The first column on Table D.1 reports the standard errors of our benchmark process using a parametric bootstrap. In calculating the bootstrap standard errors we first simulate data using our estimates reported on table IV and create target moments from simulated data. We then run the estimation for 100 different seeds of random variables by targeting the moments obtained in the previous step. For each seed of random variables we run the estimation once by employing simplex algorithm around our original parameter estimates. Using 100 sets of parameter values we then compute the standard errors.

Column 9 report the parameter estimates for 2-state income process defined in Section D.3 and their bootstrap standard errors.

The next columns—columns 10 to 17— report the estimates of 8 different specifications that are not reported in the main text. Columns (10), (11), and (12) are different versions of Column (7) of table (IV). Namely, we model nonemployment probability as a logistic function of a number of combinations of individual fixed effect α and persistent component z (similar to equation (8)). In particular, in column (10) nonemployment probability is a quadratic function of α. In column (11) nonemployment probability p_ν is assumed to be a linear function of $\alpha + z$ and age and their interaction. In column (12) nonemployment probability depends linearly on α and z and their interaction.

[50] We have also considered an alternative specification where the innovation variances are functions of earnings and age. After extensive experimentation with this formulation, we have found it to perform very poorly.

Columns (13) and (14) are similar to our benchmark specification. Again they only differ in how the nonemployment shock probability is modeled. In column (13) p_ν is a quadratic function of z. In column (14) p_ν depends on z, z^2 and age and the interaction of z and age.

In column (15) we introduce variance heterogeneity to the 1-state benchmark process. In particular, we allow the variance of each innovation from $\mathcal{N}(\mu_{z,1}, \sigma^i_{z,1})$ to the persistent component be individual-specific, with a lognormal distribution with mean $\overline{\sigma}_{z,1}$ and a standard deviation proportional to $\widetilde{\sigma}_{z,1}$, i.e., $\log\left(\sigma^i_{z,1}\right) \sim \mathcal{N}(\log \overline{\sigma}_{z,1} - \frac{\widetilde{\sigma}^2_{z,1}}{2}, \widetilde{\sigma}_{z,1})$.

In the next income process (column (16)), in the 1-state benchmark process we incorporate age and income dependence to the mixture probability in innovations to the persistent component similar to equation (20).

In column (17) we introduce ex ante variance heterogeneity in the 2-state benchmark process. Thus the variance of each innovation from $\mathcal{N}(\mu_{z_j,1}, \sigma^i_{z_j,1})$ to the persistent component j be individual-specific, with a lognormal distribution with mean $\overline{\sigma}_{z_j,1}$ and a standard deviation proportional to $\widetilde{\sigma}_{z_j,1}$, i.e., $\log\left(\sigma^i_{z_j,1}\right) \sim \mathcal{N}(\log \overline{\sigma}_{z_j,1} - \frac{\widetilde{\sigma}^2_{z,1}}{2}, \widetilde{\sigma}_{z_j,1})$.

In column (18) we replicate our parameter estimates for the Gaussian specification that is estimated in the previous version of the paper in which we didn't target the distribution of years worked.

The last column (column (19)) shows the parameter estimates for the specification presented in column (5) but without imposing a lower bound for the mean of persistent shocks.

Parameter Estimates Table D.2 contains several parameters that were not reported in the main text on table IV due to space constraints. These parameters include deterministic life cycle profile and the coefficients on age and income in the probability functions.

Since it is difficult to interpret the magnitudes of the coefficients on shock probabilities, in Figures D.10 and D.11 we visually show the estimated relationship between shock probabilities and age and income for the benchmark specification. For the 2-state income process, we also report the probability of drawing a nonemployment shock for various age and RE percentile groups for workers who satisfy the conditions of the base sample in Table D.3.

TABLE D.1 – Standard Errors and Additional Specifications

Specification:		(8)	(9)		(10)	(11)	(12)	(13)	(14)	(15)	(16)	(17)	(18)	(19)
		Benchmark Std. Errors	**2-state Process Std. Errors**		**Unemp. depends on α**	**Unemp. depends on $\alpha + z$ & age**	**Unemp. depends on α and z**	**Unemp. depends on z and z^2**	**Unemp. depends on z, z^2, & age**	**1-state Benchmark w/ var. het.**	**1-state Benchmark w/ η het.**	**2-state Benchmark w/ var. het.**	**Gaussian No Emp. CDF**	**Column (5) No Bound for $\mu_{z,1}$**
AR(1) Component		mix	2 mixtures		mix	mix	mix	mix	mix	mix	mix	2 mixtures	G	mix
↪Probability age/inc.		no/no	yes/yes		no/no	no/no	no/no	no/no	no/no	no/no	yes/yes	yes/yes	—	no/no
Nonemployment shocks		yes	yes		yes	yes	yes	yes	yes	yes	yes	yes	no	no
↪Probability age/inc.		yes/yes	yes/yes		yes/α	yes/$\alpha + z$	no/α, z	no/z, z^2	yes/z, z^2	yes/yes	yes/yes	yes/yes	—	—
Transitory Shocks		mix	mix		mix	mix	mix	mix	mix	mix	mix	mix	G	mix
HIP		yes	yes		no	no	no	yes	yes	yes	yes	yes	no	no
Variance Heterogeneity		no	no		no	no	no	no	no	yes	no	yes		no
Parameters			est.	se.										
ρ_1		0.0001	0.821	0.0004	0.847	0.978	0.976	0.968	0.964	0.960	0.965	0.824	1.0	1.005
ρ_2			0.980	0.0004								0.979		
p_{z_1}		0.0002	—	—	0.044	0.150	0.091	0.219	0.427	0.274				0.0124
$\mu_{\eta_1,1}$		0.0005	-0.373	0.0010	-0.961	-0.327	-0.576	-0.335	-0.120	-0.107	-0.436	-0.393		-5.63
$\mu_{\eta_2,1}$			-0.272	0.0006								-0.270		
$\sigma_{\eta_1,1}$		0.0006	0.597	0.0004	1.396	0.689	0.575	0.304	0.345	0.433	0.788	0.620	0.156	0.471
$\sigma_{\eta_1,2}$		0.0006	0.121	0.0004	0.066	0.083	0.168	0.174	0.061	0.0827	0.195	0.116		0.148
$\sigma_{\eta_2,1}$			0.574	0.0004								0.564		
$\sigma_{\eta_2,2}$			0.003	0.0003								0.001		
$\sigma^i_{\eta_1,1}$										0.162		0.002		
$\sigma^i_{\eta_2,1}$												0.013		
$\sigma_{z_1,0}$		0.0006	0.199	0.0004	0.339	0.154	0.193	0.719	0.689	1.5042	0.530	0.189		0.227
$\sigma_{z_2,0}$			0.605	0.0004								0.603		
λ		0.001	0.001	0.0003	0.196	0.104	0.022	0.001	0.002	0.005	0.014	0.001		
$a_\nu \times$	1	0.0047		0.0036	-3.875	-2.399	-3.191	-4.131	-3.115	-2.958	-3.773	-3.045		
$b_\nu \times$	t, α	0.0038		0.0024	-5.366	-1.241	-1.932		-1.108	-1.341	0.468	-1.092		
$c_\nu \times$	$[z_t, \alpha, (\alpha + z_t)]$	0.0039		0.0025		-3.550	-5.528	-5.477	-3.914	-4.222	-3.635	-3.219		
$d_\nu \times$	α^2, z^2				0.293			1.611	0.832					
$e_\nu \times$	$t[z_t, (\alpha + z_t)]$	0.0058		0.0043		-1.837	0.700		-2.403	-3.450	-3.560	-2.240		
$prob_\varepsilon$		0.0004	9.5%	0.0001	0.227	0.104	0.239	0.290	0.105	0.130	0.115	0.095		0.210
$\mu_{\varepsilon,1}$		0.0007	0.349	0.0007	0.115	0.249	0.146	0.159	0.296	0.223	0.340	0.340		-0.09
$\sigma_{\varepsilon,1}$		0.001	0.444	0.0006	0.449	0.562	0.239	0.142	0.247	0.384	0.283	0.438	0.494	1.024
$\sigma_{\varepsilon,2}$		0.0003	0.040	0.0002	0.061	0.042	0.072	0.022	0.064	0.048	0.081	0.025		0.024
σ_α		0.0006	0.272	0.0003	0.808	0.520	0.640	0.289	0.274	0.313	0.267	0.273	0.514	0.4367
$\sigma_\beta \times 10$		0.0002	0.180	0.0002				0.194	0.205	0.217	0.204	0.182		
$\mathrm{corr}_{\alpha\beta}$		0.0013	0.450	0.0007				0.630	0.676	0.489	0.974	0.435		
Objective value			18.78		40.33	28.28	26.19	23.32	22.11	21.81	21.74	18.43		47.9
Decomposition:														
(i) Standard deviation			4.82		4.86	4.81	5.31	5.86	5.93	5.09	5.95	4.95		7.92
(ii) Skewness			7.49		21.15	15.76	14.67	12.760	10.65	10.09	9.52	7.06		22.99
(iii) Kurtosis			4.64		6.67	6.04	4.80	5.63	5.88	6.32	6.70	4.70		13.99
(iv) Impulse resp.			12.35		27.98	18.19	15.82	13.54	13.50	13.25	13.70	11.58		36.40
(v) Inc. growth			8.22		15.10	10.81	10.39	9.21	8.39	8.24	7.84	8.37		8.94
(vi) Inequality			1.92		7.70	3.38	1.66	3.36	3.47	3.68	3.16	2.23		4.99
(vii) Nonemployment CDF			5.28		6.46	5.71	7.66	5.94	6.43	7.14	6.52	5.86		8.91

Table D.2 – Additional Parameter Estimates for Estimated Specification

Specification:		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Parameters										
					Deterministic Life Cycle Profile Parameters					
a_0	$\times 1$	0.7275	0.7902	2.6352	1.9265	2.6920	2.2661	2.7408	2.6291	2.5180
a_1	$\times t$	0.7513	0.3637	0.7213	-0.4222	-0.0786	0.2200	0.4989	0.7300	0.5666
a_2	$\times t^2$	-0.0539	0.0884	-0.1401	0.0165	-0.0746	-0.1268	-0.1137	-0.1692	-0.1296
					Nonemployment Shock Probability Function Parameters					
a_ν	$\times 1$		0.0231	-2.6429				-2.8571	-3.2740	-3.2131
b_ν	$\times t$			-1.2028				-0.7788	-0.8935	-1.0235
c_ν	$\times z_t$			-5.2467				-4.5407	-4.5692	-3.2602
d_ν	$\times t \times z_t$			-4.2236				-1.3702	-2.9203	-2.1656
					Normal Mixture Probability Function Parameters					
a_{z_1}	$\times 1$				0.0630	0.0572	-0.7834	0.2665	0.2191	-2.1797
b_{z_1}	$\times t$						1.9560			0.9374
c_{z_1}	$\times z_{t-1}$						-3.5429			-1.5598
d_{z_1}	$\times t \times z_{t-1}$						-0.0497			-1.5786
a_{z_2}	$\times 1$									-0.1858
b_{z_2}	$\times t$									-1.0438
c_{z_2}	$\times z_{t-1}$									-0.9686
d_{z_2}	$\times t \times z_{t-1}$									0.6316

Note: We define deterministic life cycle profile as a quadratic function of t, $g(t) = a_0 + a_1 t + a_2 t^2$, where $t = (age - 24)/10$. $y_t = z_t$ for the 1-state income process with 1 AR(1) component and $y_t = z_1 + z_2$ for 2-state income process with 2 AR(1) components.

Table D.3 – Mixture Probabilities for 2-State Process

	Age groups			RE (Percentile) groups				
	25–34	35–49	45–60	1–10	21–30	41–60	71–80	91–100
p_{z_1} ($\rho_{z_1} = 0.82$)	0.120	0.142	0.170	0.416	0.204	0.112	0.051	0.012
p_{z_2} ($\rho_{z_2} = 0.98$)	0.236	0.138	0.086	0.166	0.156	0.150	0.147	0.151
p_ν (nonemp.)	0.065	0.052	0.051	0.186	0.077	0.038	0.016	0.003
Pr (any large shock)	0.325	0.263	0.248	0.526	0.333	0.248	0.193	0.161

Notes: This table reports the probabilities of innovations with large standard deviations vary by age and past income. In particular, the first row reports the probability of drawing innovations to the $z_1$persistent component from the first normal distribution, $z_{1,1} \sim \mathcal{N}(-0.373, 0.597)$. And similarly, the second row presents the probability of drawing innovations to the $z_2$persistent component from the first normal distribution, $z_{2,1} \sim \mathcal{N}(-0.272, 0.574)$. The last row reports the probability of any one of the events happening in the first three rows.

Figure D.2 – Impulse Response Moments Targeted in the Estimation, Prime-Age Workers

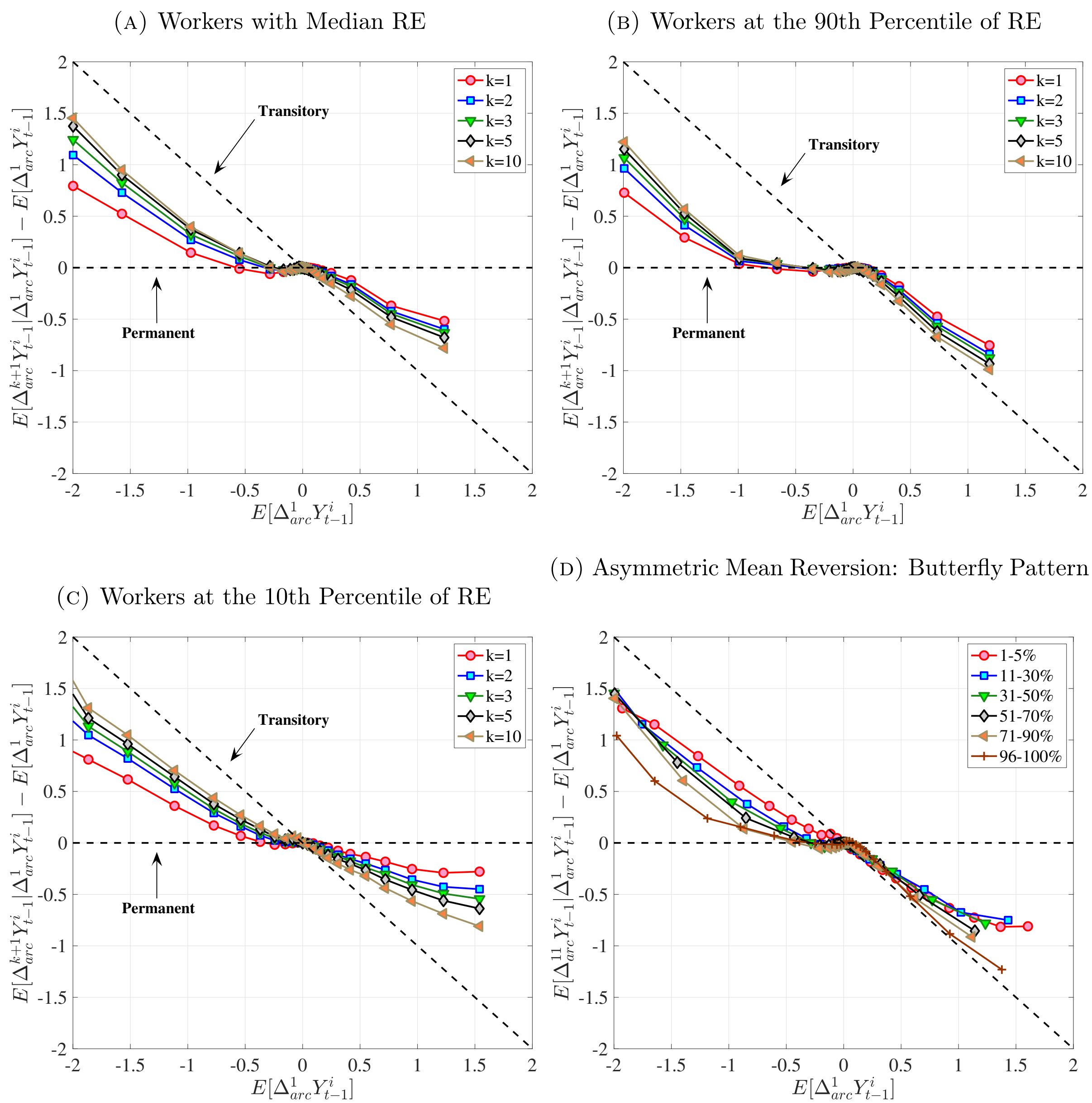


Figure D.3 – Within-Cohort Variance of Log Earnings

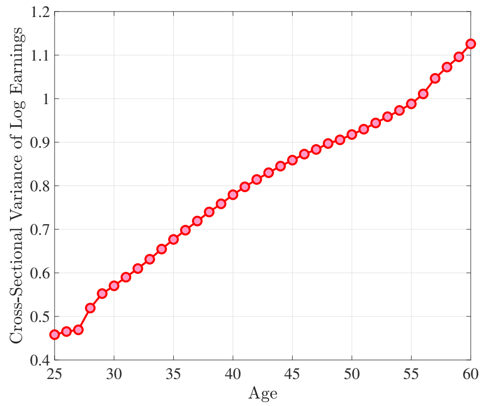


Figure D.4 – Fit of Estimated Model on Cross-Sectional Moments of One-Year Earnings Changes

(A) Std. dev. (B) Skewness (C) Kurtosis

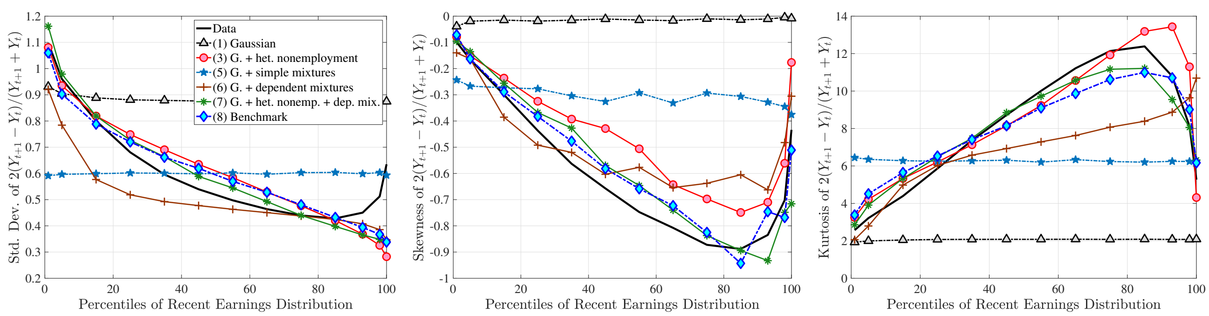


Figure D.5 – Model Fit on Cross-Sectional Moments of One-Year Earnings Changes by Age

(A) Standard deviation of $\Delta^1_{\mathrm{arc}} Y^i_t$ (B) Skewness of $\Delta^1_{\mathrm{arc}} Y^i_t$ (C) Kurtosis of $\Delta^1_{\mathrm{arc}} Y^i_t$

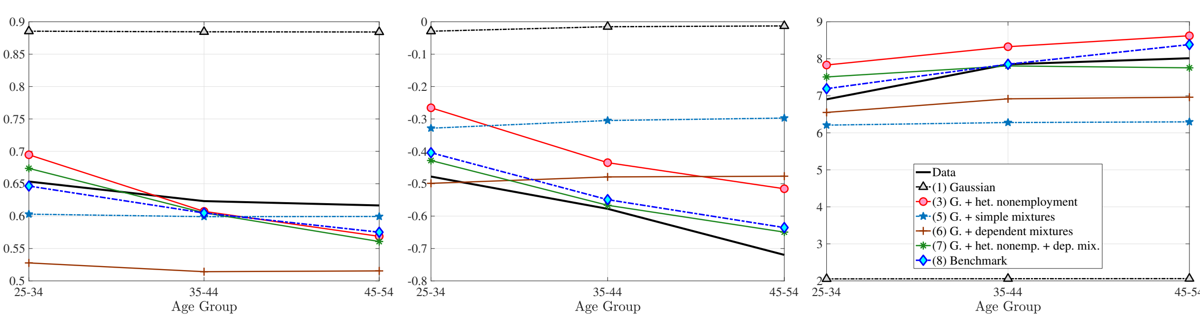


Figure D.6 – Model Fit on Cross-Sectional Moments of Five-Year Earnings Changes by Age

(a) Standard deviation of $\Delta^5_{\text{arc}} Y^i_t$ (b) Skewness of $\Delta^5_{\text{arc}} Y^i_t$ (c) Kurtosis of $\Delta^5_{\text{arc}} Y^i_t$

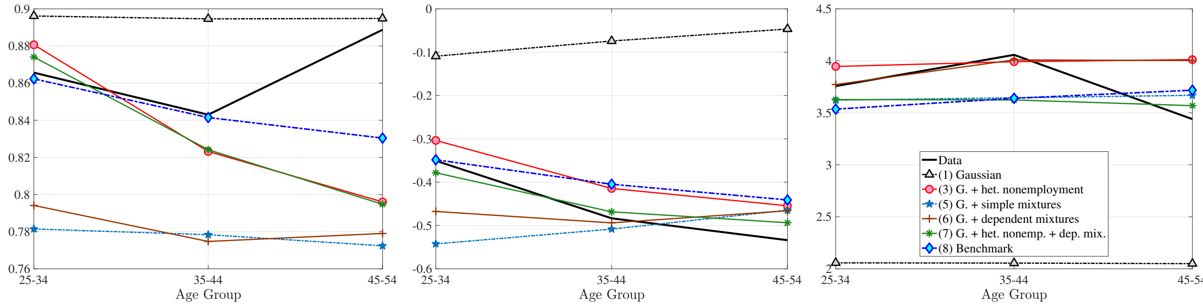


Figure D.7 – Cross-Sectional Moments of One-Year Earnings Changes by Age and Recent Earnings: Benchmark, Gaussian and Data

(a) Standard deviation of $\Delta^1_{\text{arc}} Y^i_t$ (b) Skewness of $\Delta^1_{\text{arc}} Y^i_t$ (c) Kurtosis of $\Delta^1_{\text{arc}} Y^i_t$

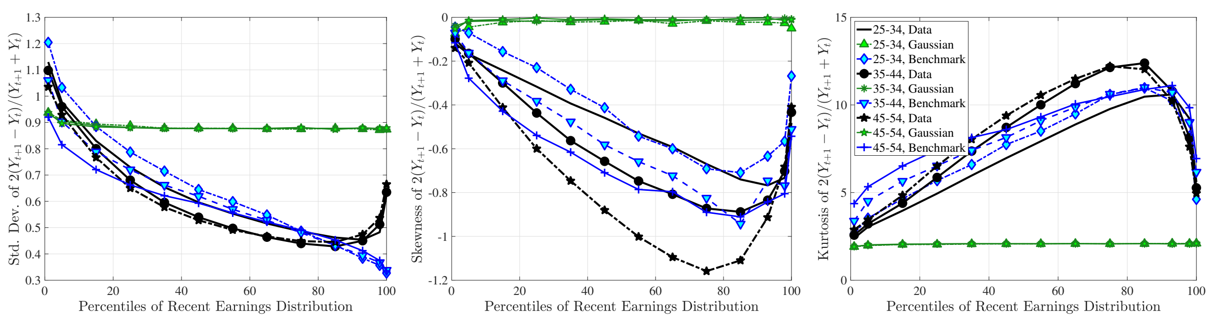


Figure D.8 – Cross-Sectional Moments of Five-Year Earnings Changes by Age and Recent Earnings: Benchmark, Gaussian and Data

(a) Standard deviation of $\Delta^5_{\text{arc}} Y^i_t$ (b) Skewness of $\Delta^5_{\text{arc}} Y^i_t$ (c) Kurtosis of $\Delta^5_{\text{arc}} Y^i_t$

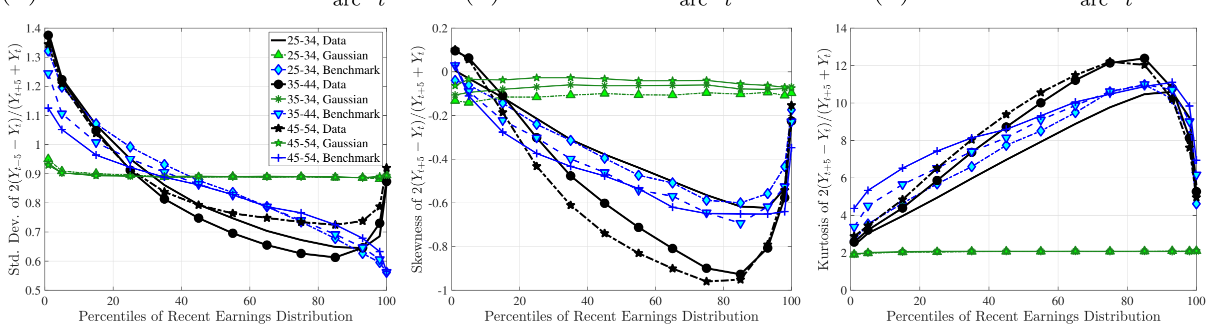
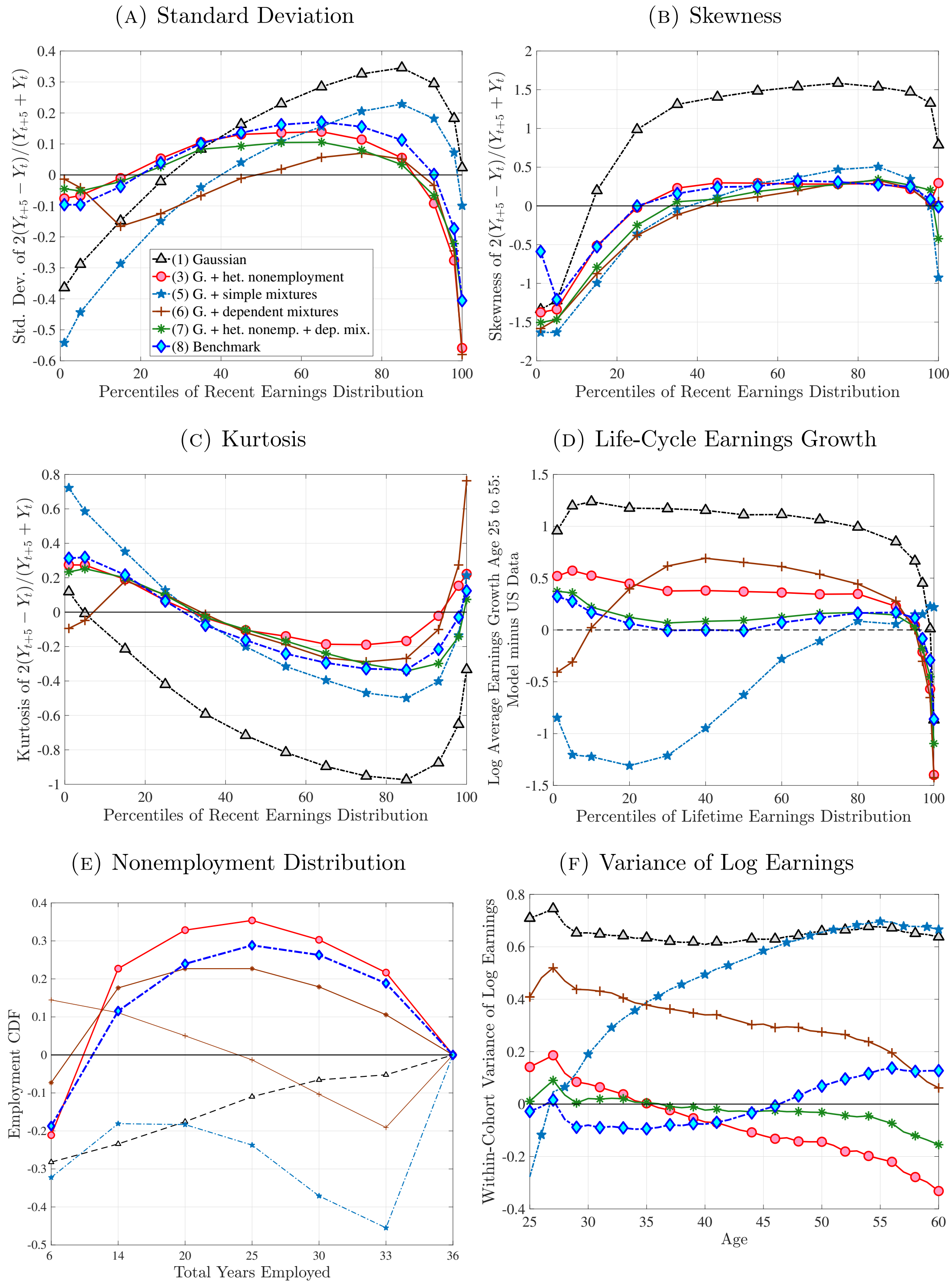


Figure D.9 – Deviations of Model Moments from Data Moments



Note: This figure shows the deviations of each estimated model through the lens of our objective function. With the exception of Figure D.9d, we plot equation (11) for each set of moments. For income growth moments in Figure D.9d, we plot the difference between the income growth of a given LE percentile in the model and the data.

FIGURE D.10 – Benchmark Process: 3-D Plot of Nonemployment Shock Probability p_ν

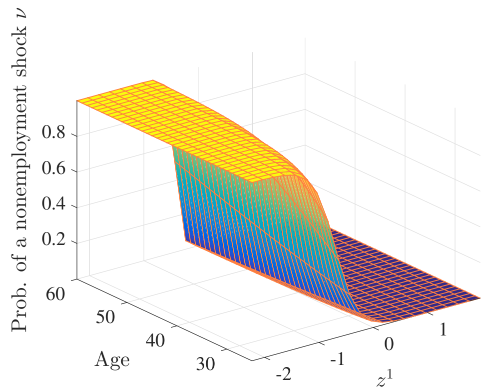
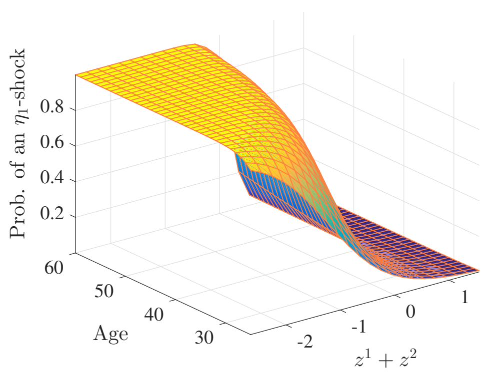
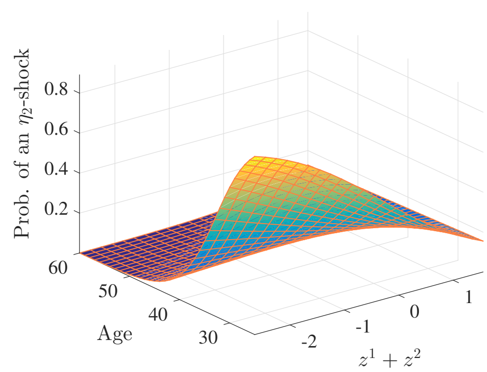


FIGURE D.11 – 2-State Process: 3-D Plot of Mixing Probabilities for the

(A) Mixing Probability p_{z^1} for z_t^1



(B) Mixing Probability p_{z^2} for z_t^2



(C) Shock Probability p_ν for ν_t