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

Fatih Guvenen   Fatih Karahan   Serdar Ozkan   Jae Song

August 7, 2016

Abstract

We study individual earnings dynamics over the life cycle using panel data on millions of U.S. workers. Using nonparametric methods, our analysis reaches two broad conclusions. First, the distribution of earnings changes displays substantial deviations from lognormality—the standard assumption in the incomplete markets literature. In particular, earnings changes display strong negative skewness and extremely high kurtosis. The high kurtosis implies that in a given year, most individuals experience very small earnings changes, and a nonnegligible number experience very large changes. Second, these deviations from nonnormality vary significantly both over the life cycle and with the earnings level of individuals. We also estimate impulse response functions and find important asymmetries: the large positive earnings changes of high-income individuals are quite transitory, whereas large negative changes are very persistent; the opposite is true for low-income individuals. We then estimate econometric processes for earnings dynamics by targeting these rich sets of moments, which allows us to link facts on earnings "changes" to underlying "shocks." We solve and simulate a life-cycle consumption-savings model using the resulting process. The idiosyncratic income fluctuations we document generate large welfare costs, and the implications for wealth inequality and partial insurance differ from those of a Gaussian process in important ways.

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

1 Introduction

This year about four million young Americans will enter the labor market for the first time. Over the next 40 years, each of these individuals will go through their unique adventure in the labor market, involving a series of surprises—finding an attractive career, being offered a dream job, getting promotions and salary raises, and so on—as well as disappointments—experiencing unemployment, failing in one career and moving on to another one, suffering health shocks, and so on. These events will vary not only in their initial significance (upon impact) but also in how durable their effects turn out to be in the long run.

An enduring question for economists is whether these wide-ranging labor market histories, experienced by a diverse set of individuals, display sufficiently simple regularities that would allow researchers to characterize some general properties of earnings dynamics over the life cycle. Despite a vast body of research since the 1970s, it is fair to say that many aspects of this question remain open. For example, what does the probability distribution of earnings shocks look like? Is it more or less symmetric, or is it skewed one way? More generally, how well is it approximated by a lognormal distribution, an assumption often made out of convenience? And, perhaps more important, how do these properties differ among low- and high-income workers or change over the life cycle? A host of questions also pertain to the dynamics of earnings. For example, how sensible is it to think of a single persistence parameter to characterize the durability of earnings shocks? Do positive shocks exhibit persistence that is different from negative shocks? Clearly, we can add many more questions to this list, but we have to stop at some point. If so, which of these many properties of earnings shocks are the most critical in terms of their economic importance and therefore should be included in this short list, and which are of second-order importance?

This paper aims to answer these questions by characterizing the most salient aspects of life-cycle earnings dynamics using a large and confidential panel data set from the U.S. Social Security Administration covering 1978 to 2013. The substantial sample size allows us to employ a fully nonparametric approach and take what amounts to high-resolution pictures of individual earnings histories.[1]

---

[1] In this paper, we focus on the earnings dynamics of men for comparability to earlier work. However, we have replicated the analysis in this paper for women and found qualitatively very similar patterns. These results are reported in an online appendix that is available from the authors' websites.

In deciding what aspects of the earnings data to focus on, we were motivated by a growing body of theoretical work (reviewed in the next section), which attributes a central role to skewness and kurtosis of economic variables for questions ranging from the effects of monetary policy to optimal taxation, and from the determinants of wealth inequality to asset prices. Therefore, we focus on the first four moments of earnings changes over the life cycle. This analysis reaches two broad conclusions. First, the distribution of individual earnings growth displays important deviations from lognormality. Second, the magnitude of these deviations (as well as a host of other statistical properties of earnings growth) varies greatly both over the life cycle and with the earnings level of individuals. Under this broad umbrella of "nonnormality and life-cycle variation," we establish four sets of empirical results.

First, earnings growth is negatively skewed, and this skewness becomes more severe as individuals get older or their earnings increase (or both). Furthermore, this increasing negativity is entirely due to upside earnings moves becoming smaller from ages 25 to 45, and to increasing "disaster" risk (i.e., the risk of a sharp fall in earnings) after age 50. Although these implications may appear quite plausible, they are not captured by a lognormal specification, which implies zero skewness. (It should be stressed that the nonzero skewness of earnings *levels* is well known since at least Pareto (1897), who documented that both earnings and log earnings are right-skewed—see Neal and Rosen (2000) for a survey. The analysis of this paper is focused on earnings *changes*—not levels—and with the few exceptions discussed in the next section, far less is known about its higher order moments.)

Second, earnings growth displays very high kurtosis. What kurtosis measures is most easily understood by looking at the histogram of log earnings changes, shown in Figure 1 (left panel: annual change; right panel: five-year change). Notice the sharpness in the peak of the empirical density, how little mass there is on the "shoulders" (i.e., the region around $\pm\sigma$), and how long the tails are compared with a normal density chosen to have the same standard deviation as in the data. Thus, there are far more people with very small earnings changes in the data compared with what would be predicted by a normal density. Furthermore, this *average* kurtosis masks significant heterogeneity across individuals by age and earnings: prime-age males with earnings of $100,000 (in 2010 dollars) face annual earnings growth rates with a kurtosis coefficient as high as 30, whereas young workers with earnings of $10,000 face a kurtosis of only 5 (relative to a Gaussian distribution, which has a kurtosis of 3).
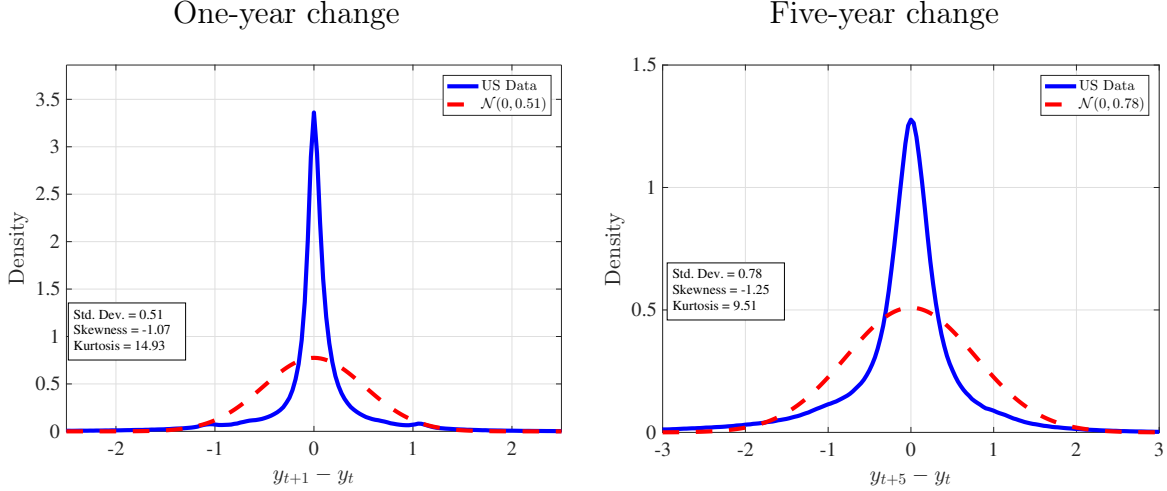
FIGURE 1 – Histogram of Log Earnings Changes. Note: The first year t is 1997, and the data are for all workers in the base sample defined in Section 2.

Third, we characterize the dynamics of earnings shocks by estimating nonparametric impulse response functions conditional on the recent earnings of individuals and on the size of the shock that hits them. We find two types of asymmetries. One, fixing the shock size, positive shocks to high-earnings individuals are quite transitory, whereas negative shocks are very persistent; the opposite is true for low-earnings individuals. Two, fixing the earnings level of individuals, the strength of mean reversion differs by the size of the shock: large shocks tend to be much more transitory than small shocks. To our knowledge, both of these findings are new in the literature. These kinds of asymmetries are hard to detect via the standard approach in the literature, which relies on the autocovariance (i.e., second moment) matrix of earnings.[2]

Fourth, there is a strong positive relationship between the level of lifetime earnings and how much earnings grow over the life cycle. For example, the median individual by lifetime earnings experiences an earnings growth of 60% from ages 25 to 55, whereas for individuals in the 95th percentile, this figure is 380%; for those in the top 1 percent, this figure is almost 2600% (or 26-fold!).[3]

While the nonparametric approach (Sections 3–5) allows us to establish key features of earnings dynamics in a robust fashion, a tractable parametric process is indispensable

---

[2]These asymmetries are difficult to detect because a covariance lumps together all sorts of earnings changes—large, small, positive, and negative—into a single statistic.

[3]A positive relationship between lifetime earnings and life-cycle earnings growth is to be expected (since, all else equal, fast earnings growth will lead to higher lifetime earnings). What is surprising is the large magnitudes involved.

for two reasons: (i) it allows us to connect earnings *changes* to underlying *innovations* or *shocks* to earnings,[4] and (ii) it allows us to conduct quantitative economic analyses in a straightforward manner. Therefore, in Section 6, we estimate a rich-yet-tractable stochastic process for earnings dynamics by targeting the four sets of empirical moments described above.[5] The two key features of this process are (i) nontrivial higher-order moments and (ii) nonlinear shock accumulation over time. Specifically, it features the sum of two AR(1) processes with normal mixture innovations as well as long-term nonemployment shocks, whose probabilities vary by age and earnings levels. This structure turns out to be important for matching the rich variation in higher-order moments as well as the asymmetric mean reversion patterns we observe in the data.

Using the estimated earnings process, in Section 7 we revisit four important questions in the incomplete markets literature. To this end, we first embed the estimated stochastic process into a life-cycle consumption savings model (and show that this can be done relatively straightforwardly). We then establish four results about how the implications of the estimated process differs from those of the more familiar Gaussian process (with the same variance): (i) the welfare costs of idiosyncratic fluctuations are about *two to three times* higher, reaching as high as 25% to 40% of lifetime consumption; (ii) wealth inequality is higher, but not high enough to match the right tail observed in the data; (iii) caution must be exercised in measuring the degree of partial insurance because different approaches yield conflicting results under nonnormality; and (iv) consumption growth displays strong higher-order moments—negative skewness and excess kurtosis—as well as higher volatility. These results show that accounting for higher-order moments of earnings dynamics can have a first-order effect on some important economic questions, and they invite more work in order to understand its implications in other settings.

---

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

[5]The nonparametric analysis yields more than 10,000 empirical moments of individual earnings data, which are made available for download as an Excel file (from the authors' websites) to enable researchers to estimate their preferred specification(s).

Related Literature

Since its inception in the late 1970s,[6] the earnings dynamics literature has worked with the implicit or explicit assumption of a Gaussian framework, thereby making no use of higher-order moments beyond the variance-covariance matrix. One of the few exceptions is an important paper by Geweke and Keane (2000), who emphasize the non-Gaussian nature of earnings shocks and fit a normal mixture model to earnings innovations. More recently, Bonhomme and Robin (2010) analyze French earnings data over short panels and model the transitory component as a mixture of normals and the dependence patterns over time using a copula model. They find the distribution of this transitory component to be left skewed and leptokurtic. In this paper, we go beyond the overall distribution and find substantial variation in the degree of nonnormality with age and earnings levels. Guvenen et al. (2014b) also documented the negative skewness of earnings changes but abstracted completely from life cycle variation, instead focusing on business cycle patterns in skewness. Furthermore, in this paper, the impulse response analysis shows the need for a different persistence parameter for large and small shocks, which is better captured as a mixture of two AR(1) processes—a step beyond the normal mixture model.

Incorporating higher-order moments of earnings dynamics into economic models is still in its infancy. Guvenen et al. (2014b) document that the skewness of individual income shocks becomes more negative in recessions (i.e., skewness is procyclical), whereas the variance is acyclical. Building on this observation, Constantinides and Ghosh (2016) and Schmidt (2016) show that an incomplete markets asset pricing model with procyclical skewness generates plausible asset pricing implications, and McKay (2014) studies aggregate consumption dynamics in a business cycle model that is calibrated to match these skewness fluctuations. Turning to fiscal policy, Golosov et al. (2016) show that using an earnings process with negative skewness and excess kurtosis (targeting the empirical moments reported in this paper) implies a marginal tax rate on labor earnings for top earners that is substantially higher than under a traditional calibration with Gaussian shocks with the same variance. Similarly, Kaplan et al. (2016) introduce leptokurtic earnings growth into a New Keynesian model with household heterogeneity. They show that excess kurtosis has important effects on the monetary transmission mechanism when calibrated to empirical values from the present paper.

---

[6]Earliest contributions include Lillard and Willis (1978) and MaCurdy (1982), among others.

Our work is related to some important recent contributions. Altonji *et al.* (2013) 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 an earnings process featuring "lots of heterogeneity" (as they call it). However, neither paper explicitly focuses on higher-order moments or their life-cycle evolution. The latter paper does model heterogeneity across individuals in innovation variances, as do we, and finds a lot of heterogeneity along that dimension in the data. In contemporaneous work, Arellano *et al.* (2016) also explore differences in the mean-reversion patterns of earnings shocks across households that differ in their earnings histories. Using data from the Panel Study of Income Dynamics (PSID), they find asymmetries in mean reversion that are consistent with those that we document in Section 4.

The analysis of consumption-savings decisions in the presence of higher-order income risk has also been conducted previously in Wang (2007), and recently in Wang *et al.* (2016) and De Nardi *et al.* (2016). Our analysis complements these papers by embedding a very general stochastic process (complete with age and income dependent shocks) with minimal modifications into a life-cycle model and analyzing a range of its properties.

2 Data

The data used in this paper are drawn from the Master Earnings File (MEF) of the U.S. Social Security Administration (SSA). Starting in 1978, the MEF combines various databases available to the SSA in one master file. For our purposes, the most important of these variables include wage/salary earnings from employees' W-2 forms (for each job held by the employee during the year), self-employment income (obtained from IRS tax forms—in particular, Schedule SE), as well as various demographics (date of birth, sex, and race). In this paper, we focus on *total annual labor earnings*, which is the sum of total annual wage/salary income (summed across all jobs) plus the labor portion of self-employment income.[7]

Wage earnings are not top coded throughout our sample, whereas self-employment income was top coded 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 Section 5, where we

---

[7]The labor portion of self-employment income is taken to be 2/3, which is a fairly standard value.

analyze average income growth over the life cycle (i.e., the first moment), and therefore a longer time series becomes essential. Thus, in that section, we use data from 1981 to 2013 and impute self-employment income above the top code for years before 1994.[8] Finally, nominal earnings records are converted into real values using the personal consumption expenditure (PCE) deflator, taking 2010 as the base year. Further details of our sample and variable construction can be found in Appendix A.

Sample Selection.

Our *base sample* is a revolving panel that maximizes sample size (important for precise computation of higher-order moments for finely defined groups) and allows a stable age structure over time. It is constructed as follows. First, an individual-year earnings observation must satisfy two criteria to be *admissible* for that year: the individual (i) must be between 25 and 60 years old (working age) and (ii) must have participated in the labor market—that is, have earnings above $Y_{\min,t}$. The threshold $Y_{\min,t}$ is chosen as the annual earnings level corresponding to one quarter of full-time work (13 weeks at 40 hours per week) at half of the legal minimum wage, which is approximately \$1,885 in 2010. This condition is fairly standard in the income dynamics literature and ensures that we select individuals with a reasonably strong labor market attachment (see, e.g., Abowd and Card (1989) and Meghir and Pistaferri (2004)).

For each year t, we will construct moments by conditioning on individuals' average earnings from $t-5$ to $t-1$ and compute statistics based on income changes from year t to $t+k$ ($k=1,5$); see Figure 2. The revolving panel for year t selects individuals for which this conditioning can be done in a sensible way. Thus, an individual is in the *base sample* if he is admissible in $t-1$ and in at least two more years between $t-5$ and $t-2$. This ensures that the individual was actively in the labor market and has earnings records from which we can compute a measure of average recent earnings, as we describe in a moment. For some statistics, such as change from y_t to y_{t+k}, we also require the individual to be admissible in t and $t+k$. This is specified later below.

Variable Construction.

Recent Earnings.   We now define a measure of what we call "recent earnings," a term used throughout the paper. Let $\tilde{y}_{t,h}^i \equiv \log(\tilde{Y}_{t,h}^i)$ denote the log earnings of individual

---

[8]Details of this imputation are provided in Appendix A.1, where we also show that only a small number of individuals are affected by this imputation (those who had very high self-employment earnings before 1994 or when the cohort was younger than age 41), so the effect on our results is minimal.
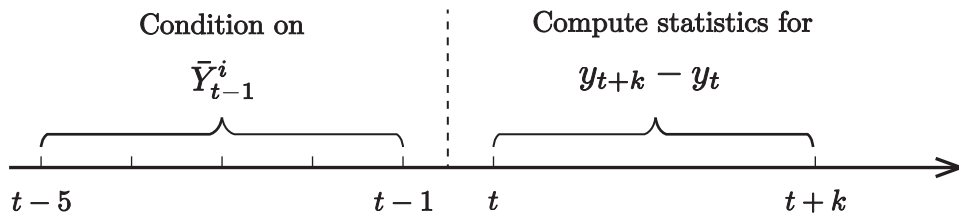
FIGURE 2 – Timeline for Rolling Panel Construction

i who is h years old in year t. (We will suppress dependence on age whenever it does not create confusion.) For a given worker, we compute his average earnings between years $t-1$ and $t-5$. We set earnings observations below $Y_{\min,t}$ to the threshold for this computation. We also control for age effects as follows. We first estimate age dummies, denoted d_h, by regressing log individual earnings on a full set of age and (year-of-birth) cohort dummies. We then construct five-year average earnings in the population from ages $h-5$ to $h-1$: $\sum_{s=1}^{5} \exp(d_{h-s})$. Finally, we normalize the worker's average past earnings with this measure to clean age effects. Thus, our measure of *recent earnings* (hereafter, RE) is

$$\bar{Y}^i_{t-1} \equiv \frac{\sum_{s=1}^{5} \tilde{Y}^i_{t-s,h-s}}{\sum_{s=1}^{5} \exp(d_{h-s})}.$$

In the next section, we will group individuals by age and by $\bar{Y}^i_{t-1}$ to investigate how the dynamics of earnings vary over the life cycle and by earnings levels.

Growth Rate Measures. We consider two measures of earnings growth rates, each with its own distinct advantages. Let $y^i_t = \tilde{y}^i_{t,h} - d_h$ denote log earnings net of age effects. The first growth measure is the familiar

$$\text{log change:} \quad \Delta_k y^i_t \equiv (y^i_{t+k,h+k} - y^i_{t,h}) = (\tilde{y}^i_{t+k,h+k} - d_{h+k}) - (\tilde{y}^i_{t,h} - d_h).$$

This measure is well known, and its higher-order moments for a lognormal distribution are familiar to readers (zero skewness and a kurtosis coefficient of 3). While this familiarity makes it a good choice for the nonparametric and descriptive analysis, it also has a well-known drawback: because log of zero is $-\infty$, we need to drop observations close to zero so as to obtain sensible statistics. Thus, when we use $\Delta_k y^i_t$ to compute a statistic, we drop individuals from the base sample whose data are not admissible (earnings below $Y_{\min,t}$) in t or $t+k$. This not only forces us to ignore valuable information in the extensive margin but also creates technical issues, by causing (little) jumps

8

in the empirical moments as workers cycle in and out of the labor force, which creates difficulties with the optimization of the objective function during estimation. For these reasons, we borrow a second and closely related measure, which is commonly used in the firm dynamics literature (see, e.g., Davis *et al.* (1996)) where firm entry and exit are key margins and lead to the same difficulties with logs. The measure is

$$\text{arc-percent change:} \quad \Delta_{\text{arc}} Y_{t,k}^i = \frac{Y_{t+k}^i - Y_t^i}{\left(Y_{t+k}^i + Y_t^i\right)/2}.$$

This measure allows computation of time differences even when the individual has zero income in one of the two years, thereby also capturing the extensive margin of earnings changes. As we discuss further below, the qualitative patterns we document are very robust regardless of the measure we use. Because of its familiarity, we report the log change measure in the descriptive analysis. In the estimation, we target the arc percent change moments and report all the analogous figures with this measure in appendix B.2.

3   Cross-Sectional Moments of Earnings Growth

In this section, we analyze the second to fourth moments of five-year earnings growth rates. Our focus on five-year (as opposed to one-year) growth rates is motivated by their closer correspondence to persistent changes in earnings and hence their greater significance for many economic questions. For completeness, we report the analogous figures for one-year earnings changes in Appendix B.1, which show the same patterns reported here for long-run changes. In Section 6, we will link these earnings "changes" to underlying "shocks" or "innovations" to an earnings process by means of a parametric estimation.

Our main focus is on how the moments of earnings growth vary with recent earnings and age. To this end, we first group individuals in the base sample into five-year age bins based on their age in year $t-1$: 25–29, 30–34,..., 50–54, and 55–60. For each year t, we group individuals based on their age and recent earnings as of time $t-1$ (RE: $\bar{Y}_{t-1}^i$). 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, we can compute the cross-sectional moments of earnings changes between t and $t+5$, which can then be viewed as corresponding to the properties of earnings changes that individuals within each bin can expect to face. This approach has

9

FIGURE 3 – Standard Deviation of Five-Year Earnings Growth

(A) Second Standardized Moment

(B) P90-P10
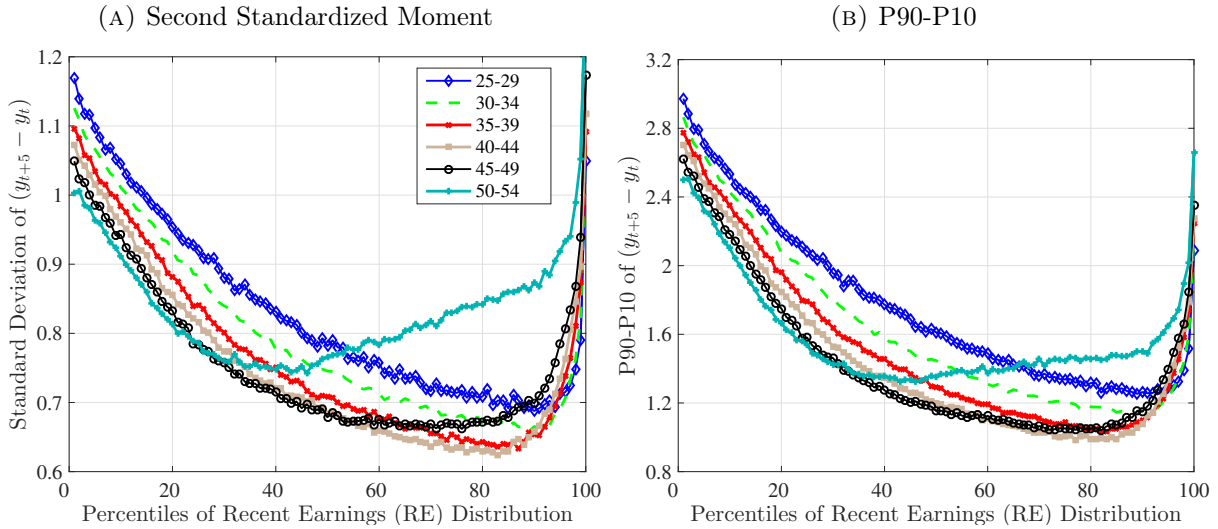
the advantage that we can compute higher-order moments precisely as the average bin contains more than 100,000 individuals. (Table A.1 in Appendix A reports sample size statistics.)

3.1 Second Moment: Variance

Figure 3a plots the standard deviation of five-year earnings growth by age and recent earnings (hereafter, RE) groups. Consistent with our nonparametric focus, we also report a quantile-based dispersion measure—the log difference between the 90th and 10th percentiles (P90 and P10, respectively) of earnings change—denoted P90-P10, in the right panel (Figure 3b). By both measures, and for every age group, there is a pronounced U-shape pattern by RE levels, implying that earnings changes are less dispersed for individuals with higher RE up to about the 90th percentile (along the x-axis). This pattern reverts itself inside the top 10% as dispersion increases rapidly with recent earnings. Further, over the life cycle, the dispersion of earnings changes declines monotonically up to about age 50 (with the exception of the very top earners) and then rises slightly for middle- to high-income individuals from ages 50 to 55.[9]

The life-cycle pattern is quite different for top earners, who experience a monotonic *increase* in dispersion of shocks over the life cycle. In particular, individuals at the 95th

---

[9]A similar pattern has been found by Baker and Solon (2003) using administrative data in Canada and by Karahan and Ozkan (2013) using data from the Panel Study of Income Dynamics.

FIGURE 4 – Skewness of Five-Year Earnings Growth

(A) Third Standardized Moment

(B) Kelley Measure
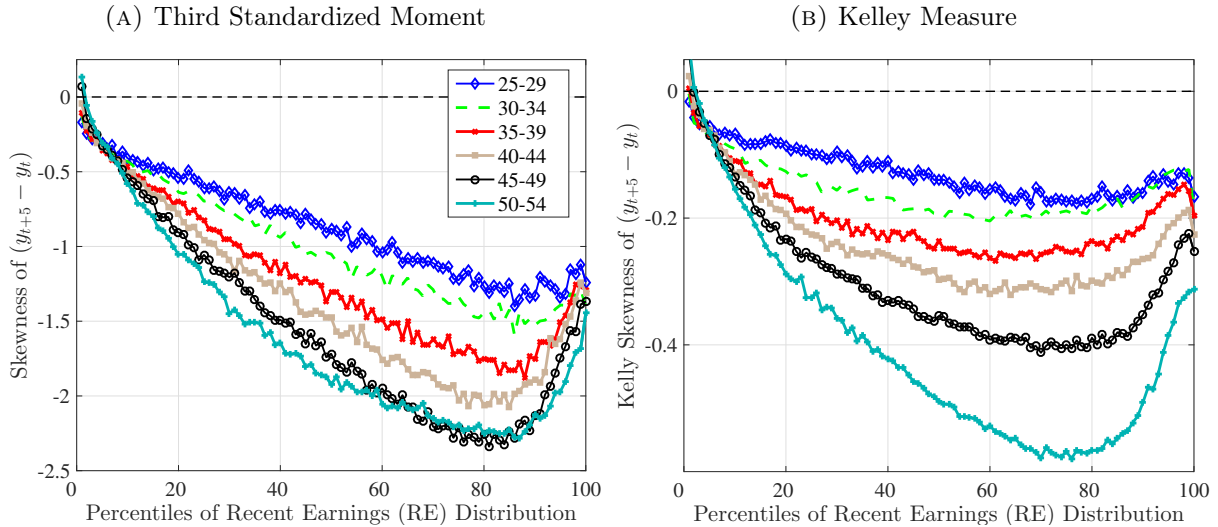
percentile of the RE distribution experience an increase from 0.71 in the youngest age group up to 0.92 in the oldest group (50–54), and those in the top 1% experience a larger increase from 1.05 in the first age group up to 1.31 in the oldest. Therefore, we conclude that the lower 95 percentiles and the top 5 percentiles display patterns with age and recent earnings that are the opposite of each other. The same theme will emerge again in our analysis of higher-order moments.

3.2 Third Moment: Skewness (or Asymmetry)

Lognormality implies that the skewness of (log) earnings growth is zero. Figure 4a plots the skewness, measured here as the third standardized moment of five-year earnings growth.[10] Two points should be observed. First, earnings changes are negatively 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 he has to fall and the less room he has left to move up. Curiously, and as was the case with the standard deviation, the life-cycle pattern in skewness becomes weaker (skewness changes less with age) at the very top of the earnings distribution.

---

[10]More precisely, for random variable X, with mean μ and standard deviation σ, the third standardized moment is $\mathbb{E}\left[(X-\mu)^3\right]/\sigma^3$.

Another measure of asymmetry is provided by Kelley (1947) measure of skewness:

$$\mathcal{S}_{\mathcal{K}} = \frac{(P90 - P50) - (P50 - P10)}{P90 - P10}. \tag{1}$$

Basically, $\mathcal{S}_{\mathcal{K}}$ measures the relative fractions of the overall dispersion (P90–P10) accounted for by the upper and lower tails. An appealing feature of Kelley's skewness is that it is bounded between -1 and $+1$, and a particular value has a simple interpretation. To see this, rearrange (1) to get

$$\frac{P90 - P50}{P90 - P10} = 0.5 + \frac{\mathcal{S}_{\mathcal{K}}}{2}.$$

Thus, $\mathcal{S}_{\mathcal{K}} < 0$ implies that the lower tail (P50-P10) is longer than the upper tail (P90-P50), indicating negative skewness. Another property of Kelley's measure is that it is robust to extremes (above the 90th or below the 10th percentile of the shock distribution). Instead, it captures the shift in the weight distribution in the middling section of a distribution, whereas the third moment also puts a large weight on the relative lengths of each tail. (We examine the tails in more detail in the next subsection.) Figure 4b plots Kelley's skewness, which reveals essentially the same pattern as with the third moment. Furthermore, the magnitude of skewness is substantial. For example, a Kelley measure of –0.40 (for 45–49 year old workers at the 80th percentile of the RE distribution) implies that the P90-P50 accounts for 30% of P90-P10, whereas P50-P10 accounts for the remaining 70%, far removed from the 50-50 distribution implied by a lognormal distribution.

A natural follow up question is whether skewness becomes more negative over the life cycle because of a compression of the upper tail (fewer opportunities to move up) or because of an expansion in the lower tail (increasing risk of income declines). This can be answered by looking at the *levels* of the P90-P50 and P50-P10 separately over the life cycle, which is plotted in Figure A.15 in Appendix B.3 to save space here. The figure shows that, from ages 25 to 50, the upper tail of the distribution compresses strongly, whereas after age 50 the bottom end starts opening up (i.e., chances of a large drop increase). As before, top earners are an exception to this pattern: the upper tail does not compress with age, but the bottom end opens up monotonically with age.

FIGURE 5 – Kurtosis of Five-Year Earnings Growth

(A) Fourth Standardized Moment

(B) Crow-Siddiqui Measure
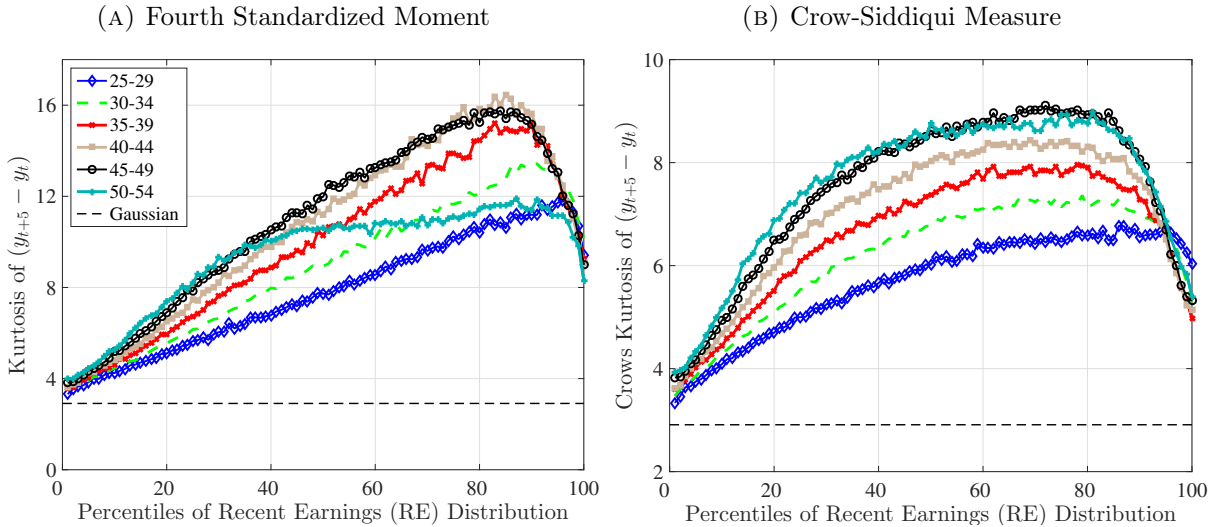
3.3 Fourth Moment: Kurtosis (Peakedness and Tailedness)

It is useful to begin by discussing what kurtosis measures. A useful interpretation has been suggested by Moors (1986), who described excess kurtosis as the tendency of a probability distribution to stay away from $\mu \pm \sigma$. This is consistent with how a distribution with excess kurtosis often looks like: 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. One familiar measure of kurtosis is the fourth standardized moment of the data: $\mathbb{E}((X - \mu)^4)/\sigma^4$. For the same reasons given for dispersion and skewness above, a nonparametric quantile-based measure is also desirable. One useful statistic is the Crow-Siddiqui measure (Crow and Siddiqui (1967)):

$$\kappa_{\text{C-S}} = \frac{P97.5 - P2.5}{P75 - P25},$$

which is equal to 2.91 for a Gaussian distribution.

With these definitions in mind, notice in Figure 5 that both measures of kurtosis increase monotonically with recent earnings up to the 80th to 90th percentiles for all age groups. That is, high earners experience *even smaller* changes of either sign, with few experiencing very large changes. Second, kurtosis increases with age for all RE levels

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

	$\text{Prob}(|y_{t+1}^i - y_t^i| < x)$		
x :	Data*	$\mathcal{N}(0, 0.51)$	Ratio
0.05	0.31	0.08	3.88
0.10	0.49	0.15	3.27
0.20	0.67	0.30	2.23
0.50	0.83	0.67	1.24
1.00	0.93	0.95	0.98

Notes: *The empirical distribution used in this calculation is for 1997-98, the same as in Figure 1.
†The intervals are defined as follows: "Center" refers to the area inside the first intersection between the two densities in Figure 1: $[-0.144, 0.194]$. "Tails" refer to the areas outside the intersection point at the tails: $(-\infty, -1.278] \cup [1.121, \infty)$. "Shoulders" refer to the remaining areas of the densities.

FIGURE 6 – Double-Pareto Tails of the U.S. Annual Earnings Growth Distribution
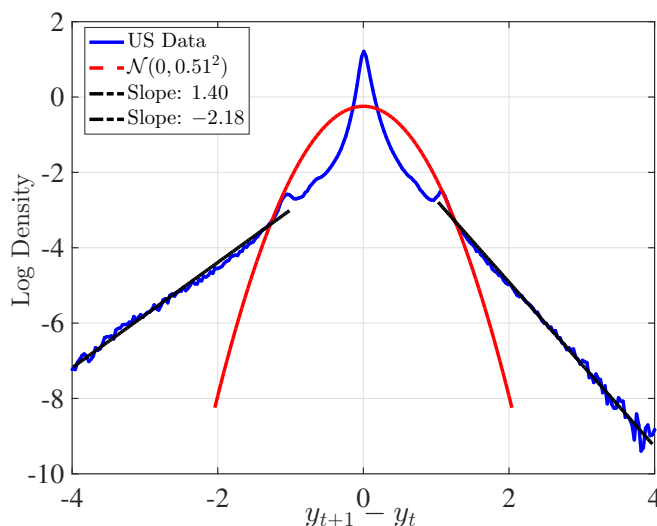
except the top 5%. Furthermore, the peak level of kurtosis ranges from a low of 12 for the youngest group all the way up to 16 for the middle-age group (40–54). The Crow-Siddiqui measure also shows very high kurtosis values, indicating that the fourth moment result is not driven by outliers (in the top and bottom 2.5% of the distribution, which is ignored by the Crow-Siddiqui measure).

To provide a familiar interpretation of kurtosis values, we calculate measures of concentration. Table I reports the fraction of individuals experiencing a log absolute earnings change of less than a threshold $x = 0.05, 0.10$, and so on, under alternative assumptions about the data-generating process. If data were drawn from a Gaussian density with a

standard deviation of 0.51 (the unconditional value for the whole sample), only 8% of individuals would experience an earnings change of less than 5%. The true fraction in the data is 31%. Similarly, the Gaussian density predicts a fraction of 15% when the threshold is 0.10, whereas the true fraction is 49%, and so on. We can also compute the probability of an extreme event (not reported in the table), which we define as drawing a shock larger than 150 log points (an almost fivefold increase, or an 80% drop). This happens approximately once in a lifetime in the data—or a 2.5% annual chance—whereas this probability is 6.3 times smaller under a normal distribution.

This high likelihood of extreme events motivates us to take a closer look at the tails of the earnings growth distribution. Figure 6 plots the empirical log density of annual earnings growth versus the Gaussian density.[11] The lognormal density is an exact quadratic, whereas the data display a more complex pattern. Two points are worth noting. One, the data distribution has much thicker *and* longer tails compared with a normal distribution, and the tails decline almost linearly, implying a Pareto distribution at both ends, with significant weight at extremes.[12] Two, the tails are asymmetric, with the left tail declining much more slowly than the right, contributing to the negative third moment documented above. In fact, fitting linear regression 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. It is worth stressing that while the Pareto tail in the earnings levels distribution is well known—indeed, going back to Pareto (1897)—here the two Pareto tails emerge in the earnings *growth* distribution. To our knowledge, the present paper is the first to document this empirical fact.

Overall, these findings show that earnings changes in the U.S. data exhibit important deviations from lognormality, and the extent of these deviations vary both over the life cycle and with the income level of individuals. In Section 7, we will investigate the implications of these features for a variety of economic questions in the context of a consumption-savings model.

3.4   Job-Stayers and Job-Switchers

Economists have documented important differences between the earnings changes that occur during an employment relationship and those that occur when workers switch jobs (see, e.g., Topel and Ward (1992), Low *et al.* (2010) and Bagger *et al.* (2014)).

---

[11]This is the same as the left panel of Figure 1 but with the y-axis now in logs.

[12]A double-Pareto distribution is one in which both tails are Pareto with possibly different tail indices.

While this literature has focused on the first moment (i.e., the average wage change), here, in the same spirit, we examine how the higher-order moments of earnings growth vary between job-stayers and job-switchers.

The SSA data set contains a unique employer identification number (EIN) for each job that a worker holds in a given year, which makes this analysis feasible. At the same time, the annual frequency of the data, together with the fact that some workers hold multiple jobs concurrently, poses a challenge for a precise identification of job-stayers and job-switchers. We have explored several plausible definitions and found qualitatively similar results. Here, we describe one reasonable classification. A worker is said to be a "job-stayer" between years t and $t+1$ if a given EIN provides the largest share of his annual earnings (out of all his EINs in that year) in years $t-1$ through $t+2$, and if the same EIN provides 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.[13] The probability of staying with the same employer increases with recent earnings and age (Figure A.19).

Figure 7 plots the quantile-based 2nd to 4th moments of earnings growth for stayers and switchers. A couple of remarks are in order. First, 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. Second, and more important, job-stayers (i) face a dispersion of earnings growth that is much smaller than that for job-switchers (about one-third for median-income workers); (ii) face shocks that have zero or slightly positive skewness as opposed to job-switchers, who face shocks that are very negatively skewed; and (iii) experience shocks with higher kurtosis than that for job-switchers, which is especially true for low RE workers. Figure A.20 plots the analogous graphs for annual earnings growth, and Figure A.21 plots the 2nd to 4th centralized moments, all confirming these conclusions.

These large and systematic differences in the higher-order moments between stayers and switchers raise the question of whether existing models of worker flows across jobs are consistent with these patterns. In ongoing work, Hubmer (2016) and Kapon *et al.* (2016) explore the ability of various frameworks (such as an Aiyagari (1994) model augmented

---

[13]Because the self employed (i.e., sole proprietors in our data set) do not have employers, we cannot check these criteria for those workers. We have experimented with treating them as switchers (the idea being that some of these individuals have less well defined employment boundaries) or as stayers and found that it made little difference. The results in this section follow the former approach: that is, a worker is classified as a switcher if more than 10% of his total annual earnings come from self employment in t or $t+1$.

FIGURE 7 – Higher-Order Moments of Earnings Growth: Stayers vs Switchers

(A) P90-P10

(B) Kelley's Skewness

(C) Crow-Siddiqui Kurtosis
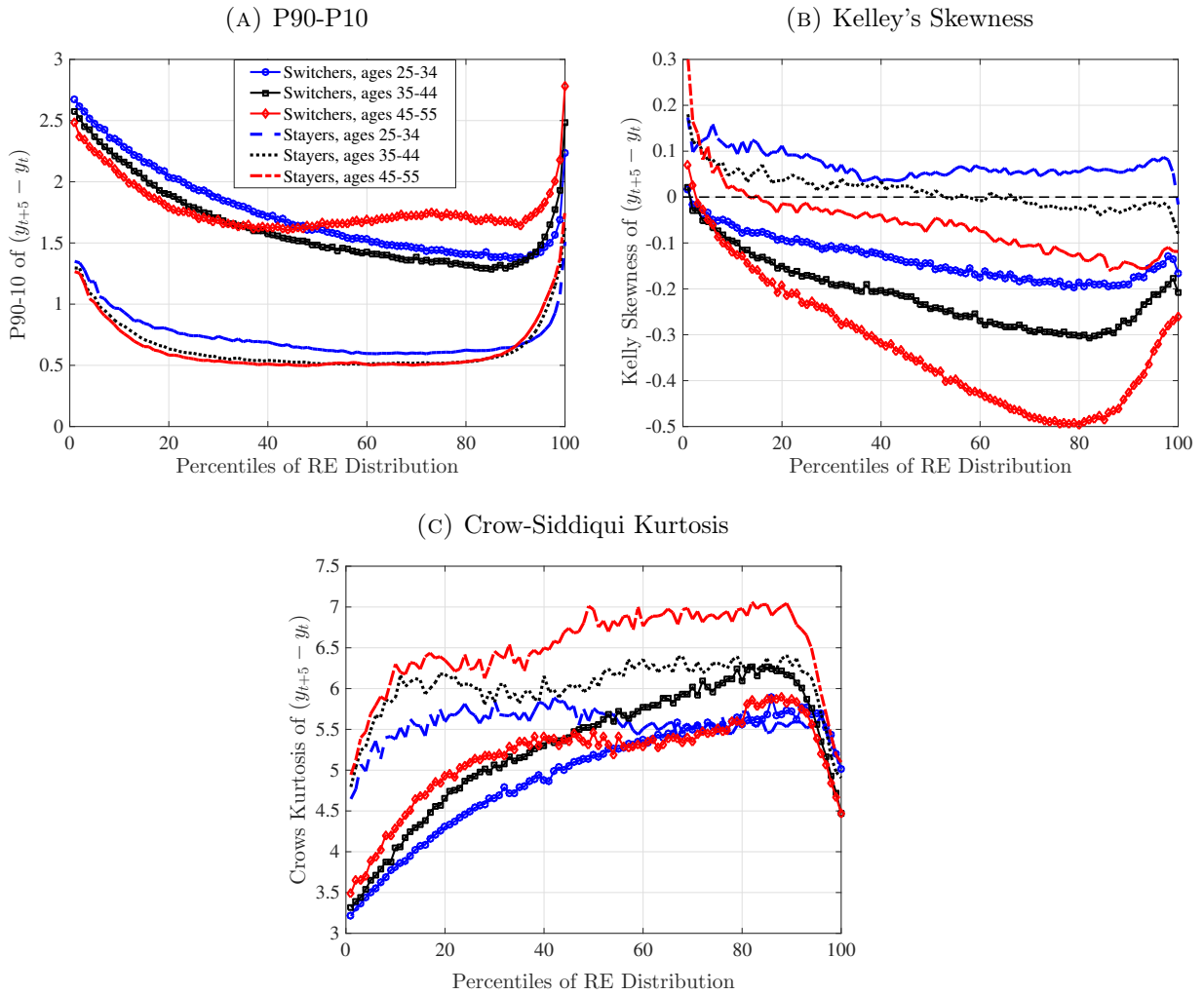
with search frictions or versions of the seminal Postel-Vinay and Robin (2002) model)
to generate these patterns.

4   Dynamics of Earnings

Another key dimension of life-cycle earnings risk is the persistence of earnings changes.
Typically, this persistence is modeled as an AR(1) process or a low-order ARMA process
(typically, ARMA(1,1)), and the persistence parameter is pinned down by the rate of
decline of autocovariances with the lag order. The AR(1) structure, for example, pre-
dicts a geometric decline and the rate of decline is directly given by the mean reversion

17

parameter. While this approach might be appropriate in survey data with small sample sizes, it imposes restrictions on the data that might be too strong, such as the uniformity of mean reversion for positive and negative shocks, for large and small shocks, and so on. Here, the substantial sample size allows us to characterize persistence without making parametric assumptions.

To reduce the number of graphs to a manageable level, we aggregate individuals across demographic groups. First, 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." Then, for each group, we rank and group individuals based on their average earnings from $t-5$ to $t-1$, and then within each group, we rank and group again by the size of the earnings change between $t-1$ and t. Hence, all individuals within a given group (obtained by crossing the two conditions) have the same average earnings up to time $t-1$ and experience the same earnings "shock" from $t-1$ to t. For each such group of individuals, we then compute the average change in log earnings from t to $t+k$, for $k = 1, 2, 3, 5, 10.$[14]

4.1 Impulse Response Functions Conditional on Recent Earnings

Figure 8a plots the "shock," $y_t^i - y_{t-1}^i$, on the x-axis and the fraction of each shock that has mean-reverted, $y_{t+k}^i - y_t^i$, on the y-axis for the median RE group. 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.

Several remarks are in order. First, the degree of mean reversion varies with the size of the shock, with much stronger mean reversion in $t+1$ for large shocks and smaller reversion for smaller shocks. Furthermore, even at the 10-year horizon, a nonnegligible fraction of the shocks' effect is still present, indicating a very persistent component to these shocks. Second, negative earnings changes tend to be less persistent than positive earnings changes for median-income workers. For example, a worker whose earnings rise by 100 log points between $t-1$ and t loses about 50% of this gain in the following 10

---

[14]Specifically, we construct 21 groups based on their RE percentiles: 1–5, 6–10, 11–15, . . . , 86–90, 91–95, 96–99, 100. Then, we construct 20 equally divided groups based on the percentiles of the shock, $y_t^i - y_{t-1}^i$: percentiles 1–5, 6–10, 11–15, . . . , 91–95, 96–100. Therefore, for every year t, we have 2 age groups, 21 RE groups, and 20 groups for shock size (for a total of 840 groups). As before, we construct these groups separately for each t and assign workers based on these averages. Then, for workers in each group, we compute the average of log k-year earnings growth, $\mathbb{E}\left[y_{t+k}^i - y_t^i | \overline{Y}_{t-1}^i, y_t^i - y_{t-1}^i\right]$, for $k = 1, 2, 3, 5, 10$.

years. Almost all of this mean reversion happens after one year, implying that whatever mean reversion there is happens very quickly. Turning to earnings losses: a worker whose earnings fall by 100 log points recovers one-third of that loss by $t + 1$ and recovers more than two-thirds of the total within 10 years. Moreover, unlike with positive shocks, the recovery (hence mean reversion) is more gradual in response to negative shocks.

Second, the degree of mean reversion varies with the magnitude of earnings shocks. This is evident in Figure 8a, where small shocks (i.e., those less than 10 log points in absolute value) look very persistent, whereas there is substantial mean reversion following larger earnings changes. A univariate autoregressive process with a single persistence parameter will fail to capture this behavior. In the next section, we will allow for multiple AR(1) processes to accommodate this variation in persistence by shock size.

The next two panels (Figures 8b and 8c) plot the analogous impulse response functions for low-income (i.e., $\overline{Y}_{t-1} \in P6 - P10$) and high-income ($P91 - P95$) workers. For low-income individuals, negative shocks mean-revert much more quickly, whereas positive shocks are more persistent than before. The opposite is true for high-income individuals.

We now extend the results to the entire distribution of recent earnings. To accommodate more income groups, we focus on a fixed horizon, 10 years, and plot the cumulative mean reversion from t to $t + 10$ for the 6 RE groups in Figure 8d. Starting from the lowest RE group (those individuals in the bottom 5% of the recent earnings distribution), notice that negative shocks are transitory, with an almost 75% mean reversion rate at the 10-year horizon. But positive shocks 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, we have the opposite pattern: only 20 to 25% of negative shocks mean-revert at the 10-year horizon, whereas around 80% of positive shocks mean-revert at the same horizon. We refer to this shape as the "butterfly pattern."

5   Life-Cycle Income Growth

We conclude our analysis of life-cycle earnings with the first moment—average earnings growth—and examine how it varies over the life cycle and, for reasons that will become clear in a moment, across groups of individuals that differ in their *lifetime earnings* (and not recent earnings). For this analysis, we leverage the long panel dimension

FIGURE 8 – Impulse Responses, Prime-Age Workers

(A) Median-Income Workers

(B) Low-Income Workers

(C) High-Income Workers

(D) Butterfly Pattern

Notes: Median-, low-, and high-income in panels A, B, and C refer to workers with, respectively, $\overline{Y}_{t-1}$ in $(P45 - P55)$, $(P6 - P10)$, and $(P91 - P95)$.
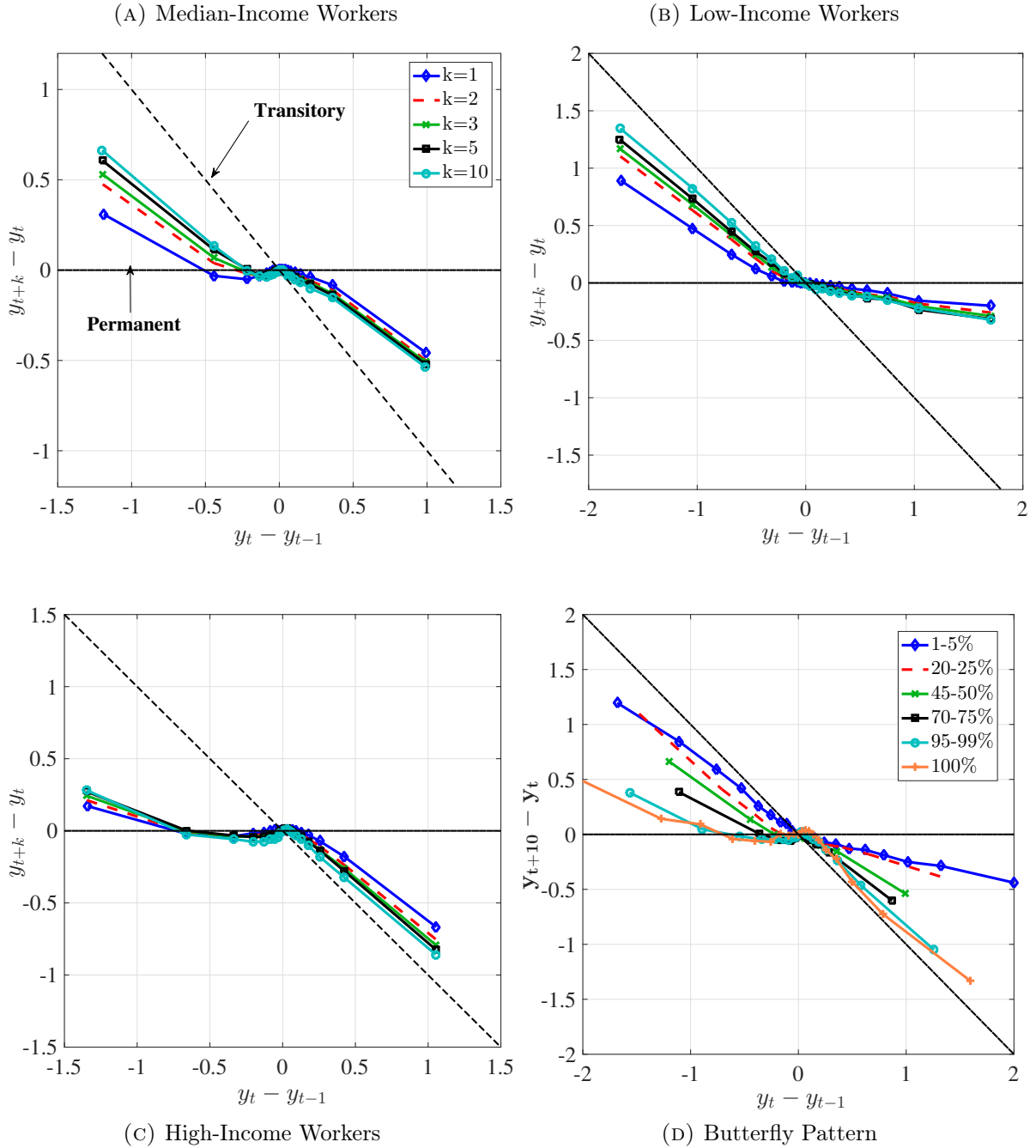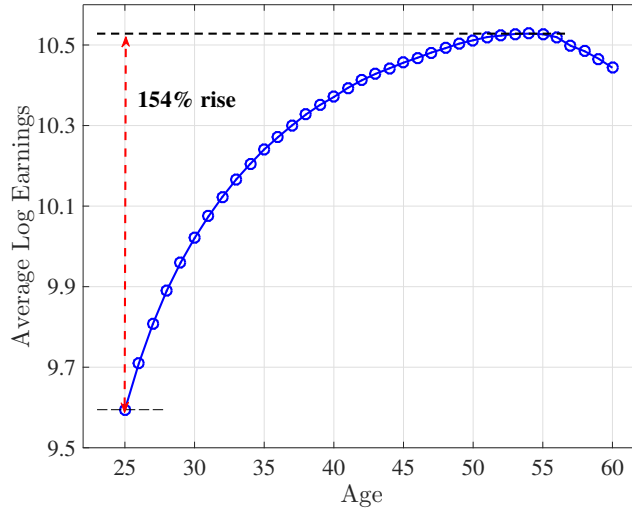
FIGURE 9 – Life-Cycle Profile of Average Log Earnings

of our data set and follow individuals over their entire working life. To this end, we employ a balanced panel that selects all individuals (i) who are between ages 25 and 28 in 1981 and are still alive by 2013 and (ii) have annual earnings above a minimum threshold, denoted $Y_{\min,t}$, for at least 15 of these years. The first condition ensures that we have 33-year earnings histories between ages 25 and 60 for each individual (which might include years with zero earnings). These two conditions are the same as in the base sample, but now extended to cover the entire life cycle.

To provide a benchmark, we estimate the average life-cycle profile of log earnings by regressing log individual earnings on a full set of age and (year-of-birth) cohort dummies.[15] The estimated age dummies are plotted as circles in Figure 9 and represent the average life-cycle profile of log earnings. It has the usual hump-shaped pattern that peaks around age 50.[16]

It is well understood that the steepness of the life-cycle income profile (i.e., the cumulative growth in income from ages 25 to 55) matters greatly for many economic questions, because it is a strong determinant of borrowing and saving motives. In Figure 9, this cumulative growth is about 93 log points, which is about 154%. Notice that feeding this life-cycle profile into a calibrated model will imply that the median individual in the

---

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

[16]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 $h = (\text{age} - 24)/10$.

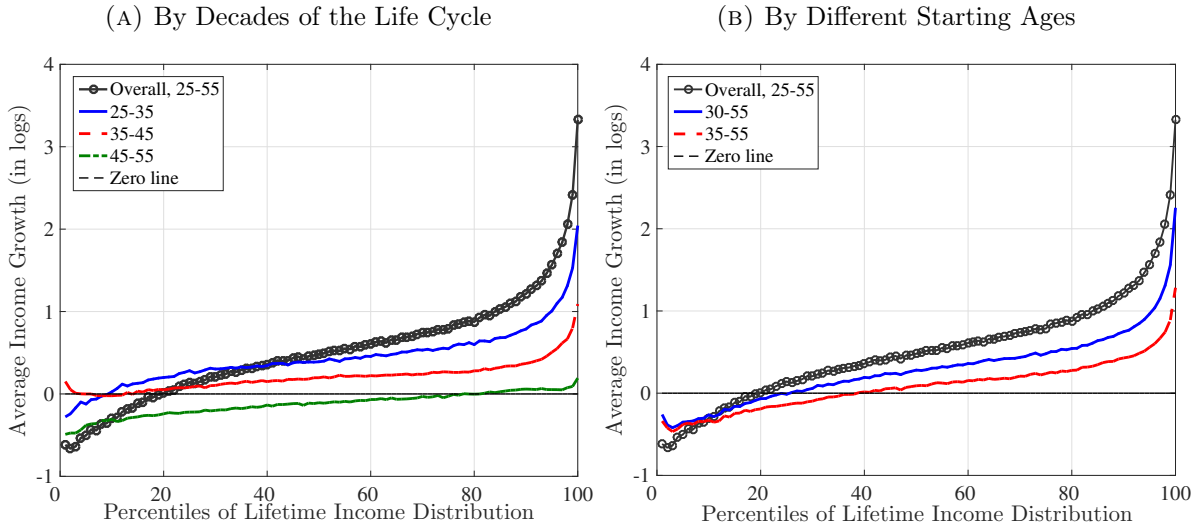FIGURE 10 – Life-Cycle Earnings Growth Rates, by Lifetime Earnings Group
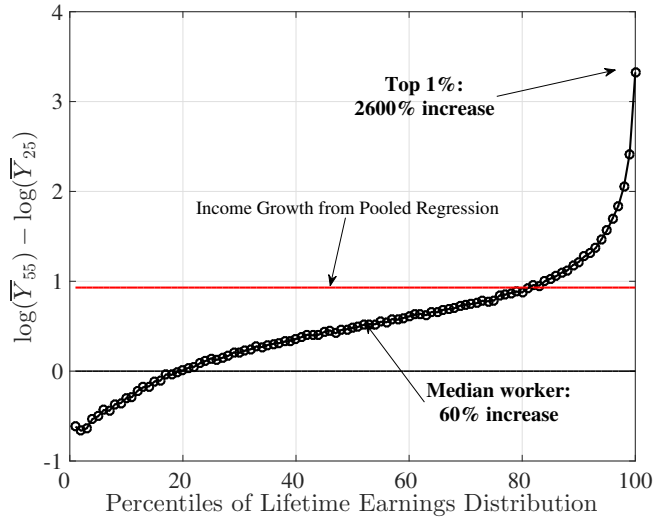
simulated sample experiences (on average) a rise of this magnitude from ages 25 to 55. Is this implication consistent with what we see in the data? In other words, if we rank male workers by their lifetime earnings, does the median worker in the United States experience an earnings growth of approximately 154%?

To answer this question, we begin by computing lifetime earnings for each individual as follows. We rank individuals based on their lifetime earnings, computed by summing their earnings from ages 25 through 60. Earnings observations lower than $Y_{\min,t}$ are set to this threshold. For individuals in a given lifetime earnings (hereafter, LE) percentile group, denoted LEj, $j = 1, 2, ..., 99, 100$, we compute growth in average earnings between any two ages h_1 and h_2 as $\log(\overline{Y}_{h_2,j}) - \log(\overline{Y}_{h_1,j})$, where $\overline{Y}_{h,j} \equiv \mathbb{E}(Y_h^i | i \in \text{LE}j)$ and Y_h^i for a given individual may be zero.

Figure 10 plots the results for $h_1 = 25$ and $h_2 = 55$. There are several takeaways. First, individuals in the median LE group experience a growth rate of 60%, slightly more than one-third of what was predicted by the profile in Figure 9. Second, we have to look above LE80 to find an average growth rate of 154%. However, earnings growth is very high for high-LE individuals, with those in the 95th percentile experiencing a growth rate of 380% and those in the 99th percentile experiencing a massive growth rate of 2600%. Although some of this variation could be expected because individuals with high earnings growth are more likely to have high lifetime earnings, these magnitudes are too large to be accounted for by that channel, as we show below.

22

FIGURE 11 – Log Earnings Growth over Subperiods of Life Cycle

(A) By Decades of the Life Cycle   (B) By Different Starting Ages

Earnings Growth by Decades. How is earnings growth distributed over different decades of the life cycle? Figure 11a answers this question by separately plotting growth rates from ages 25 to 35, 35 to 45, and 45 to 55. The bulk of earnings growth happens during the first decade. In fact, for the median LE group, average cumulative earnings growth from ages 35 to 55 is only 9%. Second, with the exception of those in the top 20% by LE, all groups experience negative growth from ages 45 to 55. How do the results change if we consider a later starting age? Figure 11b plots earnings growth starting at age 30 (solid blue line) and 35 (dashed red line). As just mentioned, from ages 35 to 55, the average cumulative growth rate is very low for all workers below L70. Top earners still do well though, experiencing a rise of 226 log points (or 849%) from ages 30 to 55 and a rise of 128 log points (or 260%) from ages 35 to 55. Those at the bottom of the LE distribution display the opposite pattern: average earnings drops by 34 log points (or 29%) from ages 35 to 55.

6   Estimating Stochastic Processes for Earnings

With few exceptions, the earnings dynamics literature relies on the (often implicit) assumption that earnings shocks can be approximated reasonably well with a lognormal distribution. This assumption, combined with a linear time-series model to capture the accumulation of such shocks, allowed researchers to focus their estimation to match

the covariance matrix (of log earnings), abstracting from higher-order moments.[17] This approach faces two difficulties. First, the empirical evidence documented in this paper shows large deviations from lognormality, in which case the third and fourth moments contain valuable information. Second, the covariance matrix estimation makes it difficult to select among alternative models of earnings risk, because it is difficult to judge the relative importance—from an economic standpoint—of the covariances that a given model matches well and those that it does not.

With these considerations in mind, we follow a different approach that relies on matching the kinds of moments presented above.[18] We believe that economists can more easily judge whether or not each one of these moments is relevant for the economic questions they have in hand. We conducted a battery of diagnostic tests on each set of moments to get a better idea about what elementary components (e.g., persistent shocks, normal mixtures, heterogeneous profiles, etc.) have the potential to generate those features. We conducted similar diagnostic analyses on the other cross-sectional moments (standard deviation, skewness, and kurtosis) as well as on impulse response moments. The rich persistence pattern necessitated mixing multiple AR(1) processes, whereas the variation in the second to fourth moments over the life cycle and with earnings levels seemed impossible to match without introducing explicit dependence of mixing probabilities on age and earnings. Therefore, we make these features part of our benchmark.

6.1   A Flexible Stochastic Process

The most general econometric process we estimate has the following features: (i) a heterogeneous income profiles (hereafter, HIP) component of linear form; (ii) a mixture of two AR(1) processes,[19] denoted by z_1 and z_2, where each component receives a new innovation in a given year with probability $p_j \in [0, 1]$ for $j = 1, 2$; (iii) a nonemployment

---

[17]To be clear, GMM or minimum distance estimation that is used to match such moments does not require the assumption of lognormality for consistency. But abstracting away from moments higher than covariances is a reflection of the belief that higher-order moments do not contain independent information, which relies on the lognormality assumption.

[18]This approach is in the spirit of Browning *et al.* (2010), Altonji *et al.* (2013), and Guvenen and Smith (2014).

[19]Geweke and Keane (2007) study how regression models can be smoothly mixed, and our modeling approach shares some similarities with their framework.

shock ν; and (iv) an i.i.d. shock ε. Here is the full specification:

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

$$z_{1,t}^i = \rho_1 z_{1,t-1}^i + \eta_{1t}^i \tag{3}$$

$$z_{2,t}^i = \rho_2 z_{2,t-1}^i + \eta_{2t}^i, \tag{4}$$

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

Innovations to AR(1)'s: $\quad \eta_{jt} \sim \begin{cases} -p_{jt}^i \mu_j & \text{with pr. } 1 - p_{jt}^i \\ \mathcal{N}((1 - p_{jt}^i)\mu_j, \sigma_j^i) & \text{with pr. } p_{jt}^i \end{cases} \tag{5}$

Indiv.-specific innov. var. : $\quad \log\left(\sigma_j^i\right) \sim \mathcal{N}(\log \overline{\sigma}_j - \dfrac{\widetilde{\sigma}_j^2}{2}, \widetilde{\sigma}_j) \tag{6}$

Initial value of AR(1) process: $\quad z_{j0}^i \sim \mathcal{N}(0, \sigma_j^i \sigma_{j0}). \tag{7}$

Nonemployment shock: $\quad \nu_t^i \sim \begin{cases} 0 & \text{with pr. } 1 - p_{\nu t}^i \\ \min\{1, \text{Expon.}(\lambda)\} & \text{with pr. } p_{\nu t}^i \end{cases} \tag{8}$

The term $g(t)$ in (2) is a quadratic polynomial in age and captures the life-cycle component of earnings that is common to all individuals. The random vector (α^i, β^i) is drawn from a multivariate normal distribution with zero mean and a covariance matrix to be estimated.[20] Each AR(1) component, z_1 and z_2, receives a shock drawn from a mixture of a Gaussian distribution and a mass point chosen so that the innovations have zero mean. Furthermore, we allow the variance of each innovation to be individual-specific, with a lognormal distribution with mean $\overline{\sigma}_j$ and a standard deviation proportional to $\widetilde{\sigma}_j$.[21] 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 (8).[22] 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

---

[20]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).

[21]Chamberlain and Hirano (1999) and Browning *et al.* (2010) find strong evidence of such heterogeneity. Our preliminary analysis also confirmed this prediction, which is why we allow for this feature. Furthermore, such heterogeneity in innovation variances can also give rise to excess kurtosis in the cross-sectional distribution of income changes, as we found in Section 3, even when each individual receives Gaussian innovations. Thus, modeling this feature is important for understanding the determinants of kurtosis we documented earlier.

[22]This specification allows the standard deviation of the initial condition to be proportional to the individual-specific standard deviation σ_j^i.

loss of generality, we impose $\rho_1 < \rho_2$. Finally, with probability $p_{\nu t}^i$, an individual receives a nonemployment shock ($\nu_t > 0$) whose duration has an exponential distribution with mean δ, truncated at 1, the interpretation being of full year nonemployment (i.e., zero annual income).

The age and income dependence of moments is captured by allowing the mixture probabilities to depend on age and the persistent component of earnings ($z_1 + z_2$):[23]

$$
\begin{aligned}
p_{jt}^i &= \frac{\exp\left(\xi_{j,t-1}^i\right)}{1 + \exp\left(\xi_{j,t-1}^i\right)}, \\
\xi_{jt}^i &= a_j + b_j \times t + c_j \times \left(z_{1,t}^i + z_{2,t}^i\right) + d_j \times \left(z_{1,t}^i + z_{2,t}^i\right) \times t,
\end{aligned}
\tag{9}
$$

for $j = 1, 2$. The equation for $p_{\nu t}$ is the same as (9) but $\xi_{j,t-1}^i$ is replaced with ξ_{jt}^i. This completes the description of the stochastic process.

Estimation Procedure

We estimate the parameters of the stochastic process described by equations (2) to (9) using the method of simulated moments (MSM). The empirical targets are: (i) the standard deviation, skewness, and kurtosis of one-year and five-year earnings growth; and (ii) moments describing the impulse response functions; and (iii) moments from the life-cycle profile of average earnings. In addition, the within-cohort variance of log earnings levels is a key dimension of the data that has been extensively studied in previous research. For both completeness, and consistency with earlier work, we include these variances (summarized in Figure A.18) as a fourth set of moments.

As discussed earlier, the descriptive analysis relied on log earnings changes because of its familiarity and required us to drop individuals with low earnings—below $\overline{Y}_{\min}$—in year t or $t + k$. In the estimation, we target the same set of moments but this time computed using arc-percent changes, which allows us to include individuals who enter and exit market work. This way, the estimated income process will capture not only income shocks in the intensive margin but also the nature and persistence of shocks that move workers in and out of the labor market. The counterparts of all the figures in Sections 3 and 4 using arc-percent change are reported in Appendix B.2. They exhibit the same qualitative patterns as the log difference versions presented above, but their

---

[23]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.

magnitudes capture the extensive margin too, allowing us to streamline the estimation process. Appendix C contains the details about which moments are targeted, how they are constructed, and the MSM objective function, and includes plots that show the targeted cross-sectional and impulse response moments.

6.2 Results: Estimates of Stochastic Processes

Table II reports the estimation results. We begin with an overview of what the specification in each column aims to capture and how these specifications fit various moments. To save space here, we plot the fit of each specification to the moments we target in Appendix C.3 (Figures A.25 and A.26).

Column (1) is the full specification and our benchmark process. It has the best fit of all the specifications we have estimated. Each of the subsequent columns, except the very last, eliminates only one feature from the benchmark so as to facilitate comparison. For example, column (2) shuts down one of the AR(1) processes ($z_2 \equiv 0$) but keeps all other features of the benchmark. The advantage of this specification is that it reduces the number of state variables required for a dynamic programming model from 2 to 1. As seen in the bottom panel, this simplification comes at some cost—a worse fit to the cross-sectional moments and impulse responses. However, relatively speaking this deterioration is not nearly as bad as when we eliminate other features later on. So column (2) can be viewed as a simplified benchmark. Column (3) eliminates the HIP component ($\beta^i \equiv 0$) from the benchmark process. This leads to a worse fit as expected, but again not dramatically so.

In Column (4), we eliminate the nonemployment shock ($\nu_t \equiv 0$) from the benchmark, which leads to a large jump in the objective value, indicating a significant deterioration in the fit of the model, especially in cross-sectional and impulse response moments. This shows that extensive margin shocks are an essential component of an empirically plausible income process. In the next column (5), we eliminate the dependence of mixture probabilities on past earnings, which shows an even worse deterioration in fit. This result is perhaps not very surprising, given the large systematic variation with recent earnings that we found in the previous sections.

Column (6) does the analogous experiment but this time shuts down age dependence in mixture probabilities. Interestingly, the fit is not nearly as bad and is actually comparable to column (2). This shows that the life-cycle changes in earnings dynamics we

27

TABLE II – Estimates of Stochastic Process Parameters

Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Benchmark process						**Gaussian process**
HIP order	1	1	0	1	1	1	0
AR(1)	2	1	2	2	2	2	1
Nonemployment shocks	yes	yes	yes	no	yes	yes	no
Heterogenous variances?	yes	yes	yes	yes	yes	yes	no
Probability age/inc.	yes/yes	yes/yes	yes/yes	yes/yes	yes/no	no/yes	no/no
Parameters							
σ_α	0.362	0.477	0.491	0.149	0.500	0.419	0.372
$\sigma_\beta \times 10$	0.200	0.208	—	0.045	0.164	0.203	—
$\text{corr}_{\alpha\beta}$	0.603	0.307	—	--0.323	0.762	0.410	—
ρ_1	0.556	0.934	0.539	0.592	0.005	0.590	1.00
ρ_2	0.956	—	0.996	0.996	0.775	0.951	—
$\bar{\sigma}_1$	0.290	0.417	0.325	1.852	0.579	0.461	0.156
$\bar{\sigma}_2$	0.447	—	0.365	0.767	2.048	0.051	—
σ_1^i	0.721	0.764	0.800	0.137	0.716	0.565	—
σ_2^i	0.600	—	0.476	0.309	0.052	0.429	—
$\sigma_{1,0}$	0.260	0.426	0.187	0.191	0.196	0.218	0.354
$\sigma_{2,0}$	0.765	—	0.509	0.195	0.071	0.294	—
σ_ε	0.118	0.204	0.071	0.214	0.011	0.133	0.494
Objective value	**20.78**	**26.78**	**23.32**	**37.34**	**45.37**	**25.43**	**72.28**
Decomposition:							
(i) Cross-section	1.49	2.18	2.18	2.87	8.06	3.09	36.75
(ii) Impulse resp.	1.84	3.96	2.27	7.76	10.81	2.25	14.17
(iii) Inc. growth	0.86	0.98	0.89	1.41	1.20	0.99	0.84
(iv) Inequality	0.13	0.05	0.09	1.90	0.53	0.14	0.48

Notes: This table reports a subset of the estimated parameters, which are discussed in the main text. The rest of the parameters are reported in Table A.3.

documented can be mimicked with dependence on recent earnings, since earnings rise with age until late in life. Therefore, allowing mixture probabilities to depend on recent earnings is more critical than their dependence on age. This finding is also important from a modeling perspective because, being age independent, the process in column (6) can be used to calibrate an infinite horizon model (or a perpetual youth version) and still capture a lot of the variation in earnings dynamics documented here.

Finally, column (7) goes back to a bare bones—persistent (AR1) plus transitory model—which is the most common income process used in calibrated macro models. Because the resulting process is Gaussian, it clearly has no chance of fitting the cross-sectional higher-order moments nor of generating the asymmetry in the impulse responses. Therefore, the estimation does what it can and does a fairly good job of matching the life-cycle growth moments and the life-cycle profile of the variance of log earnings at the expense of completely missing the cross-sectional and impulse response moments. In the next section, we shall examine the implications of the benchmark process and the Gaussian process for consumption-savings decisions to understand how they differ. In Appendix C.3, we report five more versions of the estimation in which we examine the effects of heterogeneous variances, allowing for a random walk component, allowing age and income dependence in mixture probabilities, and so on.

6.3 Parameter Estimates

We now discuss the estimates from the benchmark process in some detail, and for the remaining specifications, we point out the differences from the benchmark. The estimates for the life-cycle component ($\sigma_\alpha = 0.36$, $\sigma_\beta = 2\%$) are reasonable and quite consistent with earlier estimates in the literature, even though we target a very different set of moments here and use SSA data as opposed to the PSID data used in most previous work.[24] Turning to the dynamics of income, the two AR(1) components are vastly different from each other: the first displays only a modest annual persistence of $\rho_1 = 0.556$, whereas the second one is quite persistent: $\rho_2 = 0.956$. Both processes also receive large innovations, with mean standard deviation sizes of $\bar{\sigma}_1 = 0.29$ and $\bar{\sigma}_2 = 0.47$, and much higher variances for some individuals given the large individual-specific distribution around these mean values (given by $\tilde{\sigma}_1 = 0.72$ and $\tilde{\sigma}_2 = 0.60$). However, it would be premature to conclude from these numbers (higher persistence and larger variance) that the second AR(1) process is more important than the first; for that, we also need to know how often each process gets hit by a new innovation, which is given by the mixture probabilities, p_1 and p_2, discussed in a moment.

Turning to the nonemployment shock, the estimated mean of the exponential distribution for the duration is $\lambda = 0.07$, which implies that more than 95% of these shocks are full-year nonemployment shocks that reduce earnings below the participation threshold ($Y_{\min}$). To the extent that these shocks are realized with nonnegligible probability, they

---

[24]Cf., Haider (2001), Heathcote *et al.* (2010), and Guvenen and Smith (2014).

TABLE III – Mixture Probabilities of Persistent Components, Benchmark Process

| | Age groups | | | RE (Percentile) groups | | | | |
	25–34	35–49	45–59	1	10	50	90	100
p_{z_1} $(\rho_z = 0.56)$	0.101	0.186	0.292	0.062	0.110	0.197	0.254	0.477
p_{z_2} $(\rho_x = 0.96)$	0.264	0.201	0.187	0.696	0.476	0.185	0.100	0.044
p_ν (nonemp.)	0.088	0.065	0.059	0.374	0.208	0.050	0.020	0.014
Pr (any shock)	0.351	0.369	0.458	0.738	0.563	0.364	0.341	0.515

can have important effects on the earnings dynamics of an individual. This brings up the question of how often each of these three shocks is realized, which we turn to now.

The coefficients of the probability function in (9) are hard to interpret on their own, so instead in Table III we report the probabilities for various age and RE percentile groups for workers who satisfy the conditions of the base sample.[25] As seen here, the overall likelihood of receiving a shock to either component is fairly low. The probability of receiving the first AR(1) innovation, η_1, is only 10% at young ages and gradually climbs to 29% for workers after age 45. The opposite pattern is seen for the second and more persistent AR(1) innovation, η_2, which has a 27% chance of being received in early years, which falls to 19% after age 45. Table III also reports the probability of drawing the nonemployment shock, which falls modestly from 8.8% per year at young ages to 5.9% after age 45. Overall, considering all three shocks simultaneously (last row), we see that young workers (ages 25-35) receive a shock with about 35% probability (i.e., once every three years), and this probability rises to about 46% for those aged 45-60. Although the probability rises slightly, the composition of shocks changes from (hard-to-insure) persistent η_2 shocks and large nonemployment shocks to less persistent η_1 shocks. Finally, in addition to these large and persistent shocks that hit every other year or so, each individual receives smaller i.i.d. income shocks (ε_t) every year, with 11.8% standard deviation.

The last five columns provide a different perspective by reporting the probability of each shock by RE percentiles. Here we see much larger differences. The less persistent AR(1) shock is very unlikely to hit low-income individuals but is much more likely for top 1% individuals. The opposite happens for the second AR(1) shock, which hits the lowest-income group with 70% probability per year but has only a 4.4% chance for the top 1%. Similarly, individuals in the bottom 10 percentiles receive the full year nonemployment

---

[25]Figure A.27 plots 3D graphs of p_j $(j = 1, 2, \nu)$ and Table A.3 reports the estimated coefficients.

FIGURE 12 – Model Fit: Nontargeted Statistics

(A) Future Nonemployment by RE Quintiles

(B) Log Density of Earnings
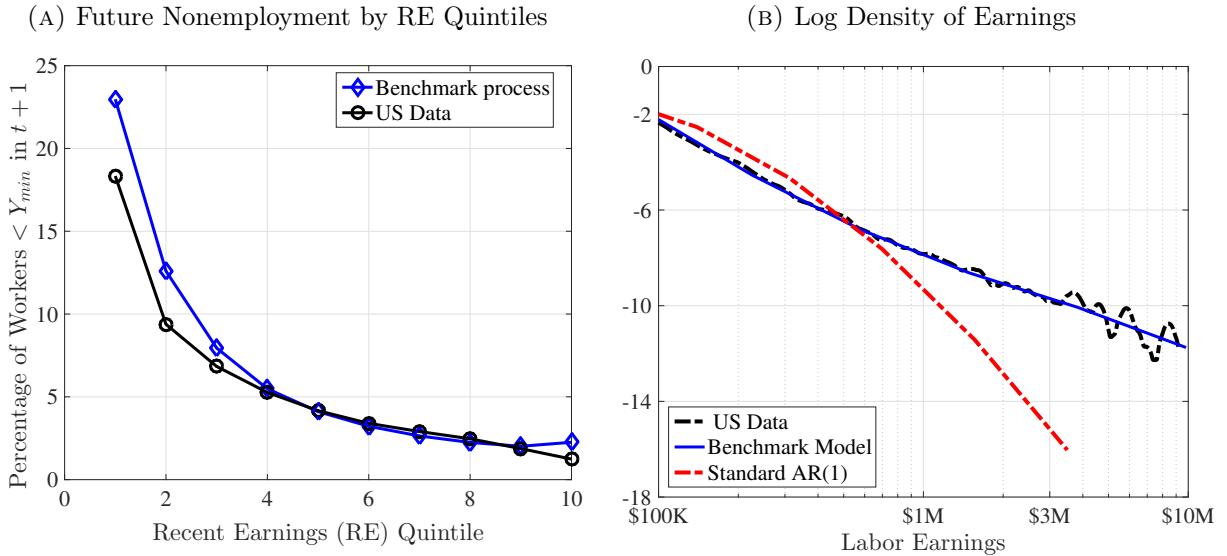
shock with a high probability, ranging from 21% to 37%, but individuals above the median receive this shock with only a 5% chance (twice in a lifetime), and top earners receive it with only 1.4% probability per year. Overall, the probability of receiving *at least* one of the three shocks in a given year (i.e., $p_1 + p_2 + p_\nu$) is U-shaped in recent earnings, ranging from 74% at the low end to 34% for individuals in the 90th percentile and up again to 52% for the top 1% individuals.

Another important point to note is that the dependence of shock probabilities on z_t induces nonlinearities in the earnings process. A key implication is that the autocorrelation of shocks is not precisely captured by the coefficients ρ_1 and ρ_2, and the nonemployment shock is autocorrelated even though v_t is drawn in an i.i.d. manner. This is because a shock in the current period affects the value of z_t, thereby changing the probabilities of shocks tomorrow. For example, for low-income individuals, the correlation of ν_t and ν_{t+1} is 0.60.[26] This persistence has important implications for consumption-savings behavior, which we analyze in the next section.

Because there is so much heterogeneity in these nonemployment shock probabilities and these probabilities were not targeted directly in the estimation, it is worth comparing the implied nonemployment rates directly to data. Figure 12a plots the fraction of

---

[26]Incidentally, this asymmetric persistence coming from nonemployment shocks contributes to the overall asymmetry we find in the impulse response analysis. Arellano *et al.* (2016) document similar asymmetry in persistence using data from the PSID.

individuals who are full-year nonemployed in year t conditional on recent earnings in the data and in the model. The fit is quite good, with a small discrepancy at the lowest end.[27] This large heterogeneity in nonemployment rates is behind the estimated model's ability to generate the substantially different patterns documented for high- and low-income individuals in the previous sections.

Finally, Figure 12b shows another moment of the data that was not targeted in the estimation but is important for the analysis of wealth concentration in the next section. It shows the log density of earnings (levels—not changes) in the data, along with the counterparts from the benchmark estimation and the Gaussian process. As seen here, the benchmark process generates a Pareto tail that matches the data almost perfectly whereas the Gaussian process (as expected) generates a lognormal tail that underestimates the number of individuals with very high earnings. We will return to this point in the next section.

7 Life-Cycle Consumption-Savings Model

We now investigate some key implications of the rich earnings process estimated in the previous section for life-cycle consumption-savings behavior. Consider a continuum of individuals that participate in the labor market for the first T_W years of their life, retire at that age, and die at age $T > T_W$. Individuals have standard CRRA preferences over consumption and hence supply labor inelastically.

Individual earnings follow the stochastic process described by equations (2) to (9). Although this process has many parameters, all dynamics are captured through only two state variables ($z_{1,t}^i$ and $z_{2,t}^i$), which makes it relatively straightforward to embed it into a dynamic programming problem. Let $\Upsilon^i \equiv (\alpha^i, \beta^i, \sigma_1^i, \sigma_2^i)$ denote the parameters capturing fixed (ex ante) heterogeneity. We construct a four-dimensional discrete grid over Υ^i where each grid point corresponds to a worker type k. Therefore, some aspects of an individual's problem depends on his (discrete) ex ante type k, whereas others—including the dynamics of income—are drawn from continuous distributions that are also fully individual specific.

Individuals can borrow or save using a risk-free asset with gross return R, where

---

[27]We have also generated similar graphs from the Survey of Income and Program Participation (SIPP) data set for various unemployment outcomes (duration, incidence, etc.) as a function of past two-year average income, which shows the same strong decline in unemployment outcomes with past earnings.

borrowing is subject to an age-dependent worker-type-specific limit, denoted by $\underline{A}_t^k$. Then the budget constraint is given by

$$c_t^i + a_{t+1}^i = a_t^i R + Y_t^{\text{disp},i}, \qquad \forall t, \tag{10}$$

$$Y_t^{\text{disp},i} = \max\left\{\underline{Y}, \widetilde{Y}_t^i\right\}^{1-\tau}, \qquad t = 1, \dots, T_W, \tag{11}$$

$$Y_t^{\text{disp},i} = \left(\widetilde{Y}_R^k\right)^{1-\tau}, \qquad t = T_W + 1, \dots, T, \tag{12}$$

$$a_{t+1}^i \geq \underline{A}_t^k, \qquad \forall t, \tag{13}$$

where c_t^i and a_t^i denote consumption and asset holdings, respectively, and $Y_t^{\text{disp},i}$ is disposable income, which differs from gross income, $\widetilde{Y}_t^i$, in two ways. First, the government provides social insurance by guaranteeing a minimum level of income, $\underline{Y}$, to all individuals—more on this in a moment. Second, the government imposes a progressive tax on after-transfer income. Following Benabou (2000) and Heathcote *et al.* (2014), we take this tax function to have a power form, with exponent $1 - \tau$. This specification has been found to fit the U.S. data quite well. Finally, retirees receive pension income, $\widetilde{Y}_R^k$, specified to mimic the U.S. Social Security Administration's OASDI system.[28] Appendix D contains further details. The dynamic programming problem of an individual is

$$V_t^i\left(a_t^i, z_{1,t}^i, z_{2,t}^i; \Upsilon^k\right) = \frac{(c_t^i)^{1-\gamma}}{1-\gamma} + \beta \mathbb{E}_t\left[V_{t+1}^i\left(a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i; \Upsilon^k\right)\right]$$

$$\text{s.t.}$$

$$\text{equations (10)-(13), and } V_{T+1}^i \equiv 0,$$

where β is the subjective discount factor, and γ governs risk aversion. This problem can be solved using standard numerical techniques; see computational Appendix D.

7.1 Calibration

Households enter the labor market at age 25, retire at 60 ($T_W = 36$), and die at 85 ($T = 60$). The coefficient of relative risk aversion γ is set to 2 as a conservative choice. The discount factor, β, is calibrated to replicate the wealth-to-income ratio of 4 as reported in the Flow of Funds Z1 tables.[29] R is set to 3%, and $\underline{A}_t^k$ is set equal to

---

[28]The dependence of pension on worker type k rather than individual income is to avoid introducing another state variable.

[29]The total wealth-to-income ratio is defined to be total asset holdings in the population relative to the sum of total before-tax labor income and capital income.

each type-k's average annual earnings over the life cycle. The parameter determining progressivity, τ, is set to 0.185 following Heathcote *et al.* (2014).

As for $\underline{Y}$, Hubbard *et al.* (1995) estimate the total value of various government programs (food stamps, AFDC, housing subsidies, etc.) for a female-headed family with two children and no outside earnings and assets. They obtain a value of about $7,000 in 1984 dollars—or about $12,800 in 2010 dollars. Using the OECD equivalence scale for one adult plus two children, this comes to $6,400 per adult person. However, as we shall see, the effects of higher-order income risk are mitigated as $\underline{Y}$ is increased, so for our baseline calibration we take a higher—and thus conservative—value of $10,000 per year. (To put this number in context, the 10th percentile of the RE distribution in Section 3 is about $9,000 per year.) For comparison, we shall also report and discuss results when $\underline{Y}$ is reduced to $2,000.

The key element in the calibration is the choice of income process parameters. Here we consider a few cases. First, we simulate the model using the process and parameters taken from column (1) of Table II.[30] We call this calibration the *benchmark model* in reference to the *benchmark process* on which it builds. Second, however, recall that the benchmark process features a HIP component, and the implications of HIP for consumption-savings decisions have been analyzed in the previous literature (e.g., Guvenen (2007) and Guvenen and Smith (2014)). Therefore, to focus on the novel aspects of this income process coming from higher-order income risk, we consider a second calibration in which we eliminate HIP by setting $\beta^i \equiv 0$ (hereafter NOHIP model) but keep the remaining parameters from column (1). In what follows, this second case will be the main focus of our analysis (although we will also report the results for the benchmark model for completeness and discuss them briefly).

Third, to provide a comparison, we also simulate the same life-cycle model with the Gaussian process in column (7) of Table II with one modification: because we want to compare this model with the NOHIP model, we recalibrate ρ such that the resulting process generates the *same* rise in the variance of log disposable income over the life cycle as the NOHIP model. See panel A of Figure 13. This yields $\rho = 0.985$. We refer to this process as the "Gaussian model" in reference to the Gaussian process on which it builds.

---

[30]The only exception is the variance of i.i.d. shocks, which we set to zero to simplify computation (i.e., $\sigma_\varepsilon \equiv 0$).

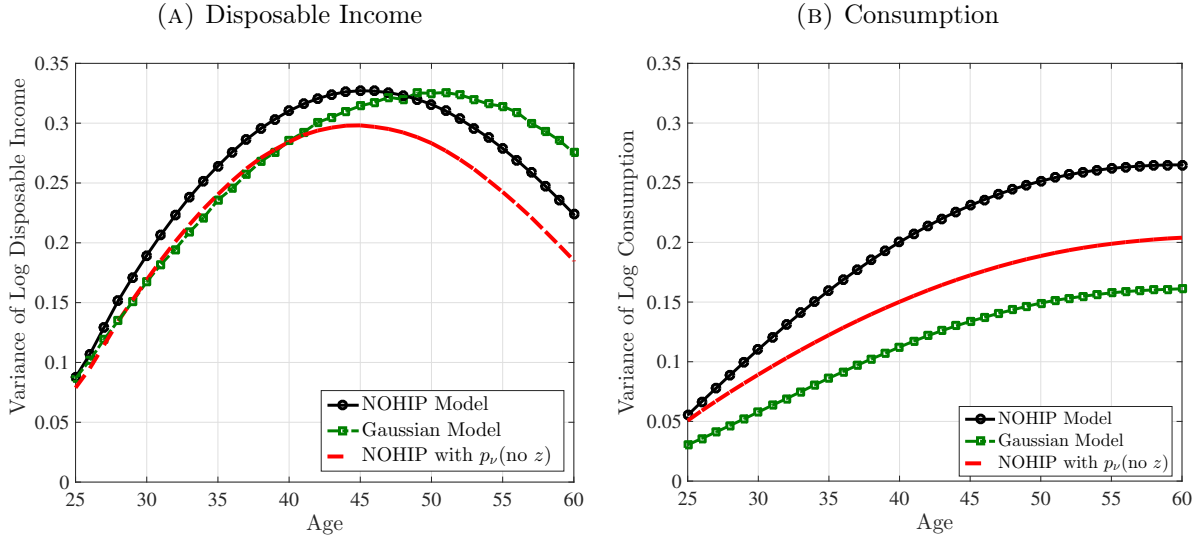FIGURE 13 – Within-Cohort Variance of Log Disposable Income and Log Consumption

(A) Disposable Income

(B) Consumption

TABLE IV – Welfare Costs of Idiosyncratic Shocks

Model	Benchmark	NOHIP	p_ν (no z)	Gaussian
	(1)	(2)	(3)	(4)
$\underline{Y} = \$10,000$	25.3%	18.1%	14.4%	11.3%
$\underline{Y} = \$2,000$	40.1%	34.5%	24.8%	12.5%

7.2 Results

We now study four sets of implications of the benchmark process with higher-order moments: (i) the welfare costs of idiosyncratic fluctuations, (ii) wealth inequality, (iii) measuring the insurability of income shocks, and (iv) the higher-order moments of consumption growth.

I. Welfare Costs of Idiosyncratic Income Shocks

The first exercise aims to quantify the amount of income risk implied by the estimated income processes. Specifically, we ask: *What fraction of consumption at every date and state would an individual in the benchmark model be willing to give up to live in a hypothetical world with no income uncertainty?* This hypothetical world is defined as one with the income process set to its average value at each age. This experiment is conducted behind the veil of ignorance, that is, before the individual learns his type k. Because this is a well-understood experiment, we relegate the equations to Appendix D.

In the benchmark model (Table IV, first column, first row), the individual is willing to give up 25.3% of consumption at every date and state, which indicates very large welfare costs of idiosyncratic income fluctuations. The last column reports the corresponding number for the Gaussian model, which is 11.3%, or less than half of what we find for the benchmark model. Second, these welfare costs turn out to be fairly sensitive to the value of $\underline{Y}$. Reducing $\underline{Y}$ to $2,000$ (second row) raises the welfare cost to 40.1% (from 25.3%) in the benchmark model. Interestingly, the welfare cost rises only by a little in the Gaussian model, to 12.5% from 11.3%. The reason for this asymmetry is that in the benchmark model, individuals occasionally receive large negative shocks (including full-year nonemployment shocks) and therefore benefit from the insurance provided by the income floor, whereas this is less of a case in the Gaussian model, where tail shocks are much less likely. Therefore, weakening the safety net is more costly when the true income process is as in the benchmark model.

Third, the benchmark process includes a HIP component in addition to higher-order risk. To isolate from the effects of the former and focus on the more novel latter component, in column (2) we shut down the HIP component ($\beta^i \equiv 0$). The welfare costs in the NOHIP model are slightly smaller—falling from 25.3% to 18.1% when $\underline{Y} = \$10,000$ and from 40.1% to 34.5% when $\underline{Y} = \$2,000$—but are still substantial. Therefore, higher-order risk on its own generates larger welfare costs than a Gaussian process. The cost is nearly three times higher: 34.5% (NOHIP) vs. 12.5% (Gaussian).

Fourth, to distinguish between the risk posed by the persistent AR(1) shocks and nonemployment shocks, we conduct the following experiment. We eliminate the dependence of the shock probability function, p_ν, on z_1 and z_2, only allowing it to vary by age (and rescale its level so that the average nonemployment rate by age remains the same as in the NOHIP case). This change has two effects: one, because z_1 and z_2 are quite persistent, eliminating the dependence on them makes nonemployment shocks completely transitory; and two, nonemployment ceases to vary with the income level and hits all workers of a given age with the same probability. As seen in column (3), this leads to a fall in the welfare costs from 18.1% to 14.4% (and from 34.5% to 24.8% with a low safety net).[31]

---

[31]Interestingly, the welfare costs in this case are quite close to the case if we were to eliminate nonemployment shocks completely, suggesting that most of the welfare costs of nonemployment shocks are due to their persistence and concentration among already low-income individuals.

Fifth, and finally, the average welfare costs reported in Table IV mask significant heterogeneity across ex ante types (not reported in the table). For example, ranking all individual types k by the welfare costs of the idiosyncratic risk they face in the benchmark model, we find that the 90th percentile of this distribution is close to 26%, whereas the 10th percentile is a mere 2.4%. As can perhaps be anticipated, the highest welfare costs are associated with types who have high values of (σ_1, σ_2) and, perhaps less obvious, with those who have high values of (α^i, β^i). That is, high-income individuals suffer more from idiosyncratic risk. This is because they are less protected by the social safety net, the magnitude of which is too small to make a difference in their income fluctuations.

Welfare Costs of Higher-Order Income Risk. To better understand how higher-order income risk leads to large welfare costs, consider the well-known thought experiment in which a decision maker chooses between (i) a static gamble that changes consumption, c, by a random proportion $(1 + \tilde{\delta})$ and (ii) a fixed payment π, called the risk premium, to avoid the gamble. An expected utility maximizer solves

$$U(c \times (1 - \pi)) = \mathbb{E}\left[U(c \times (1 + \tilde{\delta}))\right]. \tag{14}$$

To provide a decomposition for how higher-order moments affect the risk premium, the following equation is helpful. Taking a first-order Taylor-series approximation to the left-hand side and a *fourth-order* approximation to the right-hand side, we get

$$\pi^* \approx \underbrace{\frac{\gamma}{2} \times \sigma_\delta^2}_{\text{variance aversion}} - \underbrace{\left[\frac{(\gamma + 1)\gamma}{6} \times \sigma_\delta^3\right] \times s_\delta}_{\text{(negative) skewness aversion}} + \underbrace{\left[\frac{(\gamma + 2)(\gamma + 1)\gamma}{24} \times \sigma_\delta^4\right] \times \kappa_\delta}_{\text{kurtosis aversion}}, \tag{15}$$

where s_δ and κ_δ are the skewness and kurtosis coefficients, respectively.[32] This expression is a generalization of the well-known formula that includes only the first term. Among other things, it shows that γ is perhaps more appropriately called the coefficient of "variance aversion," because with higher-order moments, the individual also displays aversion to negative skewness and to excess kurtosis. Moreover, the coefficient in front of skewness is a quadratic in γ and a cubic in σ_δ. Therefore, for a fixed level of skewness, a higher curvature or dispersion can greatly amplify the risk premium. For kurtosis, the situation is even more extreme: the coefficient is a cubic in γ and a quartic in dispersion. Therefore, income shocks that are negatively skewed or exhibit excess kurtosis (or both)

---

[32]See Appendix D.2 for derivation.

will generate a higher risk premium than a Gaussian shock with the same variance. This explains how the NOHIP (and benchmark) model generates a higher welfare cost even though it has the same variance as in the Gaussian model. Notice also that our choice of $\gamma = 2$ was deliberately conservative; larger values would imply even larger welfare costs from higher-order income risk.

II. Wealth Inequality

An important use of the consumption-savings model is for studying wealth inequality (cf. Aiyagari (1994) and Huggett (1996)). Some recent papers have pointed out that income changes that display negative skewness (Lise (2013)) or excess kurtosis (Castañeda et al. (2003)) can generate higher wealth inequality. To complement this work, here we explore the potential of higher-order risk whose magnitude is estimated from the U.S. data to generate a realistic wealth distribution.

As seen in Table V, the benchmark model exhibits the highest wealth inequality, whereas the Gaussian model exhibits the lowest. In particular, going from the Gaussian model to the benchmark, the Gini coefficient rises from 0.58 to 0.69, and the top 10% share rises from 37.9% to 47.4%. These differences are not just due to HIP: the NOHIP model in column 3 has a Gini coefficient of 0.66 compared with 0.58 for the Gaussian model, and the top 0.1% share of aggregate wealth is 2.2% in the NOHIP model relative to 1.1% for the Gaussian model. Therefore, higher-order income risk provides nonnegligible amplification of wealth concentration, especially at the top, where the share is doubled for the top 0.1%.[33]

That said, the improvement is not nearly large enough to generate the massive concentration of wealth at the very top observed in the U.S. data, where the top 1% holds 37% of aggregate wealth. The corresponding figure is 7% in the Gaussian model and 9.2% in the NOHIP model. Furthermore, in unreported simulations we found that although increasing risk aversion from 2 to 10 increases the gap in the Gini between the Gaussian model and the NOHIP model (generating a Gini in excess of 0.70 in the latter), it has little effect on the top-end concentration. Thus, although the benchmark model provides a step in the right direction, we conclude that empirically measured income

---

[33]This is due to the fact that the benchmark process generates a Pareto tail for the income distribution (at least over a wide range of top income levels) unlike the Gaussian process—recall Figure 12b. For example, the number of individuals with $3 million in wealth is ($e^{15/10} \approx$) about 150 times larger under the benchmark model than under the Gaussian one, and the gap increases monotonically for higher income levels. This thicker Pareto tail of the income distribution translates into a thicker tail for the wealth distribution.

TABLE V – Key Statistics of Wealth Distribution

	U.S. Data	Simulated Model		
		Benchmark	NOHIP	Gaussian
	(1)	(2)	(3)	(4)
Gini	0.85	0.69	0.66	0.58
Top 10%	50.8%	47.4%	41.6%	37.9%
Top 1%	37.0%	10.0%	9.2%	7.0%
Top 0.1%	14.8%	2.2%	2.2%	1.1%

risk—even with negatively skewed and leptokurtic income changes—cannot generate the thick right tail of the U.S. wealth distribution.[34] That said, higher-order risk as in the benchmark model (column 2) could be important for explaining wealth accumulation of the top 10% excluding the top 1%, which compares favorably with the U.S. data (unlike the Gaussian model—see Table V).

III. Measuring the Insurability of Income Shocks

A third set of implications we explore concerns the transmission of income shocks into consumption. This transmission rate has received attention in the recent literature because it is interpreted as a measure of partial insurance—that is, insurance above and beyond self-insurance (see, e.g., Blundell *et al.* (2008), Primiceri and van Rens (2009), and Kaplan and Violante (2010)).

Blundell *et al.* (2008) (hereafter BPP) propose using two simple moment conditions to estimate the insurance coefficients in response to permanent shocks and transitory shocks, respectively.[35] The insurance coefficients range from 0 to 1, where 0 means no partial insurance above self-insurance (or full transmission) and 1 means perfect insurance (or no transmission). Here, the experiment we consider is the following. Suppose we give a panel data set of income and consumption simulated from the benchmark model to an econometrician and ask her to estimate the insurance coefficients for permanent and transitory shocks using BPP's moments. What would the econometrician conclude about the extent of the insurability of income shocks?

---

[34]It seems that one needs other mechanisms, such as rate of return heterogeneity, to generate empirically plausible levels of wealth inequality at the top (see, e.g., Gabaix *et al.* (2015) and Benhabib and Bisin (2016)).

[35]The exact formulas for the BPP insurance coefficients are $\phi^\eta = 1 - \frac{\text{cov}(\Delta c_t^i, y_{t+1}^{\text{disp},i} - y_{t-2}^{\text{disp},i})}{\text{cov}(\Delta y_t^{\text{disp},i}, y_{t+1}^{\text{disp},i} - y_{t-2}^{\text{disp},i})}$ for permanent shocks and $\phi^\varepsilon = 1 - \frac{\text{cov}(\Delta c_t^i, \Delta y_{t+1}^{\text{disp},i})}{\text{cov}(\Delta y_t^{\text{disp},i}, \Delta y_{t+1}^{\text{disp},i})}$ for transitory shocks.

TABLE VI – Measures of Insurability of Shocks

Model:	BPP insurance coefficients					
	NOHIP			Gaussian		
Age:	30	40	50	30	40	50
Permanent	0.60	0.70	0.79	0.39	0.50	0.63
Transitory	0.84	0.86	0.87	0.97	0.95	0.95

Table VI reports the results for different ages. When the true data-generating process is the NOHIP model, the BPP procedure estimates that 60% of permanent shocks are insured and the remaining 40% is transmitted to consumption at age 30.[36] The corresponding insurance coefficient for the Gaussian process is lower—39% (so 61% transmitted). The same gap between the insurance coefficients in the two models—of approximately 20%—is stable across different ages. Taken at face value, this finding would indicate that persistent income shocks in the benchmark model are more insurable relative to the Gaussian model. This could happen if the NOHIP *process* features less persistent shocks (which is partly true, at least judging coarsely based on the values of ρ_1 and ρ_2) or if the NOHIP *life-cycle model* features more insurance opportunities relative to the Gaussian model (which is not true—they both have self-insurance only).

However, a second way economists have approached the question of the degree of insurance is by studying how within-cohort consumption inequality evolves over the life cycle. The idea is that if shocks are easily insurable—because either they are not persistent or the economy features rich smoothing opportunities—then consumption inequality should not rise much with age. To explore this, Figure 13b plots the cross-sectional variance of log consumption in the NOHIP and Gaussian models. In the NOHIP model, consumption inequality rises by about 22 log points from age 25 to age 60, which is about twice as large as the rise in the Gaussian model (of 13 log points). Therefore, this second way to look at the degree of partial insurance leads to the opposite conclusion: shocks in the NOHIP model are harder to insure, which causes consumption inequality to rise more than in the Gaussian model.[37] This latter evidence is also more consistent

---

[36]For brevity, we omit the statistics of the benchmark model from here on, which look very similar to the NOHIP model discussed here.

[37]Further analysis has shown that the key feature of the NOHIP model in generating the large rise in inequality is the nonlinearity and persistence of nonemployment shocks. Removing the dependence of p_ν on z as we did above (and which breaks both the autocorrelation and income dependence of nonemployment) leads to a much smaller rise in consumption inequality—though still higher than the Gaussian model—see Figure 13b.

TABLE VII – Cross-Sectional Moments of Consumption

Model	NOHIP			Gaussian		
Age	30	40	50	30	40	50
Variance ($\times 10^2$)	0.86	0.63	0.34	0.50	0.41	0.20
Kelley skewness	−0.33	−0.10	0.06	0.00	0.06	0.09
Crow-Siddiqui kurtosis	5.3	7.8	8.5	3.0	3.0	3.1

with the higher welfare costs of idiosyncratic shocks that we found for the NOHIP model relative to the Gaussian model above.

Taken together, we conclude from these two pieces of evidence that once we move beyond Gaussian shocks and linear models, extra care is needed to properly measure the extent of insurability of shocks. The nonlinear dynamics and higher-order moments generate interesting new patterns. For example, why does the transmission parameter indicate a low rate of transmission in the NOHIP model? An important reason is that the partial insurance coefficient measures the *average* response of consumption growth to income shocks. But it is plausible to expect that the consumption response varies by the size of the shock, by its sign, and by many of its other properties established in the previous sections. So, the average response coefficient provides an incomplete picture of the transmission of income shocks to consumption. Instead, we would like to know the properties of the entire distribution of consumption growth rates implied by a model,[38] including its higher-order moments. We turn to this next.

IV. Higher-Order Moments of Consumption Growth

Table VII reports the variance, Kelley skewness, and Crow-Siddiqui kurtosis of consumption growth in the NOHIP model and the Gaussian model. First, the variance of consumption growth declines with age in both models but is about 50% larger in the benchmark model as in the Gaussian one. Recall that both income processes have similar variances for income growth. So this suggests that consumption growth is much more volatile under the benchmark model.

Second, consumption growth is negatively skewed before age 50 in the benchmark model but has slight positive skewness in the Gaussian model. The minimum income floor induces slight positive skewness in both models—without it, the benchmark model

---

[38]This point is also emphasized contemporaneously in Arellano *et al.* (2016).

delivers even more negatively skewed consumption growth. Interestingly, consumption growth becomes *less* negatively skewed with age, even though income shocks become *more* negatively skewed, which seems to be due to precautionary wealth allowing better smoothing at older ages.

Finally, consumption growth has very high excess kurtosis, as measured by the robust Crow-Siddiqui measure. The Gaussian model delivers almost no excess kurtosis (the Gaussian distribution has a Crow-Siddiqui kurtosis of 2.9). (The third and fourth standardized moment measures of skewness and kurtosis—which we do not report here for brevity—confirm the same finding). In further analysis, we have also found that the excess kurtosis of consumption growth increases when risk aversion is increased.

To our knowledge, there are only a few papers that have considered deviations from lognormality of the consumption growth distribution. Among these, Brav *et al.* (2002) argued that accounting for the skewness of consumption growth distribution is critical for reconciling the observed equity premium with the data, and Constantinides and Ghosh (2016) provide evidence that household consumption growth is negatively skewed in the US data. More recently, Toda and Walsh (2015) found that the household consumption growth distribution displays a double Pareto distribution, which is consistent with the implications of the benchmark model studied here displaying excess kurtosis of consumption growth. However, neither paper studies the life cycle patterns of these moments shown in Table VII. An exciting future research avenue would be to confront these implications with high quality micro consumption data from which higher-order moments can be measured precisely.

8   Concluding Thoughts

Our analysis of the life-cycle earnings histories of millions of U.S. workers has reached two broad conclusions. First, the higher-order moments of individual earnings shocks display large deviations from lognormality. In particular, earnings shocks display strong negative skewness (what can be thought of as individual-level disaster shocks) and extremely high kurtosis. The high kurtosis implies that in a given year, most individuals experience very small earnings changes, few experience middling changes, and a small but nonnegligible number experience extremely large changes. The second conclusion is that these statistical properties of earnings changes vary substantially both over the life cycle and with the earnings level of individuals.

Although the lognormal framework is often adopted in the literature on the grounds of tractability, negative skewness and excess kurtosis are naturally generated by standard economic models of job search over the life cycle. For example, job ladder models in which workers do on-the-job search and move from job to job as they receive better offers and fall off the job ladder after nonemployment not only will generate negative skewness but also will imply that skewness becomes more negative with age. This is because as the worker climbs the job ladder, the probability of receiving a wage offer that is much higher than the current wage will be declining. At the same time, as the worker moves higher up the wage ladder, falling down to a flat nonemployment surface (or disability) implies that there is more room to fall. Furthermore, as the attractiveness of job offers declines with the current wage, more job offers will be rejected, and therefore the frequency of job-to-job transitions will also decline with age, implying that most wage changes will be small (within-job) changes. This increases the concentration of earnings changes near zero, which in turn raises the kurtosis of changes. That said, the magnitudes of variation over the life cycle and by earnings levels that we documented in these moments are so large that it is an open question whether existing models of job search can be consistent with these magnitudes, and, if not, what kinds of modifications should be undertaken to make them consistent. In ongoing work, Hubmer (2016) and Kapon et al. (2016) explore the ability of various frameworks to generate some of these patterns.

A limitation of the SSA data set used in this paper is the lack of information on hours and wages separately. Consequently, we are not able to distinguish the extent to which the patterns documented here are due to changes in hours versus changes in the wage rate. In ongoing work, Busch et al. (2016) use detailed panel data on French workers (with information on daily hours, days in an employment spell, daily wages, employer ID, and so on), which allows them to make progress on these questions.

We have also estimated nonparametric impulse response functions of earnings shocks and found significant asymmetries: positive shocks to high-earnings individuals are quite transitory, whereas negative shocks are very persistent; the opposite is true for low-earnings individuals. Although these statistical properties are typically ignored in quantitative analyses of life-cycle models (with the standard persistent-transitory lognormal income processes), they are fully consistent with search-theoretic models of careers over the life cycle. After establishing these empirical facts nonparametrically, we estimated what we think is the simplest earnings process that is broadly consistent with these salient features of the data.

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.

Finally, the consumption-savings model we studied here showed that higher-order income risk can have quite different implications for various distributional phenomena compared with the commonly used Gaussian processes. We believe that this analysis only scratched the surface of potential implications of higher-order income risk. We hope the empirical findings about the nature of earnings dynamics documented in this paper feed back into economic research and policy analyses.

References

ABOWD, J. M. and CARD, D. (1989). On the covariance structure of earnings and hours changes. *Econometrica*, **57** (2), 411–45.

AIYAGARI, R. (1994). Uninsured idiosyncratic risk and aggregate saving. *Quarterly Journal of Economics*, **109** (3), 659–684.

ALTONJI, J., SMITH, A. A. and VIDANGOS, I. (2013). Modeling earnings dynamics. *Econometrica*, **81** (4), 1395–1454.

ARELLANO, M., BLUNDELL, R. and BONHOMME, S. (2016). *Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework*. mimeo, University College London.

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.

BENABOU, R. (2000). Unequal societies: Income distribution and the social contract. *American Economic Review*, **90** (1), 96–129.

BENHABIB, J. and BISIN, A. (2016). *Skewed Wealth Distributions: Theory and Empirics.* Working Paper 21924, National Bureau of Economic Research.

BLUNDELL, R., PISTAFERRI, L. and PRESTON, I. (2008). Consumption inequality and partial insurance. *American Economic Review*, **98** (5), 1887–1921.

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

BRAV, A., CONSTANTINIDES, G. and GECZY, C. (2002). Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, **110** (4), 793–824.

BROWNING, M., EJRNAES, M. and ALVAREZ, J. (2010). Modelling income processes with lots of heterogeneity. *Review of Economic Studies*, **77**, 1353–1381.

BUSCH, C., FIALHO, P. and GUVENEN, F. (2016). *Life-Cycle Wage and Hours Dynamics: Higher Order Moments.* Tech. rep., University of Minnesota.

CASTAÑEDA, A., DÍAZ-GIMÉNEZ, J. and RÍOS-RULL, J.-V. (2003). Accounting for the U.S. earnings and wealth inequality. *The Journal of Political Economy*, **111** (4), 818–857.

CHAMBERLAIN, G. and HIRANO, K. (1999). Predictive distributions based on longitudinal earnings data. *Annales d'Économie et de Statistique*, **55-56**, 211–242.

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.

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. (2016). *The Implications of Richer Earnings Dynamics for Consumption, Wealth, and Welfare.* Working paper 21917, National Bureau of Economic Research.

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

GABAIX, X., LASRY, J.-M., LIONS, P.-L. and MOLL, B. (2015). *The Dynamics of Inequality.* Tech. rep., NYU Finance.

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.

— and — (2007). Smoothly mixing regressions. *Journal of Econometrics*, **138**, 252–290.

GOLOSOV, M., TROSHKIN, M. and TSYVINSKI, A. (2016). Redistribution and social insurance. *American Economic Review*, **106** (2), 359–386.

GUVENEN, F. (2007). Learning Your Earning: Are Labor Income Shocks Really Very Persistent? *American Economic Review*, **97** (3), 687–712.

—, 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.

HEATHCOTE, J., STORESLETTEN, K. and VIOLANTE, G. L. (2010). The macroeconomic implications of rising wage inequality in the united states. *Journal of Political Economy*, **118** (4), 681–722.

—, — and — (2014). Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review*.

HUBBARD, R. G., SKINNER, J. and ZELDES, S. P. (1995). Precautionary saving and social insurance. *The Journal of Political Economy*, **103** (2), 360–399.

HUBMER, J. (2016). *The Job Ladder and its Implications for Earnings Risk*. Tech. rep., Yale University.

HUGGETT, M. (1996). Wealth distribution in life-cycle economies. *Journal of Monetary Economics*, **38** (3), 469–494.

—, VENTURA, G. and YARON, A. (2011). Sources of lifetime inequality. *American Economic Review*, **101** (7), 2923–54.

KAPLAN, G., MOLL, B. and VIOLANTE, G. L. (2016). *Monetary Policy According to HANK*. Working paper, Princeton University.

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

KAPON, S., KARAHAN, F., OZKAN, S. and SONG, J. (2016). *Understanding Lifetime Earnings Inequality: A Search-Theoretic Approach*. Working paper.

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.

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

LILLARD, L. A. and WILLIS, R. J. (1978). Dynamic aspects of earnings mobility. *Econometrica*, **46** (5), 985–1012.

LISE, J. (2013). On the job search and precautionary savings. *Review of Economic Studies*, **80**, 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.

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.

NEAL, D. and ROSEN, S. (2000). Theories of labor earnings. In A. B. Atkinson and F. Bourguignon (eds.), *Handbook of Income Distribution*, North-Holland.

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).

POSTEL-VINAY, F. and ROBIN, J.-M. (2002). Equilibrium Wage Dispersion with Worker and Employer Heterogeneity. *Econometrica*, **70** (6), 2295–2350.

PRIMICERI, G. E. and VAN RENS, T. (2009). Heterogeneous life-cycle profiles, income risk and consumption inequality. *Journal of Monetary Economics*, **56** (1), 20 – 39, carnegie-Rochester Conference Series on Public Policy: The Causes and Consequences of Rising Economic Inequality April 25-26, 2008.

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.

TODA, A. A. and WALSH, K. (2015). The double power law in consumption and implications for testing euler equations,. *Journal of Political Economy*, **123** (5), 1177–1200.

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.

WANG, C., WANG, N. and YANG, J. (2016). Optimal consumption and savings with stochastic income and recursive utility. *Journal of Economic Theory*.

WANG, N. (2007). An equilibrium model of wealth distribution. *Journal of Monetary Economics*, **54** (7), 1882–1904.

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 straight-forward. 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.[39] 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.

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).

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

| | # Observations in Each RE Percentile Group | | | |
Age group	Median	Min	Max	Total ('000s)
25-29	69,287	43,500	71,580	6,728
30-34	124,707	74,913	127,996	12,048
35-39	133,598	81,547	136,480	12,897
40-44	138,568	85,502	141,336	13,358
45-49	133,355	84,145	136,090	12,868
50-54	117,533	75,954	120,072	11,361
55-59	73,618	49,723	76,442	7,182

Note: Each entry reports the statistics of the number of observations in each of the 100 RE percentile groups for each age. Cross-sectional moments are computed for each year and then averaged over all years, so 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.

---

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

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 40,000 on which the computation of higher-order moments is based.

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

Our main sample used to construct cross-sectional moments and impulse response moments covers a period 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 are constructed using the 1978-2013 sample that covers 36 years. During this sample period self employment income is capped by the SSA maximum taxable earnings limit before 1994. In this section we discuss our methodology of imputing self employment income that is capped by this limit.

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 above the threshold χy_t^{max} reported in the MFE data; i.e., $y_{it}^{SE} \geq \chi y_t^{max}$, where $\chi < 1$.[40] For this purpose we take the first year in our sample with uncapped self employment income, 1996, and regress self employment income, $y_{i,1996}^{SE}$ on observables that can be constructed for the period before 1994 as well.[41] In particular, we estimate the following specification using quantile regression on 75 quantiles:

$$\log y_{i,1996}^{SE} =$$
$$\sum_{k=0}^{3} \alpha_{1,k}^h \mathbb{I}\left\{y_{i,1996-k}^W < Y_{\min,1996-k}\right\} + \sum_{k=0}^{3} \alpha_{2,k}^h \mathbb{I}\left\{y_{i,1996-k}^W \geq Y_{\min,1996-k}\right\} \log y_{i,1996-k}^W$$
$$+ \sum_{k=1}^{3} \alpha_{3,k}^h \mathbb{I}\{y_{i,1996-k}^{SE} > \chi y_{1996-k}^{max}\} + \sum_{k=1}^{3} \alpha_{4,k}^h \mathbb{I}\left\{y_{i,1996-k}^{SE} < Y_{\min,1996-k}\right\}$$
$$+ \sum_{k=1}^{3} \alpha_{5,k}^h \mathbb{I}\left\{y_{i,1996-k}^{SE} \geq Y_{\min,1996-k}\right\} \min\left(\log y_{i,1996-k}^{SE}, \log \chi y_{1996-k}^{max}\right) + \varepsilon_{it}$$
$$(16)$$

where $y_{i,t}^W$ 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. We run this regression 7 times for each age group based on their age in year 1996: 25–29, 30–34,..., 50–54, and 55–60. As a result we have regression coefficients indexed by $h = 1, 2, ...7$ and $q = 1, 2, ..., 75$.

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_t^{max} reported in the MFE data. Then, for an individual who is in age group $h = 1, 2, ...7$ in year $t = 1981, 1981, ...1993$

---

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

[41]The first year in our sample with uncapped self employment income is 1994 but we rather use 1996 self employment income in this regression because of low quality in 1992 self employment income.

FIGURE A.1 – Fraction of Top-Coded Self Employment Income Observations
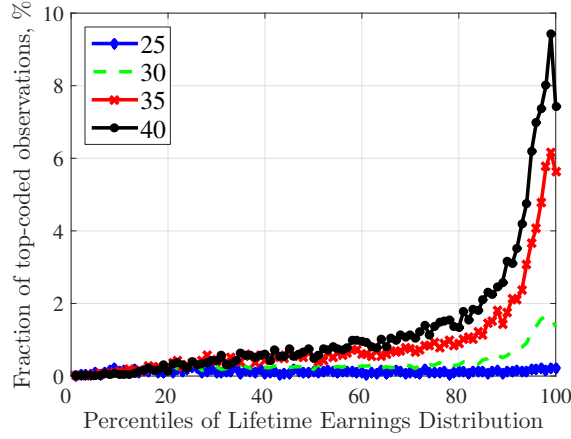
and has self employment income above the limit χy_t^{max}:[42]

$$\log y_{i,t,h}^{SE} =$$
$$\sum_{k=0}^{3} \alpha_{1,k,\tau}^{h} \mathbb{I}\left\{y_{i,t-k}^{W} < Y_{\min,t-k}\right\} + \sum_{k=0}^{3} \alpha_{2,k,\tau}^{h} \mathbb{I}\left\{y_{i,t-k}^{W} \geq Y_{\min,t-k}\right\} \log y_{i,t-k}^{W}$$
$$+ \sum_{k=1}^{3} \alpha_{3,k,\tau}^{h} \mathbb{I}\{y_{i,t-k}^{SE} > \chi y_{t-k}^{max}\} + \sum_{k=1}^{3} \alpha_{4,k,\tau}^{h} \mathbb{I}\left\{y_{i,t-k}^{SE} < Y_{\min,t-k}\right\} \tag{17}$$
$$+ \sum_{k=1}^{3} \alpha_{5,k,\tau}^{h} \mathbb{I}\left\{y_{i,t-k}^{SE} \geq Y_{\min,t-k}\right\} \min\left(\log y_{i,t-k}^{SE}, \log \chi y_{t-k}^{max}\right),$$

where τ is drawn from a uniform distribution on the set of integers from 1 to 75.

FIGURE A.2 – Income Growth for Imputed and Non-Imputed Data
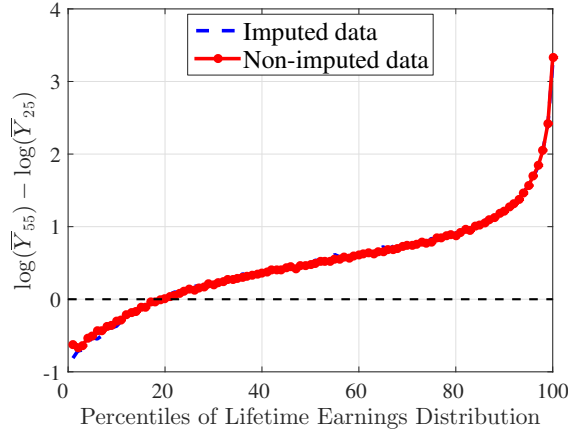
Figure A.2 plots the fraction of top coded self employment income observations against the percentiles of lifetime earnings distribution at ages 25, 30, 35, and 40. Almost no observations

---

[42]We start with year 1981 because we need to observe wage and self employment income in the previous three years.

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.

Figure A.2 plots lifetime income growth against lifetime earnings percentiles using imputed and non-imputed data. The two series are indistinguishable, indicating that top-coding has very little effect on lifetime income growth.

B   Appendix: Robustness and Additional Figures

This section reports additional results from the data. Section B.1 reports the cross-sectional moments of one-year earnings growth. Section B.2 shows the cross-sectional moments of one-year and five-year arc percent changes of earnings. Section B.3 investigates why skewness becomes more negative over the life cycle. Finally, Section B.4 shows several features of the data that are mentioned in the paper but are relegated to the appendix.

B.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 A.3 – Standard Deviation of One-Year Earnings Growth

(A) Second Standardized Moment

(B) P90-10
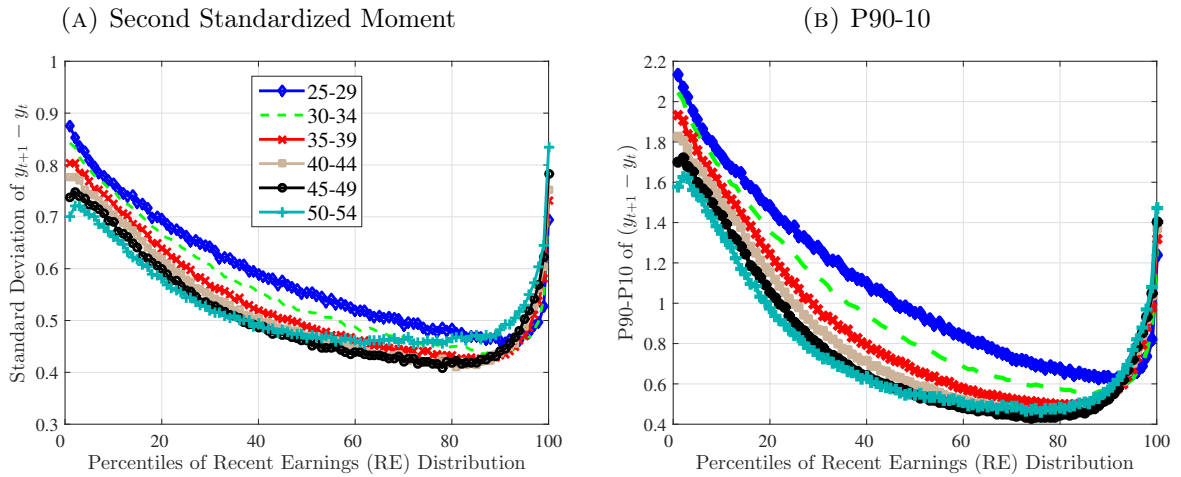
FIGURE A.4 – Skewness of One-Year Earnings Growth

(A) Third Standardized Moment

(B) Kelley Measure
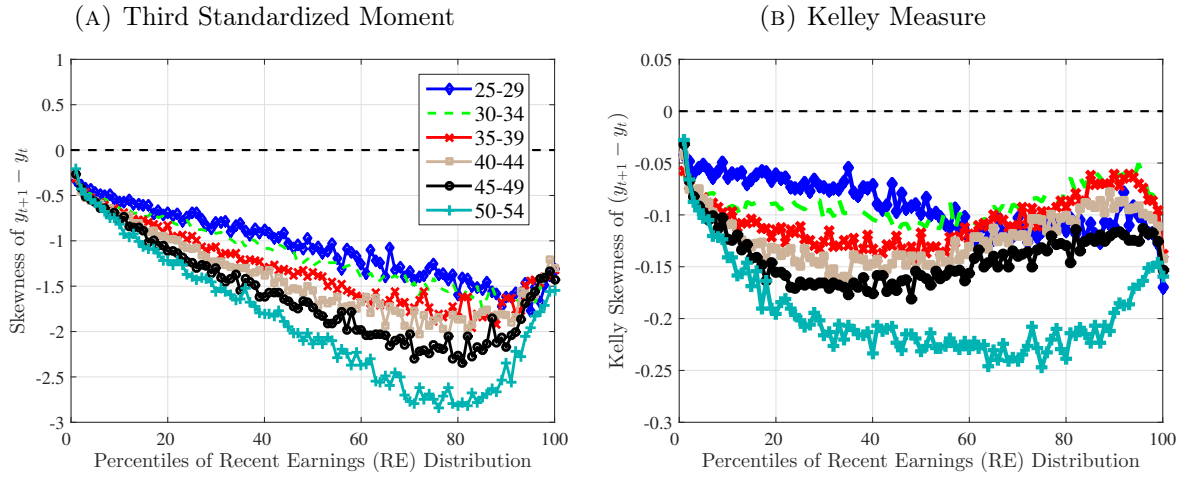
FIGURE A.5 – Kelley's Skewness Decomposed: Change in P90-P50 and P50-P10 Relative to Age 25–29

(A) P90-P50

(B) P50-P10
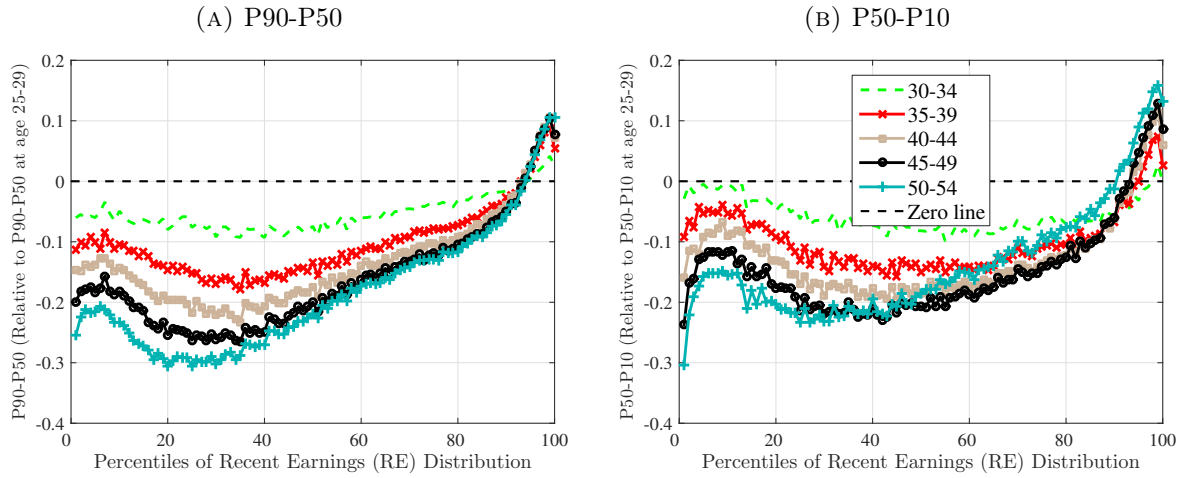
FIGURE A.6 – Kurtosis of One-Year Earnings Growth

(A) Fourth Standardized Moment   (B) Crow-Siddiqi Measure
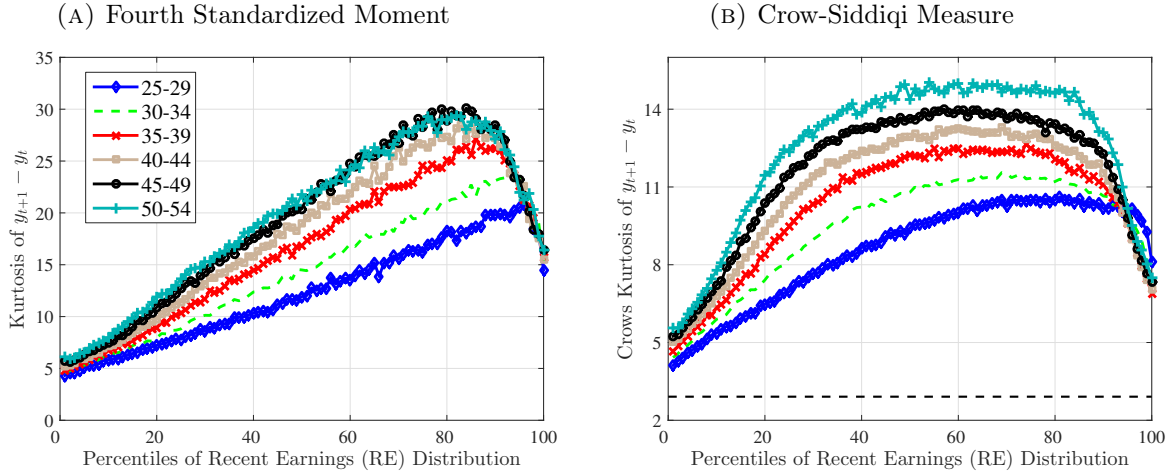
B.2   Arc-Percent Moments

Throughout the paper, we documented moments of log earnings changes. To do 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, Section B.2 reports the cross-sectional moments of arc percent changes defined in Section 2, which we reproduce here for convenience:

$$\text{arc percent change:} \quad \Delta_{\text{arc}}Y^i_{t,k} = \frac{Y^i_{t+k} - Y^i_t}{\left(Y^i_{t+k} + Y^i_t\right)/2}. \tag{18}$$

This measure allows computation of time differences even when the individual has zero income in one of the two years. Section B.2.1 shows the moments for one-year arc percent change, whereas B.2.2 shows the moments for five-year change.

B.2.1 Moments of One-Year Arc Percent Changes

FIGURE A.7 – Standard Deviation of Annual Earnings Change (Arc percent)

(A) Second Standardized Moment

(B) P90-10
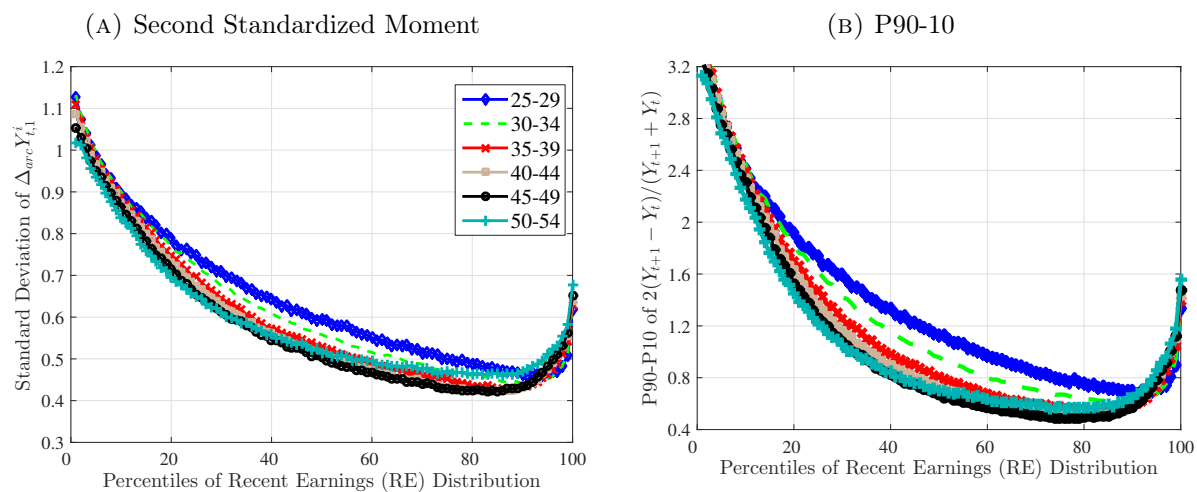
FIGURE A.8 – Skewness of Annual Earnings Change (Arc percent)

(A) Third Standardized Moment

(B) Kelley Measure
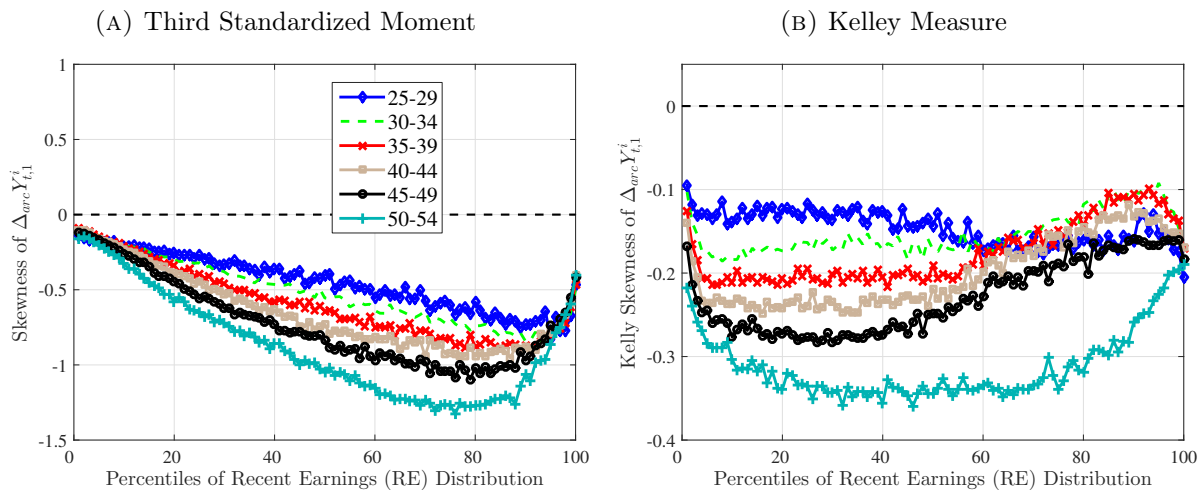
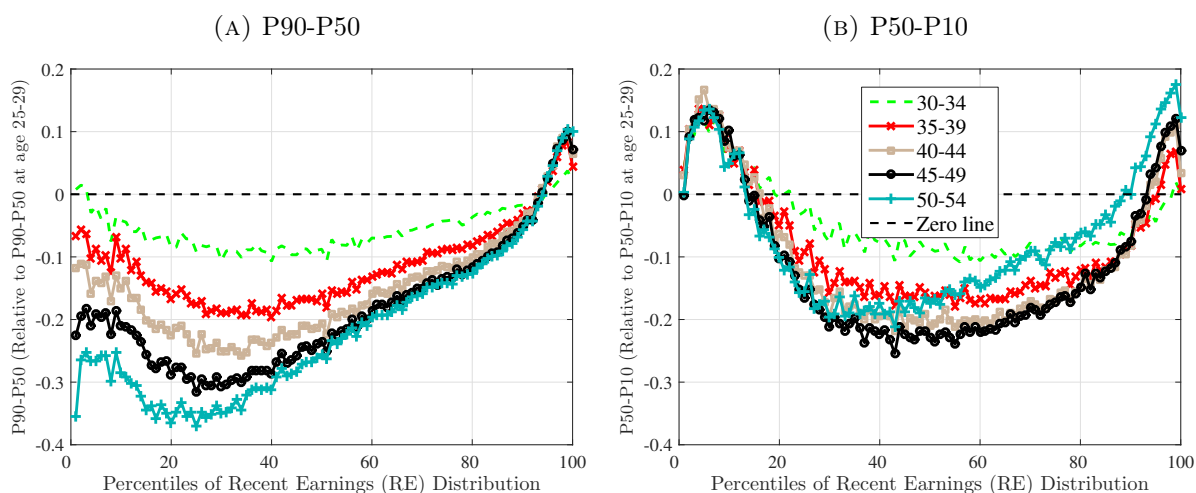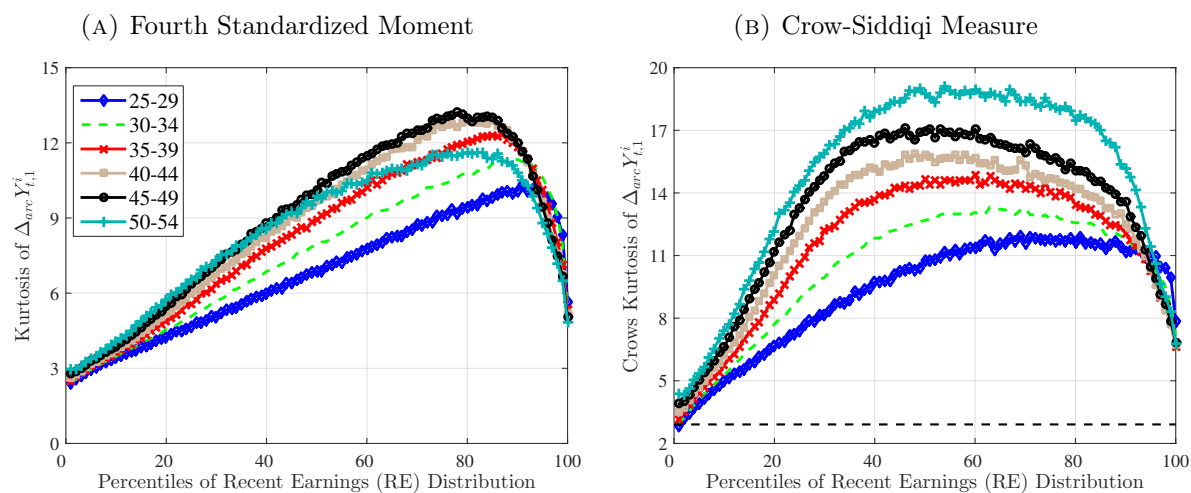FIGURE A.9 – Kelley's Skewness Decomposed: Change in P90-P50 and P50-P10 Relative to Age 25–30 (Arc percent)

(A) P90-P50

(B) P50-P10

FIGURE A.10 – Kurtosis of Annual Earnings Change (Arc percent)

(A) Fourth Standardized Moment

(B) Crow-Siddiqi Measure

B.2.2 Moments of Five-Year Arc Percent Changes

FIGURE A.11 – Standard Deviation of Five-Year Earnings Change (Arc percent)

(A) Second Standardized Moment   (B) P90-10
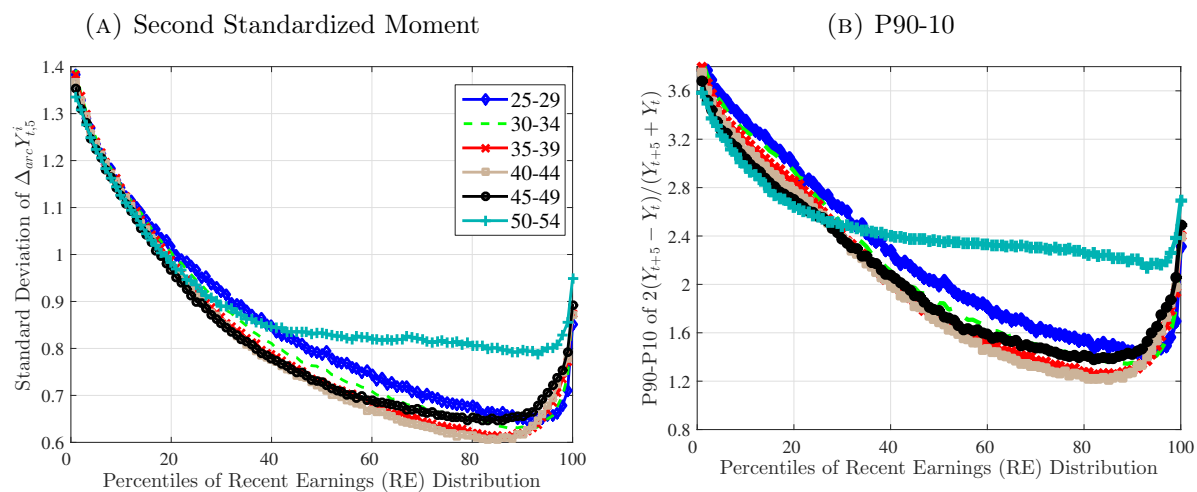
FIGURE A.12 – Skewness of Five-Year Earnings Change (Arc percent)

(A) Third Standardized Moment   (B) Kelley Measure
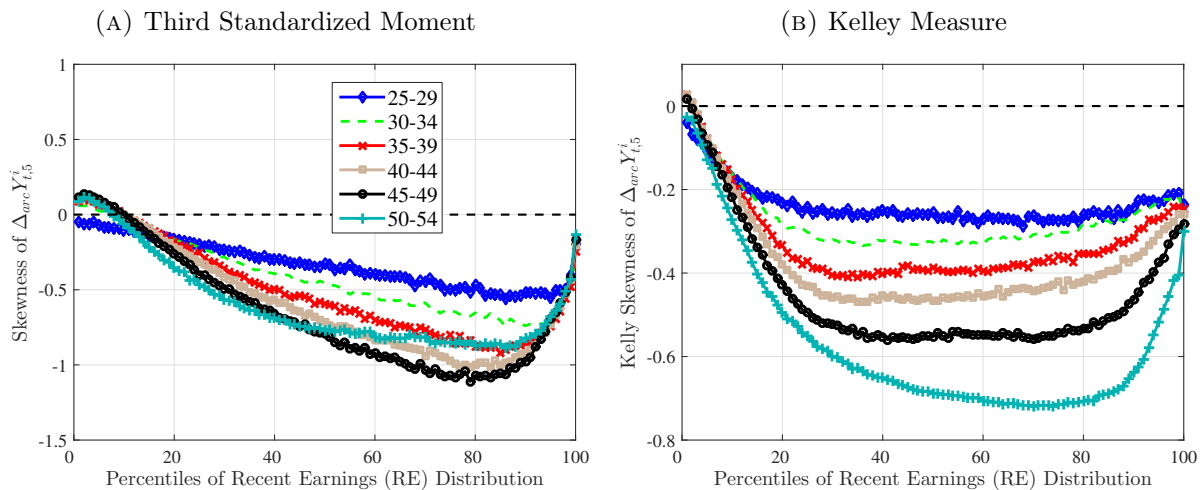
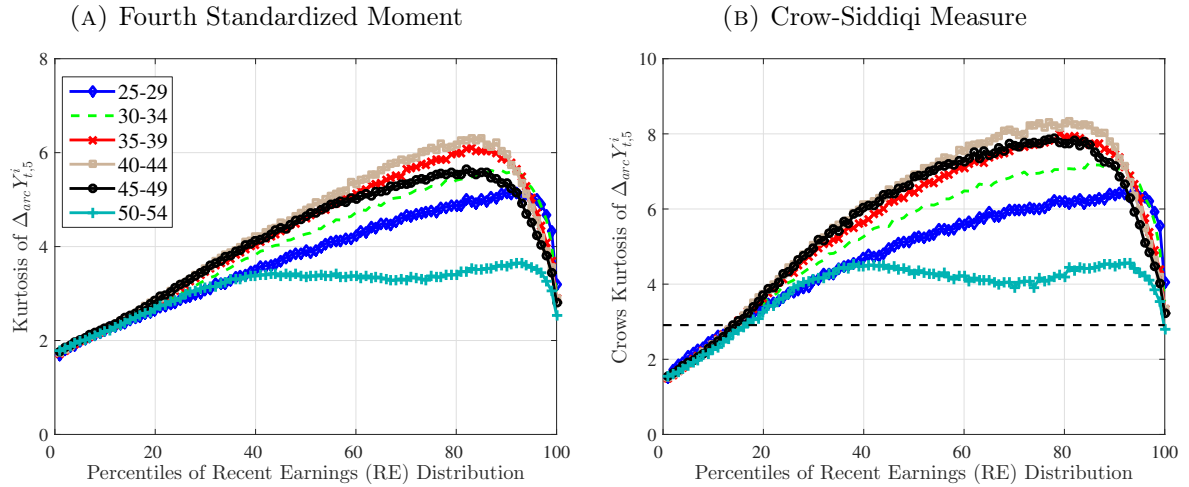FIGURE A.13 – Kelley's Skewness Decomposed: Change in P90-P50 and P50-P10 Relative to Age 25–30 (Arc percent)

(A) P90-P50

(B) P50-P10
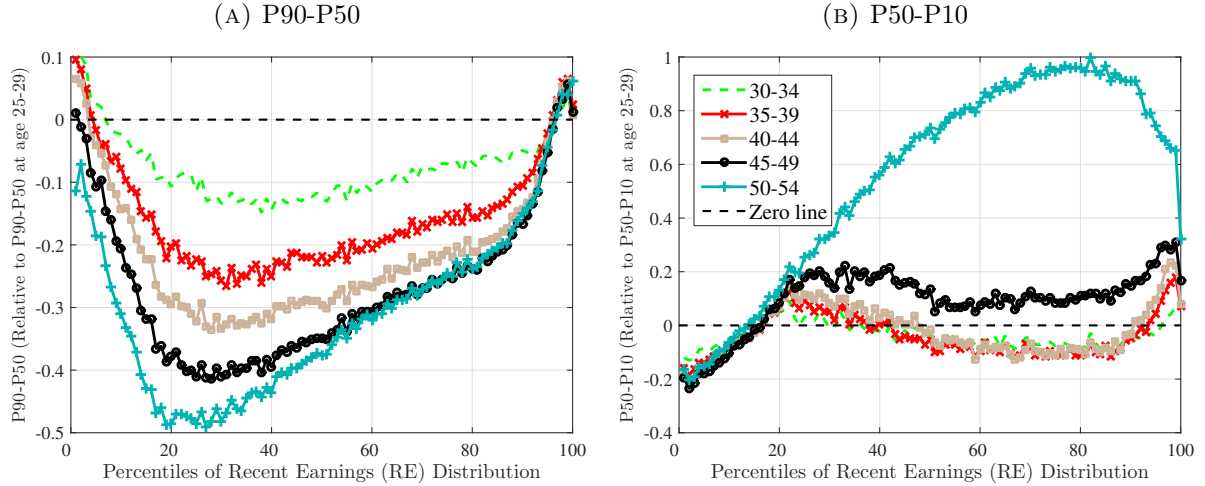
FIGURE A.14 – Kurtosis of Five-Year Earnings Change (Arc percent)

(A) Fourth Standardized Moment

(B) Crow-Siddiqi Measure

B.3   Why Does Skewness Become More Negative With Age?

A natural follow up question is whether the increasingly more negative skewness over the life cycle is primarily due to a compression of the upper tail (fewer opportunities to move up) or due to an expansion in the lower tail (increasing risk of falling a lot)? For the answer, we need to look at the *levels* of the P90-P50 and P50-P10 separately over the life cycle. The left panel of Figure A.15 plots P90-P50 for different age groups minus the P90-P50 for 25- to 29-year-olds, which serves as a normalization. The right panel plots the same for P50-P10. One way to understand the link between these two graphs and skewness is that keeping P50-P10 fixed over

the life cycle, if P90-P50 (left panel) declines with age, this causes Kelley's skewness to become more negative. Similarly, keeping P90-P50 fixed, a rise in P50-P10 (right panel) has the same effect.

Turning to the data, up until age 49, both P90-P50 and P50-P10 decline with age (across most of the RE distribution). This leads to the declining dispersion that we have seen above. The shrinking P50-P10 would also lead to *rising* skewness if it were not for the faster compression of P90-P50 during the same time. Therefore, from ages 25 to 49, the increasing negativity of skewness is *entirely* due to the fact that the upper end of the shock distribution compresses more rapidly than the compression of the lower end. After age 50, P50-P10 starts expanding rapidly (larger earnings losses becoming more likely), whereas P90-P50 stops compressing any further (stabilized upside). Thus, during this phase of the life cycle, the increasing negativity in Kelley's skewness is due to increasing downward risks and not the disappearance of upward moves. The only exception to this pattern is, again, the top earners (RE95 and above) for whom P90-P50 actually never compresses over the life cycle, whereas the P50-P10 gradually rises as they get older. Therefore, as they climb the wage ladder, these individuals do not face a tightening ceiling, but do suffer from an increasing risk of falling a lot.

FIGURE A.15 – Kelley's Skewness Decomposed: Change in P90-P50 and P50-P10 Relative to Age 25–30

(A) P90-P50

(B) P50-P10
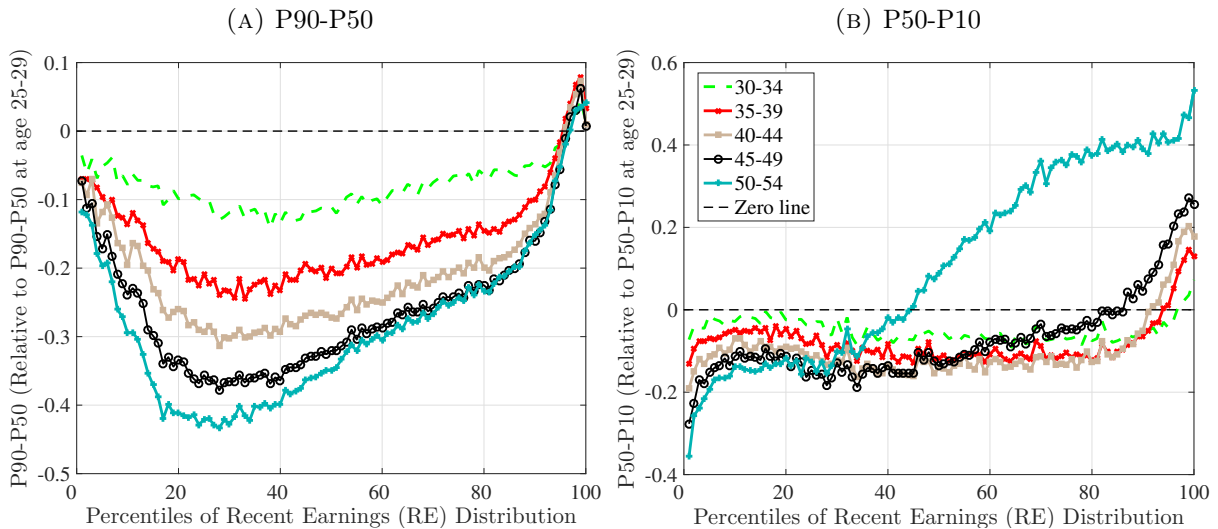
B.4   Further Figures

In this section, we report some additional figures of interest that are omitted from the main text due to space constraints. First, Figure A.16 plots selected percentiles of the future earnings change distribution for every RE percentile.
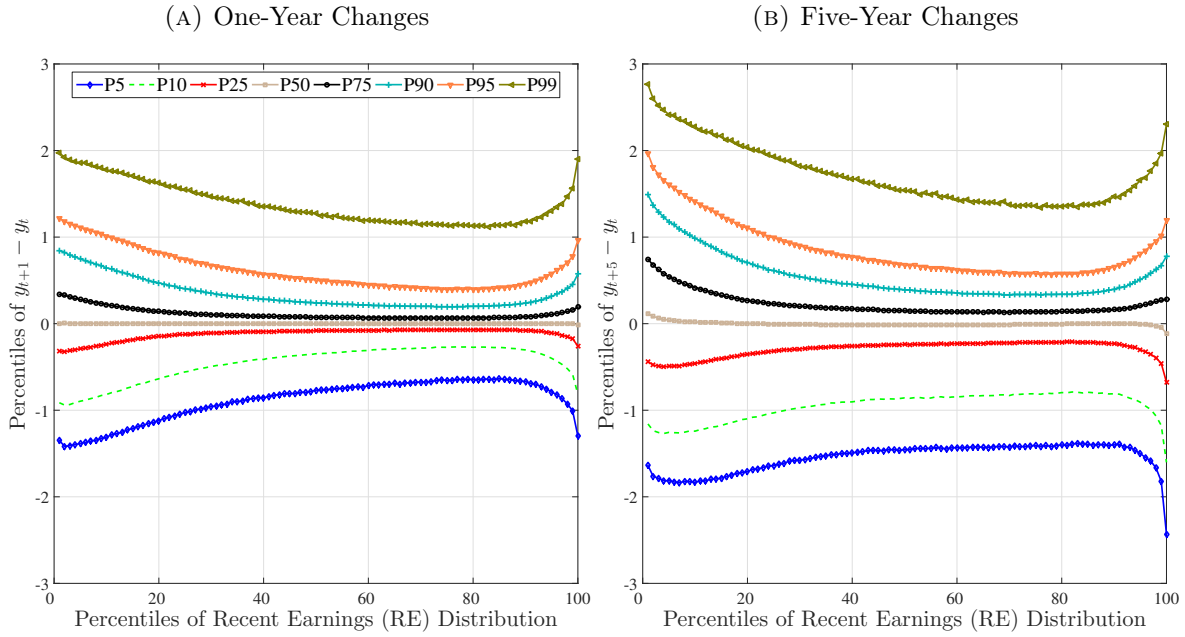
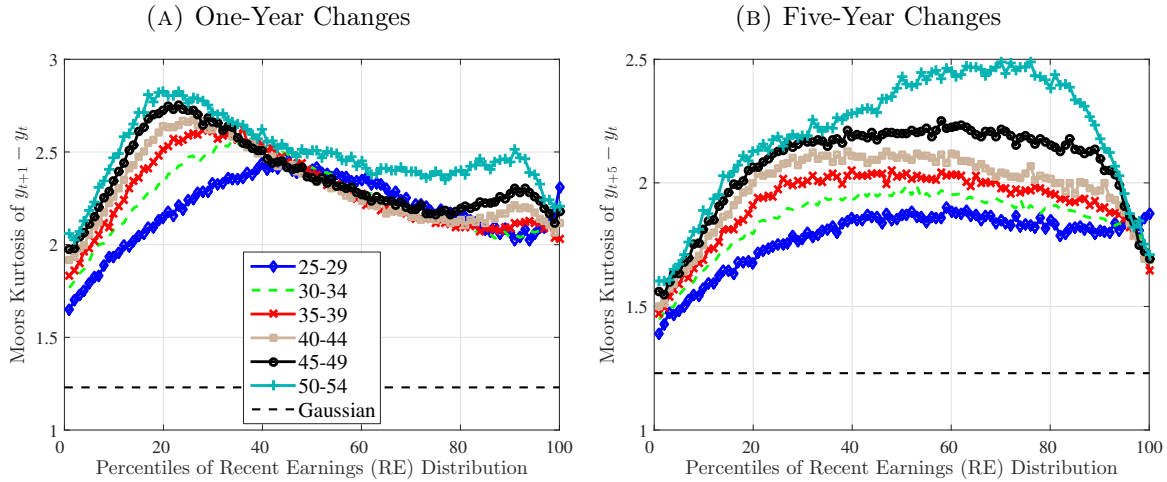FIGURE A.16 – Selected Percentiles of Earnings Changes

(A) One-Year Changes

(B) Five-Year Changes

FIGURE A.17 – Moors Kurtosis of Log Earnings Changes

(A) One-Year Changes

(B) Five-Year Changes

Second, Figure A.17 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
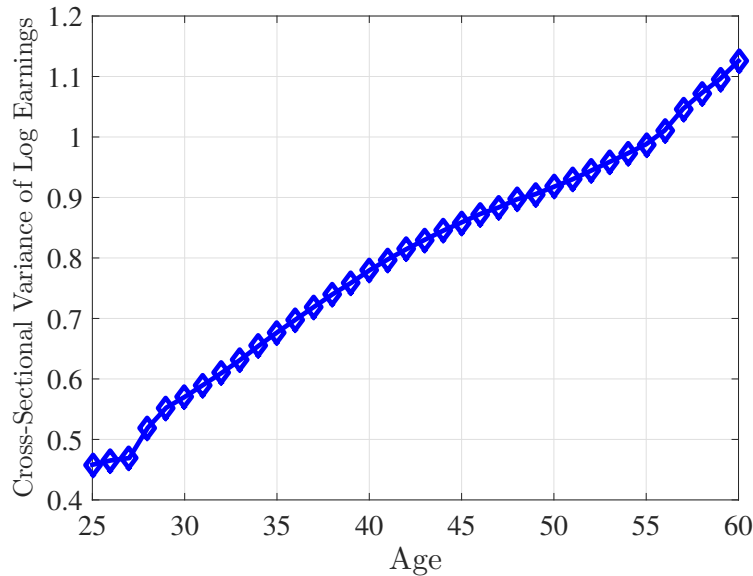
$$\kappa_M = \frac{(P87.5 - P62.5) + (P37.5 - P12.5)}{P75 - P25}.$$

For a Gaussian distribution, Moors' kurtosis always takes the value of 1.23 (shown on dashed

lines).

Lastly, Figure A.16 plots the cross-sectional variance of log earnings over the life cycle, constructed along the lines described in Deaton and Paxson (1994).

FIGURE A.18 – Within-Cohort Variance of Log Earnings

B.5 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 A.19 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 changes increases across the RE distribution.

Second, 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 A.21.

The results are consistent with what we might expect. Job-stayers (i) face a dispersion of earnings changes that is around half that of job-changers, (ii) face shocks that are less negatively skewed as opposed to job-switchers who face shocks that are very negatively skewed, and (iii) experience shocks with much higher kurtosis than job-switchers. In fact, kurtosis is as high as 38 for annual changes and 25 for five-year changes for job-stayers, but is less than 15 for job-switchers at both horizons.

FIGURE A.19 – Fraction Staying Jobs Between t and $t+1$
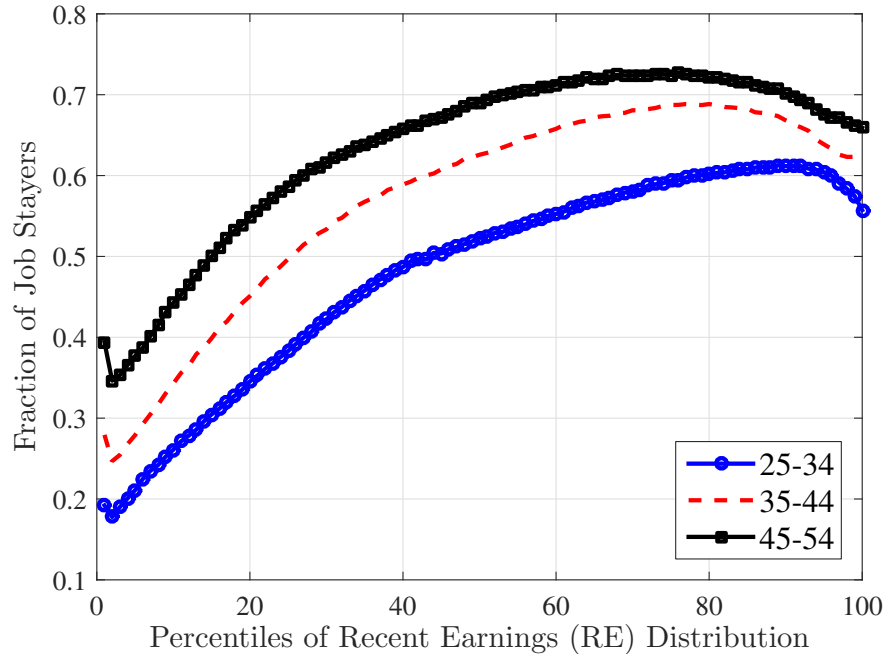
C Estimation

In this section we describe the steps of the estimation procedure in more detail.

Pretesting for Model Specification

To decide on the general stochastic process that we use in estimation, we conducted extensive pretesting as follows. In the first stage, we use each set of moments presented above as diagnostic tools to determine the basic components that should be included in the stochastic process that we will then fit to the data. Clearly, this stage requires extensive pre-testing and exploratory work. For example, to generate the life-cycle earnings growth patterns documented in Section 5, we considered three basic ingredients: (i) an AR(1) process + an i.i.d. shock, (ii) growth rate heterogeneity with no shocks, and (iii) a mixture of two AR(1) processes where each component receives a nonzero innovation with a certain probability. We picked the first two ingredients because of the widespread attention they garnered in the previous literature, and the third one based on our conjecture that it might perform well.

We found that the first ingredient, on its own, could not generate the rich patterns of earnings growth revealed by the data, whereas the HIP process performed fairly well, and the AR(1) mixture process performed the best. Therefore, we concluded that a stochastic process for earnings should include either (ii) or (iii) (or both) as one of its components. Although earnings growth data on their own could not determine which one of these pieces is more important, when we analyze these data along with the other moments, we will be able to obtain sharper identification of the parameters of these two components.

(A) P90-P10, One-year

(B) Kelley's Skewness, One-Year

(C) Crow-Siddiqui Kurtosis, One-Year
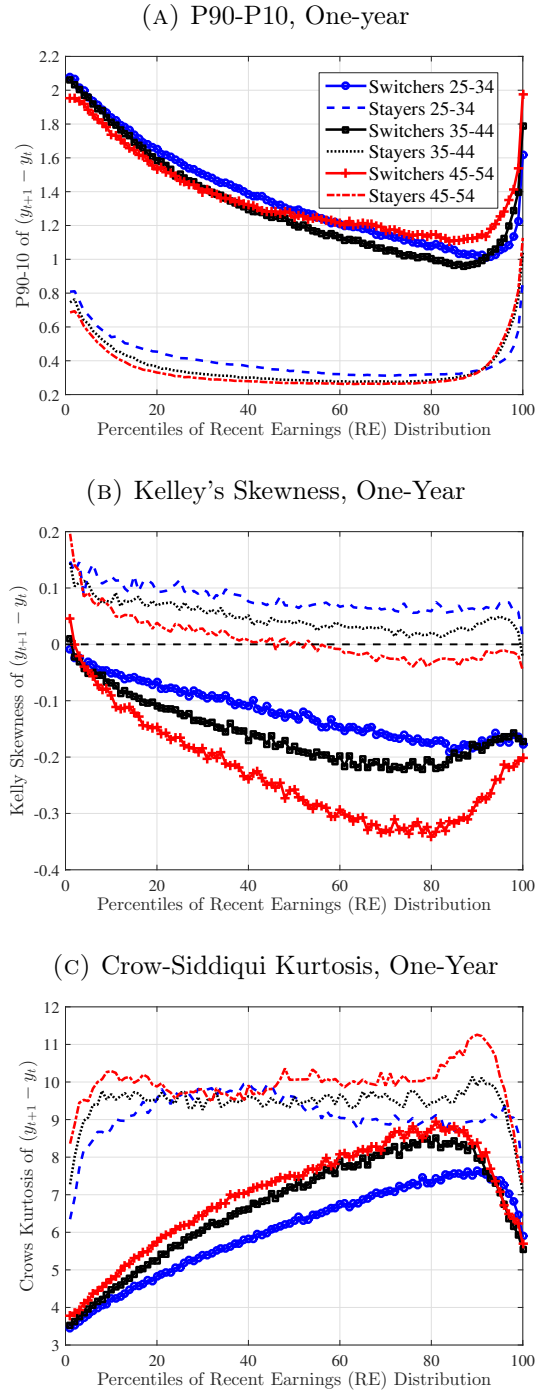
FIGURE A.21 – Second to Fourth Moments of Earnings Growth: Stayers vs Switchers

(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
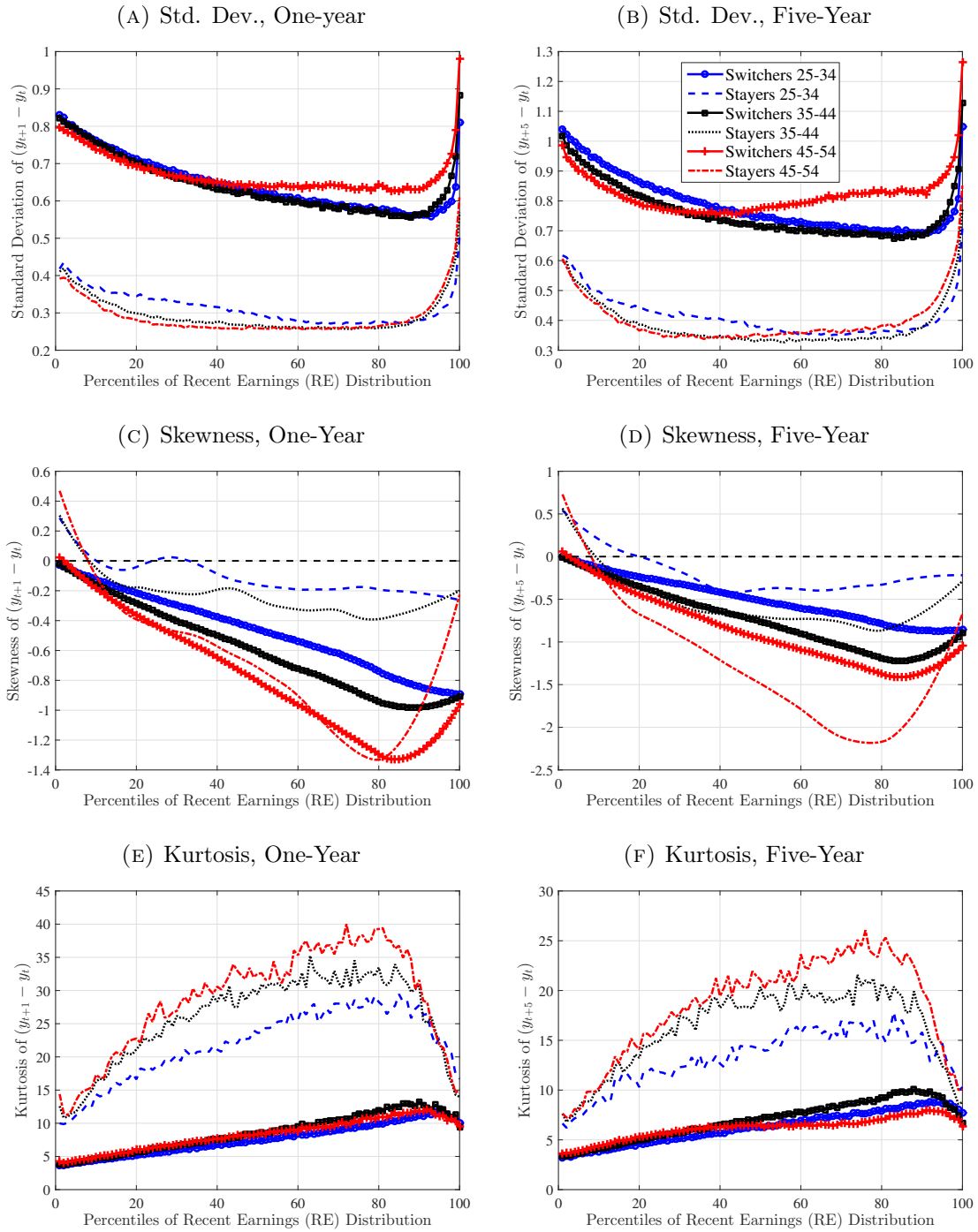
Accounting for Zeros. Recall that in order to construct the cross-sectional moments, we have dropped individuals who had very low earnings—below $\overline{Y}_{\min}$—in year t or $t+k$ so as to allow taking logarithms in a sensible manner.[43] Although this approach made sense for documenting empirical 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. To this end, instead of targeting moments of log earnings change, we target moments of arc percent change, as defined in equation (18). According to this measure, an income change from any level to 0 corresponds to an arc percent change of -2, whereas an income change from 0 to any level indicates an arc percent change of 2.

Standard Deviation of (Log) Earnings Levels. 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 A.18.

Aggregating Moments. If we were to match all data points in all these moments (i.e, 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 100 RE percentiles into 10 to 15 groups and the 6 age groups into three (ages 25–34, 35–44 and 45–55) for cross-sectional moments and 2 age groups (ages 25–34 and 35–55) for impulse response moments. Full details of this aggregation are included in Section C.1. After the aggregation procedure, we are left with 120 moments that capture average earnings over the life cycle (targeted ages are 25, 30, 35, 40, 45, 50, 55, and 60); 234 moments for cross-sectional statistics (standard deviation, skewness and kurtosis of one-year and five-year earnings growth); and 800 moments coming from the impulse response functions. Adding the 36 moments on the variance of log earnings, in sum, we target a total of $120 + 234 + 800 + 36 = 1,190$ moments.[44]

Let m_n for $n = 1, ..., N = 1,190$ 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

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

where $\psi_n > 0$ is an adjustment factor. When $\psi_n = 0$ and m_n is positive, F_n is simply the percentage deviation between data and model moments. This measure becomes problematic

---

[43] We were able to include those below the threshold in sets (i) and (iii) because for those moments it made sense to first take averages, including zeros, and then take the logarithms of those averages.

[44] 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.

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

$$\hat{\theta} = \arg\min_{\theta} \boldsymbol{F}(\theta)' W \boldsymbol{F}(\theta), \tag{19}$$

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

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

The weighting matrix, W, is chosen such that 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.[45]

C.1 Moment Selection and Aggregation

1. 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_{t-1}^i. 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 moments 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 A.22.

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 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 $k = 1, 2, 3, 5, 10$, that is, $\mathbb{E}[\Delta_{arc} Y_{t,k}^i]$, conditional on groups formed by crossing age, recent earnings $\overline{Y}_{t-1}$, and earnings change between $t-1$ and t $\Delta_{arc} Y_{t,1}^i$. We aggregate age groups into two: young workers (25-34) and prime-age workers (35-55). In each year, individuals are assigned to groups based on their ranking in the age-specific RE distribution. The following list defines these groups in terms of the RE distribution: 1–5, 6–10, 11–30, 31–50,

---

[45]More precisely, each life-cycle growth moment is weighed by 0.25/120, each cross-sectional moment by 0.35/234, each impulse response moment by 0.25/800, and each variance moment by 0.15/36.
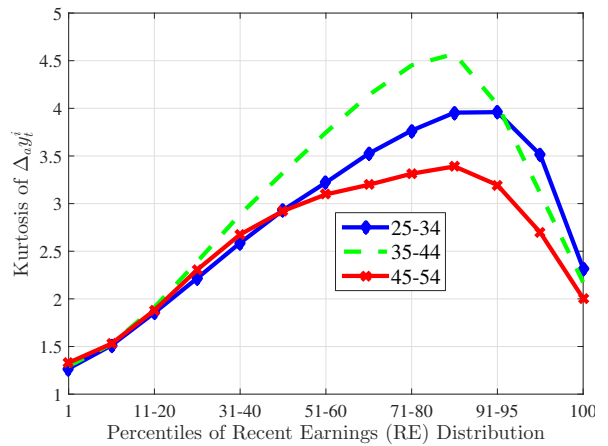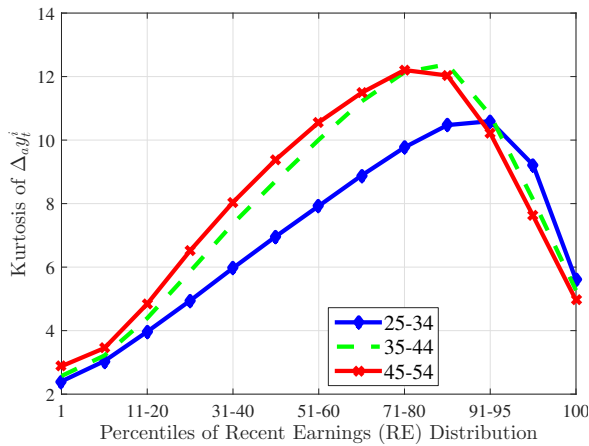
FIGURE A.22 – Cross Sectional Moments Targeted in the Estimation

(A) Standard Deviation of One-Year Earnings Growth (B) Standard Deviation of Five-Year Earnings Growth
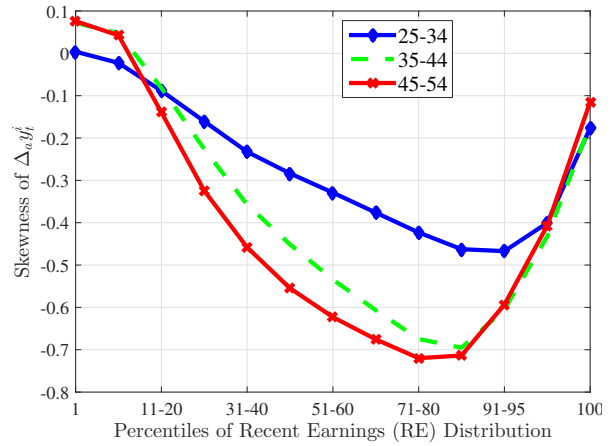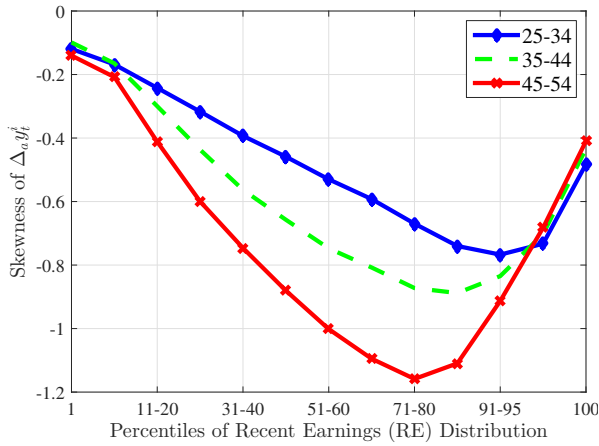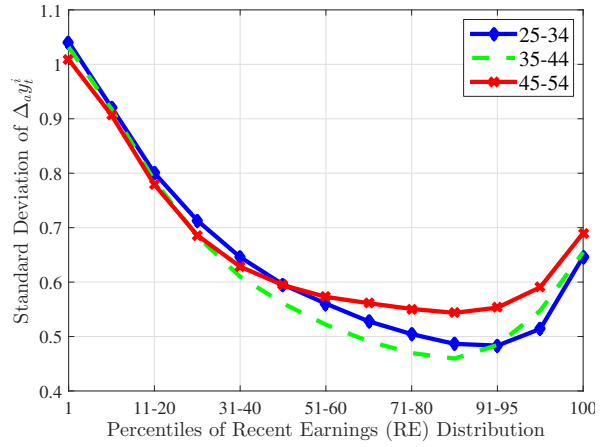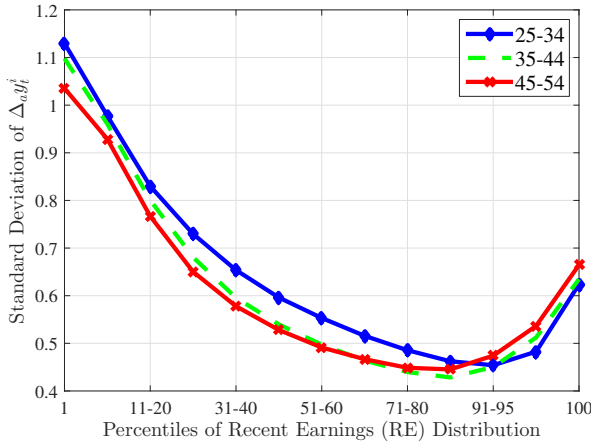
(C) Skewness of One-Year Earnings Growth

(D) Skewness of Five-Year Earnings Growth

(E) Kurtosis of One-Year Earnings Growth

(F) Kurtosis of Five-Year Earnings Growth

51–70, 71–90, 91–95, 96–100. We then group workers based on the percentiles of the age- and RE-specific income change distribution, $\Delta_{arc}Y_{t,1}^i$: 1–2, 3–5, 6–10, 11–30, 31–50, 51–70, 71–90, 91–95, 96–98, 99–100. As a result, we have a total of $2 \times 8 \times 10 \times 5 = 800$ moments based on impulse response. The impulse response functions targeted in the estimation are plotted in figures A.23a–A.24. More specifically, Figure A.23a plots for prime age workers with median recent earnings, the mean reversion patterns at various horizons. Figures A.23b and A.24c do the same for workers at the 90th and 10th percentiles of the recent earnings distribution, respectively. Lastly, figure A.24 shows the variation of these impulse response functions with recent earnings.

4. Variance of log earnings. 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. Note that the life-cycle inequality profile is plotted on Figure A.18.

C.2   Numerical Method for Estimation

The estimation objective is maximized as described now. The global stage is a multi-start algorithm where candidate parameter vectors are uniform Sobol' (quasi-random) points. We typically take about 10,000 initial Sobol' points for pre-testing and select the best 2000 points (i.e., ranked by objective value) for the multiple restart procedure. 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.

C.3   Additional Estimation Results

This section contains estimation results not reported in the main text. We first estimate additional specifications. Then, for all of the estimated income processes, we report some of the parameters that were not reported in Table II.

Model Fit: Additional Figures   Lastly, in figure A.26, we show the fit of the benchmark specification on impulse response moments. Eight panels correspond to the dynamics of earnings changes for various recent earnings percentiles. The plotted fit is for prime-age workers.

Results for Additional Specifications   Table A.2 reports the estimates of five specifications. The first column has two modifications to our benchmark process: It contains a random walk component that is subject to shocks with age- and income-varying probabilities but does not contain an nonemployment shock. To economize on parameters, we do not allow the innovation variances to the random walk component to be heterogeneous. Columns (2) through (4) are also simple variations of the benchmark process. In (2), the process does not have a HIP component, but is otherwise identical to the benchmark specification. Column (3) shuts down the heterogeneity in innovation variances, column (4) shuts down the dependence of shock probabilities on age and income. Finally, the last column estimates a version of the standard AR(1) process, where the AR(1) component gets hit by shocks with age- and income-dependent probabilities.

(A) Workers with Median RE

(B) Workers at the 90th Percentile of RE Distribution

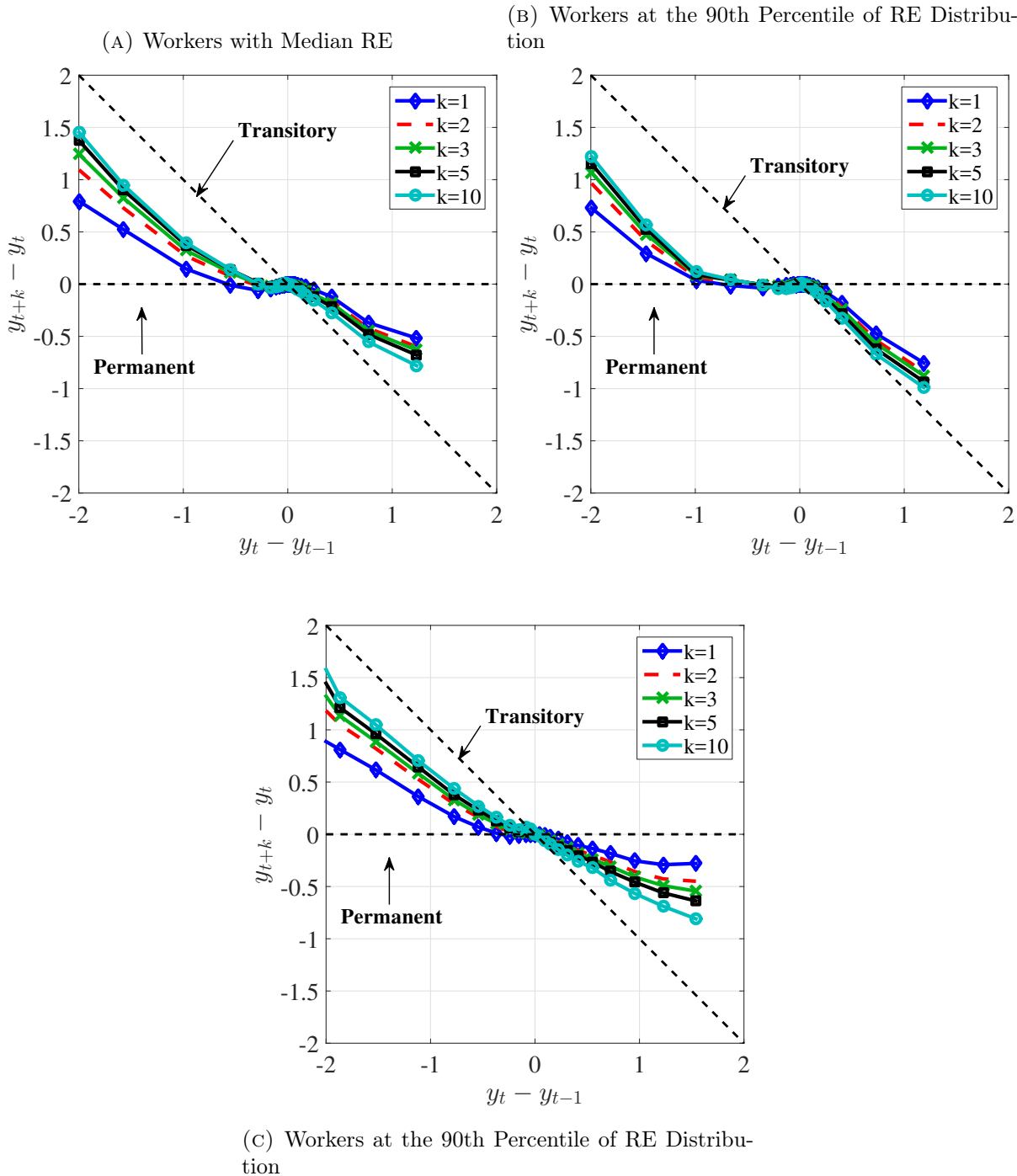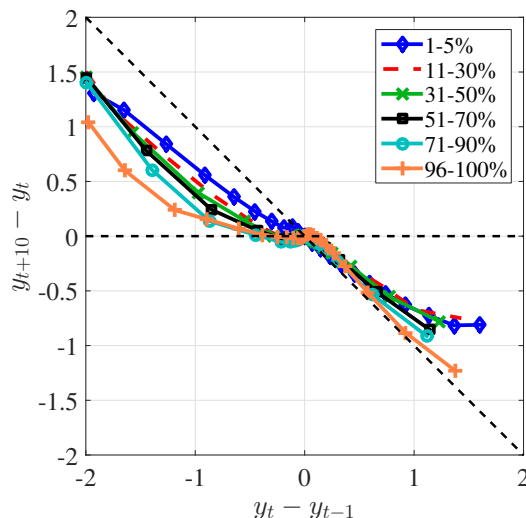(C) Workers at the 90th Percentile of RE Distribution

FIGURE A.24 – Asymmetric Mean Reversion: Butterfly Pattern

Parameter Estimates Table A.3 contains several parameters that were not reported in the main text due to space constraints. These parameters include the means of shocks and the coefficients on age and income in the probability functions for 11 estimated specifications.[46]

Since it is difficult to interpret the magnitudes of the coefficients on shock probabilities, in Figure A.27 we visually show the estimated relationship between shock probabilities and age and income for the benchmark specification.

D   Details of Consumption Model

Retirement Pension There is a social security system that pays pension after retirement. We model the retirement pension as a function of the average earnings below social security maximum taxable earnings of worker's own type k, AE^k over the life cycle. In particular, the pension system mimics the U.S. social security "*primary insurance amount (PIA)*" which is the benefit a person would receive if he begins receiving retirement benefits at the normal retirement age:

$$y_R^k = \begin{cases} 0.9AE^k & \frac{AE^k}{AE} < 0.23 \\ 0.2AE^k + 0.32(AE^k - 0.23AE) & 0.23 < \frac{AE^k}{AE} < 1.38 \\ 0.57AE + 0.15(AE^k - 1.38AE) & \frac{AE^k}{AE} > 1.38 \end{cases}, \quad t = T_W + 1, ...T,$$

where AE is the average earnings in the economy.

D.1   Numerical Solution of the Consumption Model

In order to be able to fully capture the features of the rich earnings dynamics we estimated in this paper, we chose not to discretize the income process when solving the consumption

---

[46]We estimated 12 specifications in total, but all parameters have already been reported for the "Gaussian process" in the main text.

FIGURE A.25 – Fit of Estimated Model to Key Data Moments

(A) Histogram of Arc Percent Change

(B) Life-Cycle Earnings Growth

(C) Std. dev.

(D) Skewness

(E) Kurtosis

(F) Variance of Log Earnings
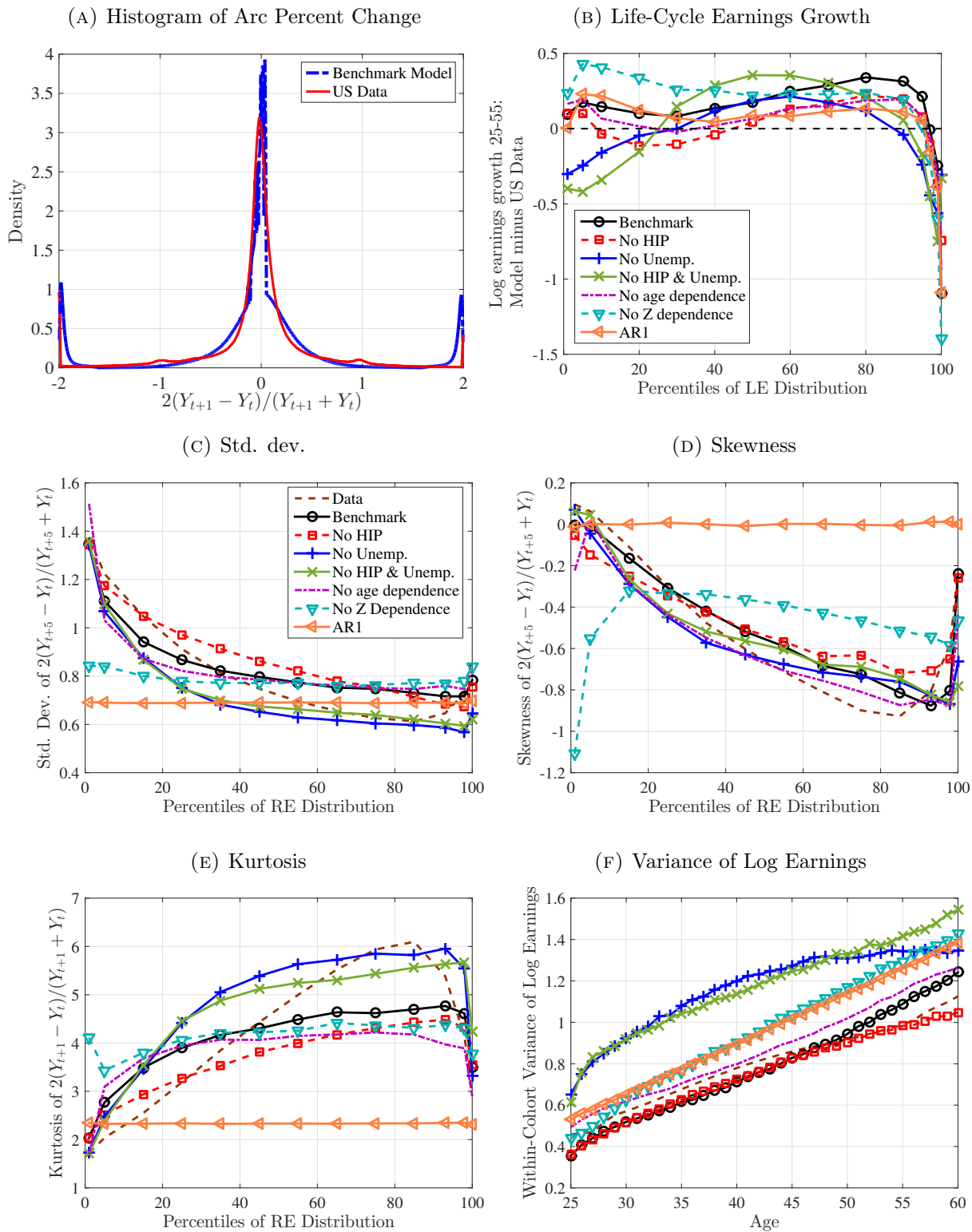
FIGURE A.26 – Fit of Estimated Model to Key Impulse Response Data Moments for the 8 RE Groups
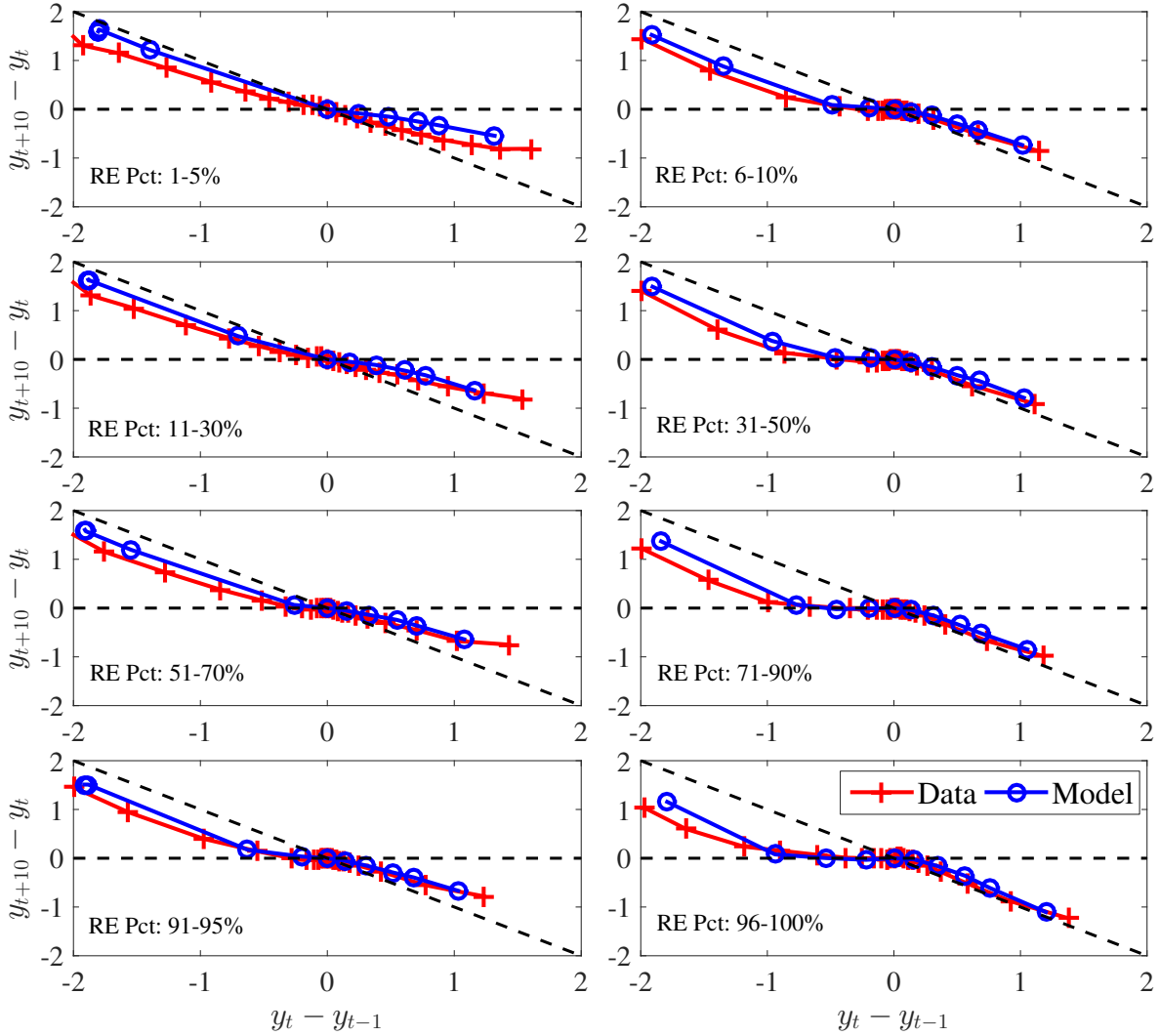
model. Thus it is worth discussing our numerical solution methodology to solve an otherwise standard Bewley model. We employ value function iteration to solve for the policy function for the consumption decision and then use the policy function to simulate consumption-savings decisions of individuals. We now discuss the details below.

D.1.1 Bounds and Grids

For a given $\left(\alpha^k, \beta^k, \sigma_{z_1}^k, \sigma_{z_2}^k\right)$ − type worker, the value function has three continuous variables for each age t, asset holdings, and two persistent components, $a_t, z_{1,t}, z_{2,t}$. We start simulating the continuous income process which is used in the simulation of the consumption path of individuals. We take the minimum and maximum values of simulated persistent components to be the bounds of the $z_{1,t}$ and $z_{2,t}$ spaces. We choose grip points in the $z_{1,t}$ and $z_{2,t}$ spaces to

TABLE A.2 – Estimates of Stochastic Process Parameters for Additional Specifications

Specification:	(1)	(2)	(3)	(4)	(5)
HIP order	1	0	1	1	0
AR(1)	2	2	2	2	1
RW	yes	no	no	no	no
Nonemployment. shock	no	no	yes	yes	no
Heterog. variances?	yes	yes	no	yes	no
Prob. age/inc.	yes/yes	yes/yes	yes/yes	no/no	yes/yes
Parameters					
σ_α	0.197	0.319	0.388	0.552	0.588
$\sigma_\beta \times 10$	0.045	—	0.228	0.180	—
$\mathrm{corr}_{\alpha\beta}$	0.712	—	0.330	0.543	—
ρ_1	-0.006	0.645	0.459	0.024	0.993
ρ_2	0.786	1.016	0.961	0.774	—
$\bar{\sigma}_1$	0.967	1.992	0.494	0.703	0.999
$\bar{\sigma}_2$	1.855	1.077	0.520	2.030	—
$\bar{\sigma}_3$	0.551				
σ_1^i	0.624	0.079	—	0.327	—
σ_2^i	0.040	0.903	—	0.006	—
$\sigma_{1,0}$	0.272	0.145	0.343	0.581	0.376
$\sigma_{2,0}$	0.300	0.250	0.662	0.017	—
σ_ε	0.055	0.242	0.085	0.024	0.303
Objective value	**23.18**	**38.15**	**23.37**	**46.78**	**43.89**
Decomposition:					
(i) Cross-section	0.83	3.15	2.03	8.77	5.93
(ii) Impulse resp.	2.97	7.44	2.48	11.33	8.75
(iii) Inc. growth	1.02	1.60	0.77	1.17	2.90
(iv) Inequality	0.55	2.35	0.19	0.61	1.68

be equally spaced.[47] We have 21 and 41 grid points for $z_{1,t}$ and $z_{2,t}$ spaces, respectively.

As for the asset space, the borrowing limit is set to be the each type's own simulated average earnings, $\bar{A}^k$. The upper bound for the asset spaced is chosen such that in simulations no individual's asset holdings come close to this upper bound, A_t. We choose grids in the a_t

---

[47]Since $z_{1,t}$ and $z_{2,t}$ are in log values polynomially spaced grid points along these dimensions performed worse.

TABLE A.3 – Additional Parameter Estimates for Estimated Specification

Specification:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
	Benchmark model										
HIP order	1	1	0	1	1	1	1	0	1	1	0
AR(1)	2	1	2	2	2	2	2	2	2	2	1
RW	no	no	no	no	no	no	yes	no	no	no	no
Nonemp. shock	yes	yes	yes	no	yes	yes	no	no	yes	yes	no
Heterog. variances?	yes	yes	yes	yes	yes	yes	yes	yes	no	yes	no
Prob. age/inc.	yes/yes	yes/yes	yes/yes	yes/yes	yes/no	no/yes	yes/yes	yes/yes	yes/yes	no/no	yes/yes
Parameters											
μ_1	0.319	-0.599	0.179	-2.162	0.569	0.300	0.803	-2.239	0.287	0.247	-2.999
μ_2	-0.576	—	-0.436	-1.445	-3.993	-1.177	-1.999	-2.028	-0.190	-3.992	—
μ_3	—	—	—	—	—	—	-1.275	—	—	—	—
Shock probabilities:											
a_1 $\times 1$	-2.987	-0.810	-1.797	-2.427	-1.893	-2.390	-2.630	-2.222	-1.730	-1.265	-0.874
b_1 $\times t$	0.024	0.765	-0.146	0.021	0.163	—	-0.927	-0.061	-1.397	—	-1.987
c_1 $\times y_{t-1}$	0.783	-1.682	0.179	-1.661	—	0.913	-0.049	-2.119	-0.349	—	-2.360
d_1 $\times t \times y_{t-1}$	0.680	-1.288	0.193	0.024	—	—	0.435	0.252	2.056	—	0.188
a_2 $\times 1$	-0.561	—	-0.972	-2.033	-4.041	-1.069	-2.188	-3.971	-0.837	-3.809	—
b_2 $\times t$	0.711	—	0.166	-1.019	0.190	—	0.381	0.139	-0.171	—	—
c_2 $\times y_{t-1}$	-1.934	—	-0.950	-3.025	—	-3.186	-1.487	0.056	-0.906	—	—
d_2 $\times t \times y_{t-1}$	-1.289	—	0.120	1.341	—	—	0.178	-0.357	-0.050	—	—
a_3 $\times 1$	—	—	—	—	—	—	-0.164	—	—	—	—
b_3 $\times t$	—	—	—	—	—	—	-0.484	—	—	—	—
c_3 $\times y_t$	—	—	—	—	—	—	-2.874	—	—	—	—
d_3 $\times t \times y_t$	—	—	—	—	—	—	0.696	—	—	—	—
λ	0.075	—	0.096	—	2.337	0.029	—	—	0.146	1.193	—
a_ν $\times 1$	-3.131	—	-3.011	—	-1.163	-4.044	—	—	-3.492	-3.430	—
b_ν $\times t$	-0.200	—	0.432	—	-0.958	—	—	—	0.071	—	—
c_ν $\times y_t$	-3.908	—	-4.182	—	—	-5.346	—	—	-4.377	—	—
d_ν $\times t \times y_t$	-2.946	—	-3.150	—	—	—	—	—	-4.594	—	—

dimension to be polynomial spaced between $a_t = \bar{A}_k$ and $a_t = A_t$, with an explicit point at $a_t = 0$. We have 60 grid points for the asset space.

D.1.2 Integration and Interpolation

The value function of a type-k individual with $\left(\alpha^k, \beta^k, \sigma_{z_1}^k, \sigma_{z_2}^k\right)$ is given by:[48]

$$V_t^{i,k}\left(a_t^i, z_{1,t}^i, z_{2,t}^i\right) = \max_{c_t^i, a_{t+1}^i} \quad u(c_t^i) + \beta \mathbb{E}_t\left[V_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right) \mid z_{1,t}^i, z_{2,t}^i\right].$$
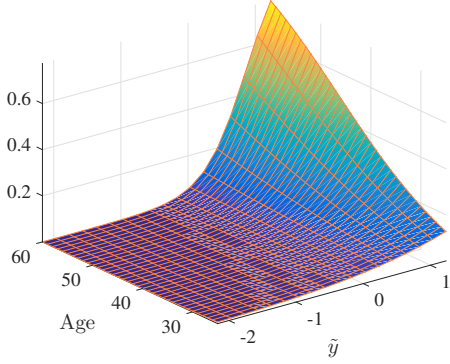
We define the conditional expectation function:

$$\tilde{V}_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t}^i, z_{2,t}^i\right) = \mathbb{E}_t\left[V_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right) \mid z_{1,t}^i, z_{2,t}^i\right].$$

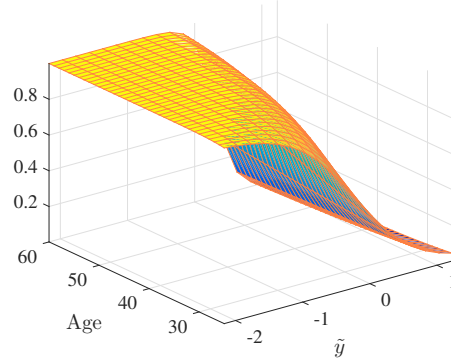[48]The budget constraint and the other equations that properly define the value function are omitted here for simplicity.
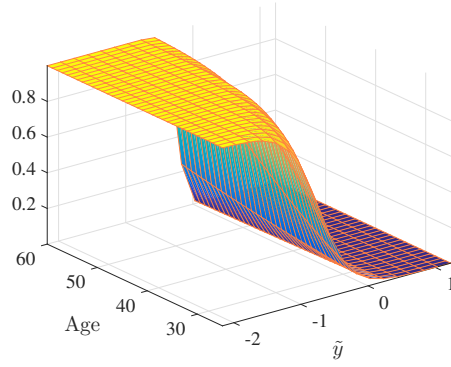
(A) Mixing Probability p_1 for z_t

(B) Mixing Probability p_1 for x_t

(C) Shock Probability p_ν for ν_t

Then the value function becomes:

$$V_t^{i,k}\left(a_t^i, z_{1,t}^i, z_{2,t}^i\right) \quad = \max_{c_t^i, a_{t+1}^i} \quad u(c_t) + \beta \tilde{V}_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t}^i, z_{2,t}^i\right),$$

which reduces the interpolation to only along the a_{t+1} dimension, for which we use one-dimensional spline interpolation.

The conditional expectation, $\tilde{V}_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t}^i, z_{2,t}^i\right) = \mathbb{E}_t\left[V_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right) \mid z_{1,t}^i, z_{2,t}^i\right]$ involves integration over the innovations to both of the persistent components, $\eta_{1,t}^i$ and $\eta_{2,t}^i$ and an nonemployment shock, ν_t^i. Nonemployment shocks are functions of $\left(z_{1,t+1}^i, z_{2,t+1}^i\right)$. Then for given $\left(z_{1,t+1}^i, z_{2,t+1}^i\right)$ we first take the expectation over ν_t^i. For this purpose, we approximate the nonemployment shocks using the Gauss-Laguerre abscissas and weights which is the quadrature that is used for exponential distribution. In particular, let $\left\{x_{\nu,t}^1, x_{\nu,t}^2, ..x_{\nu,t}^{n_\nu}, x_{\nu,t}^{n_\nu+1} = 0\right\}$ and $\left\{w_{\nu,t}^1, w_{\nu,t}^2, ..w_{\nu,t}^{n_\nu}, w_{j,t}^{n_\nu+1} = 1 - p_{\nu,t}\right\}$ be the the Gauss-Laguerre abscissas and weights for the ν_t^i, including the case of full employment, $x_{\nu,t}^{n_\nu+1} = 0$. Note that the weights depend on $\left(z_{1,t+1}^i, z_{2,t+1}^i\right)$. Then, we define the conditional expectation function over nonemployment

shocks as the following:

$$
\begin{aligned}
\mathbb{V}_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right) &= \mathbb{E}_{t,\nu}\left[V_{\nu,t+1}^{i,k}\left(\nu_t^i; a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right) \mid z_{1,t+1}^i, z_{2,t+1}^i\right] \\
&= \sum_{j}^{n_\nu+1} w_{\nu,t}^j V_{\nu,t+1}^{i,k}\left(x_{\nu,t}^j; a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right).
\end{aligned}
$$

Innovations to the persistent components, $\eta_{1,t}$ and $\eta_{2,t}$ are drawn from a mixture of two normal distributions, one of which is a degenerate distribution. The non-degenerate normal distribution is approximated using the Gauss-Hermite abscissas and weights. Let $\left\{x_{j,t}^1, x_{j,t}^2, ..x_{j,t}^{n_j}, x_{j,t}^{n_j+1}\right\}$ and $\left\{w_{j,t}^1, w_{j,t}^2, ..w_{j,t}^{n_j}, w_{j,t}^{n_j+1} = 1 - p_{j,t}\right\}$ be the Gauss-Hermite abscissas and weights for the $\eta_{i,t}$ including the degenerate distribution ($x_{j,t}^{n_j+1}$ and $w_{j,t}^{n_j+1}$). Then, the conditional expectation function over the shocks to the persistent components is defined as:

$$
\begin{aligned}
\tilde{V}_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t}^i, z_{2,t}^i\right) &= \mathbb{E}_t\left[V_{t+1}^{i,k}\left(a_{t+1}^i, z_{1,t+1}^i, z_{2,t+1}^i\right) \mid z_{1,t}^i, z_{2,t}^i\right] \\
&= \sum_{j_1}^{n_1+1}\sum_{j_1}^{n_1+1} w_{1,t}^{j_1}w_{2,t}^{j_2} V_{t+1}^{i,k}\left(a_{t+1}^i, \rho_1 z_{1,t}^i + x_{1,t}^{j_1}, \rho_2 z_{2,t}^i + x_{2,t}^{j_2}\right).
\end{aligned}
$$

The above part requires interpolation over $\left(z_{1,t+1}^i, z_{2,t+1}^i\right)$ for which we use 2-dimensional nested cubic spline. We use 7 abscissas for innovations to the each of the persistent component.

D.1.3   Simulating the Consumption Path

The policy function for the consumption decision has three continuous variables, asset holdings, and two persistent components, $a_t, z_{1,t}, z_{2,t}$. To interpolate the consumption decision we employ the 3-dimensional nested cubic spline interpolation. We do not need an additional state for the nonemployment shocks since they are transitory in nature. Namely, we store the consumption decision in the value function iteration step for only the case of full-year nonemployed shocks. In simulating the nonemployment shocks less than a year, we just add the difference in labor earnings due to the shorter nonemployment spell to the asset holdings of the individuals, which can be thought as cash-on-hand. This allows us to interpolate the consumption decision only over three dimensions, $(a_t, z_{1,t}, z_{2,t})$ rather than 4-dimensional interpolation.

D.1.4   Other Computational Details

We use MPI for the parallel programming. In particular, we use one core for each worker type$-k$ (with a specific $\left(\alpha^k, \beta^k, \sigma_{z_1}^k, \sigma_{z_2}^k\right)$). The values of $\left(\alpha^k, \beta^k, \sigma_{z_1}^k, \sigma_{z_2}^k\right)$ and the number of individuals to be simulated for each $\left(\alpha^k, \beta^k, \sigma_{z_1}^k, \sigma_{z_2}^k\right)$ $-$type are determined using the Gauss-Hermite abscissas and weights, respectively. We have four $\left(\alpha^k, \beta^k\right)$ $-$types, three $\sigma_{z_1}^k$ $-$types, and three $\sigma_{z_2}^k$ $-$types. In total, we have $4 \times 3 \times 3 = 36$ ex-ante types of workers and used 36 cores to compute for 36 different value functions. Once each core finishes simulating consumption path for its own type of worker, the main core gathers all the simulated data and compute statistics for the entire economy. It takes 4-5 mins to solve and simulate the model for a given

discount factor. It usually takes 6-7 iterations to converge to the discount factor that matches the wealth to income ratio observed in the data.

D.1.5   Welfare Analysis

The formula for welfare cost, χ, is given by

$$\chi = 1 - \left(\frac{V}{V_{Complete}} \right)^{1/(1-\gamma)},$$

where V is the expected lifetime utility at time 0 who has not drawn her income shocks for the first period yet, $V_{Complete}$ is the expected lifetime utility in the existence of financial markets against idiosyncratic risk (complete markets), and γ is the coefficient of relative risk aversion.

D.2   Derivation of the Risk Premium, Equation (15)

$$U(c(1 - \pi)) = \mathbb{E}\left(U(c(1 + \tilde{\delta})) \right)$$

Taking a first-order Taylor-series approximation to the left hand side and fourth order approximation to the right hand side we get:

$$U(c) - U'(c)c\pi = \mathbb{E}\left(U(c) + U'(c)c\tilde{\delta} + \frac{1}{2}U''(c)c^2\tilde{\delta}^2 + \frac{1}{6}U'''(c)c^3\tilde{\delta}^3 + \frac{1}{24}U''''(c)c^4\tilde{\delta}^4 \right).$$

Observing that the second term on the right hand side is zero when $\mathbb{E}(\tilde{\delta}) = 0$, and rearranging yields:

$$\pi = -\frac{1}{2}\frac{u''(c)c}{u'(c)} \times m_2 - \frac{1}{6}\frac{u'''(c)c^2}{u'(c)} \times m_3 + \frac{1}{24}\frac{u''''(c)c^3}{u'(c)} \times m_4, \tag{20}$$

where m_n denotes the nth central moment of $\tilde{\delta}$. To convert these into statistics that are reported in the paper, we write $m_2 = \sigma_\delta^2$, $m_3 = s_\delta \times \sigma_\delta^3$, where s_δ is the skewness coefficient, and $m_4 = k_\delta \times \sigma_\delta^4$, where k_δ is kurtosis. With this notation, and assuming a CRRA utility function with curvature γ, we get:

$$\pi^* = \frac{\gamma}{2} \times \sigma_\delta^2 - \frac{(\gamma + 1)\gamma}{6} \times s_\delta \times \sigma_\delta^3 + \frac{(\gamma + 2)(\gamma + 1)\gamma}{24} \times k_\delta \times \sigma_\delta^4,$$

which is equation (15). This can also be written as:

$$\pi^* = \frac{\gamma}{2} \times \sigma_\delta^2 \times \left[1 + \frac{1}{3}(\gamma + 1)\left(-s_\delta \times \sigma_\delta + \frac{1}{4}(\gamma + 2)k_\delta \times \sigma_\delta^2 \right) \right].$$