The Nature of Countercyclical Income Risk*

Fatih Guvenen† Serdar Ozkan‡ Jae Song§

June 2, 2012

Abstract

This paper studies the cyclical nature of individual income risk using a confidential dataset from the U.S. Social Security Administration, which contains (uncapped) earnings histories for millions of individuals. The base sample is a nationally representative panel containing 10 percent of all U.S. males from 1978 to 2010. We use these data to decompose individual income growth during recessions into "between-group" and "within-group" components. We begin with the behavior of within-group shocks. Contrary to past research, we do not find the variance of idiosyncratic income shocks to be countercyclical. Instead, it is the left-skewness of shocks that is strongly countercyclical. That is, during recessions, the upper end of the shock distribution collapses—large upward income movements become less likely—whereas the bottom end expands—large drops in income become more likely. Thus, while the dispersion of shocks does not increase, shocks become more left skewed and, hence, risky during recessions. Second, to study between-group differences, we group individuals based on several observable characteristics at the time a recession hits. One of these characteristics—the average income of an individual at the beginning of a business cycle episode—proves to be an especially good predictor of fortunes during a recession: prime-age workers that enter a recession with high average earnings suffer substantially less compared with those who enter with low average earnings (which is not the case during expansions). Finally, we find that the cyclical nature of income risk is dramatically different for the top 1 percent compared with all other individuals—even relative to those in the top 2 to 5 percent.

JEL codes: E24, E32, J21, J31.

Keywords: Business cycle risk, countercyclical income risk, factor structure, the top 1%, idiosyncratic shocks.

---

*For useful comments and discussions, we thank George Constantinides, Per Krusell, Jonathan Parker, Luigi Pistaferri, Martin Schneider, as well as seminar and conference participants at various institutions. The views expressed herein are those of the authors and not necessarily those of the Social Security Administration, the Federal Reserve Bank of Chicago, or the Board of Governors of the Federal Reserve System.

†University of Minnesota, Federal Reserve Bank of Chicago, and NBER; `guvenen@umn.edu`

‡Federal Reserve Board; `serdar.ozkan@frb.gov`

§Social Security Administration; `jae.song@ssa.gov`

1 Introduction

From 2007 to 2009, US male workers experienced an average decline in their annual labor earnings of 6.5 percent. While this figure represents a very steep decline compared with any other post-war recession, it is dwarfed by the dispersion of income growth rates across workers during the same recession: for example, a quarter of workers saw their labor earnings rise by more than 16 percent, one in ten saw a rise of more than 65 percent, whereas another one in ten saw their incomes fall by more than 55 percent. Moreover, despite the 6.5 percent *mean* decline in earnings just noted, the worker with *median* income change actually experienced a slight rise—of 0.1%—during these two years. The goal of this paper is to understand this wide dispersion of fortunes and how it varies over the business cycle. More specifically, we seek to decompose income changes during a business cycle "episode" (i.e., recession or expansion) into a component that can be predicted based on the observable characteristics of individuals (prior to the episode) and a separate "residual" component that represents purely idiosyncratic shocks that hit individuals that are ex ante very similar. The first one represents the "between-group" component of business cycle risk, whereas the second can be thought of as the "within-group" component.

An important advantage of our analysis is the very rich dataset that we employ. Basically, our main panel dataset is a 10 percent random sample of all US males who had a Social Security number between the ages of 25 and 60 from 1978 to 2010. This dataset has three important advantages. First, earnings records in our dataset are uncapped (no top-coding), allowing us to study individuals with very high incomes.[1] Second, the substantial sample size allows us to employ flexible non-parametric methods and still obtain extremely precise estimates. To give some idea about the size of the sample, the bulk of our analysis is conducted with a sample that has about 4.5 million individuals in each year for a total of more than 165 million individual-year observations during this period. Third, thanks to their records-based nature, the data contain very little measurement error, which is a serious issue with survey-based micro datasets. One drawback is possible underreporting (e.g., cash earnings), which can be a concern at the lower end of the earnings distribution.

The panel aspect of our dataset allows us to use individuals' earnings and employment histories to construct observable characteristics as of the beginning of a business cycle

[1] Kopczuk et al. (2010) also employ an SSA dataset with uncapped earnings (after 1978), whereas Haider and Solon (2006), Schulhofer-Wohl (2011), Bonhomme and Hospido (2012, Spain), and Bönke et al. (2011, Germany) used datasets with earnings capped at the Social Security contribution limit.

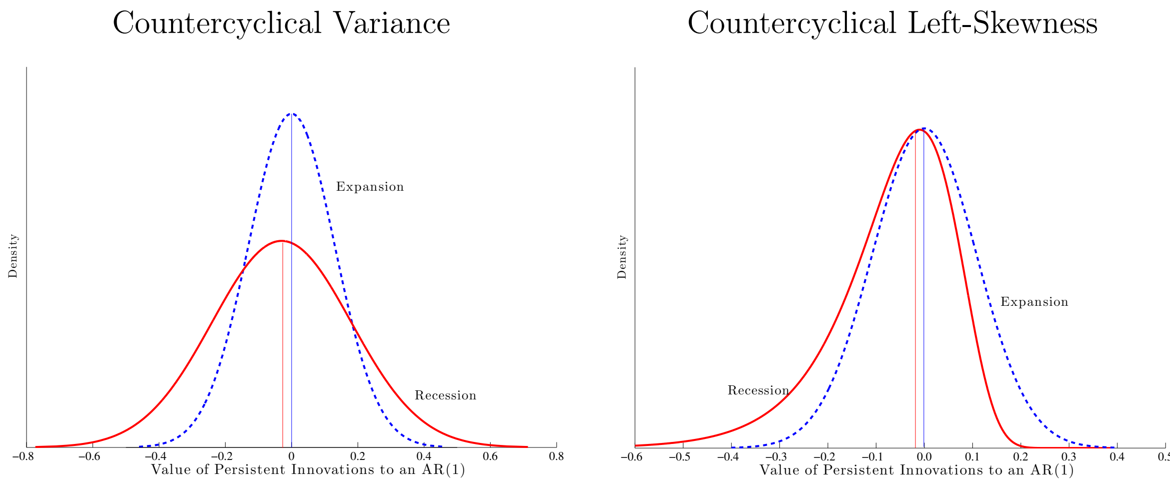


Figure 1: Countercyclical Variance or Countercyclical Left-Skewness?

episode. For example, we can ask whether individuals that entered a recession with a high average income are affected differently *during* the recession relative to those that entered with a low average income. How about individuals who were rising stars (i.e., had fast income growth rate) versus those whose careers were stagnant when the recession hit? Similarly, are some individuals' incomes inherently more sensitive to business cycle fluctuations than others', perhaps because of their occupations, the industry they work in, or the nature of their skill sets, etc.? Finally, how does age factor into any of these patterns? To answer these questions systematically, we group individuals along three observable dimensions at the time a business cycle episode begins: (i) age, (ii) pre-episode average income, and (iii) pre-episode income growth rate.

Our main findings can be summarized as follows. First, we study the cyclical nature of idiosyncratic shocks, once observable factors are accounted for. Contrary to past research, we find that income shock variances are *not* countercyclical. However, uncertainty does have a significant countercyclical component, but it comes from the *left-skewness* increasing during recessions. That is, during recessions, the upper end of the income shock distribution collapses—large upward income movements become less likely—whereas the bottom end expands—large drops in income become more likely. The two scenarios—countercyclical variance versus left-skewness—are shown in Figure 1. Relative to the earlier literature that argued for increasing variance—which results in some individuals receiving much more positive shocks during recessions—our results are even more pessimistic: Uncertainty increases in recessions without an increasing chance of upward movements.

We then turn to the systematic component of business cycle risk. We find substantial between-group variation across individuals that differ in pre-episode average income. For example, when we rank prime-age (35–54) male workers based on their 2002–06 average income, those in the 10th percentile of this distribution experienced a fall in their income during the Great Recession (2007–2010) that was about 18 percent worse than that experienced by those who ranked in the 90th percentile. In fact, average income loss during this recession was almost a linear (upward sloping) function of pre-recession average income all the way up to the 95th percentile (Figure 17). Interestingly, this good fortune of high-income workers did not extend to the very top: those in the top 1%, based on their 2002–2006 average income, experienced an average loss that was *21 percent worse* than that of workers in the 90th percentile. Although these magnitudes are largest for the Great Recession, the same general patterns emerged in the other recessions too. For example, the 1980–83 double dip recession is very similar to the Great Recession for all but the top 5 percentiles. But the large income loss for the top 1% was not observed during that recession at all. In fact, this appears to be a more recent phenomenon: The worst episode for the top 1% was the (otherwise mild) 2000–02 recession, when their average (log) income loss exceeded that of those in the 90th percentile by almost 30 log points.[2]

Our results on the business cycle behavior of top incomes complement and extend the findings in Piketty and Saez (2003), Parker and Vissing-Jørgensen (2010), and Saez (2012). In particular, these papers used repeated cross-sections to construct synthetic groups of individuals based on their income level. They then documented the strong cyclicality of high income groups over the business cycle. With panel data, we are able to track the same individuals over time, which allows us to control for compositional change and measure how persistent the effects of such fluctuations are. Our results confirm the higher cyclicality of top earners and reveal the very high persistence of these fluctuations. For example, individuals who were in the 99.9th percentile based on their 1995–99 average income experienced a 5-year average income loss between 2000 and 2005 that exceeded 50 log points! Similarly large persistent losses are found for the top income earners during the 5-year periods covering the Great Recession (2005–10) as well as the 1990–95 period.

[2] In this paper, "log points" will be used to denote the log difference between two variables multiplied by 100.

1.1 Literature Discussion

The cyclical patterns of idiosyncratic labor income risk have received attention from both macro and financial economists. In an infinite-horizon model with permanent shocks, Constantinides and Duffie (1996) showed that one can generate a high equity premium (Mehra and Prescott (1985)) if idiosyncratic shocks have countercyclical variance. Storesletten et al. (2004) used a clever empirical identification scheme to estimate the cyclicality of shock variances.[3] Using the Panel Study of Income Dynamics (PSID), they estimated the variance of AR(1) innovations to be *three* times higher during recessions. They did not, however, investigate the cyclicality of the skewness of shocks, nor did they allow for a factor structure as we do here. Moreover, note that the question of interest is "the cyclical *changes* in the *dispersion* of income *growth rates*," which involves triple-differencing. It is extremely challenging to answer such a question without a very large and clean dataset.

Our findings are more in line with the way Mankiw (1986) modeled idiosyncratic shocks. Basically, he showed that one can resolve the equity premium puzzle if idiosyncratic shocks have countercyclical left-skewness—as found in the current paper. In a related context, Brav et al. (2002) found that accounting for the countercyclical skewness of individual consumption growth helps generate a high equity premium with a reasonable risk aversion parameter.

The spirit of our analysis is similar to the literature that decomposed wage inequality trends into between-group and within-group components (among many others, Juhn et al. (1993), Lemieux (2006), and Autor et al. (2008)). But there are several notable differences. First, our focus is on growth rates rather than levels, which is feasible with the panel dimension of our dataset. Second, we focus on business cycle variation whereas that literature examined secular trends. Third, relying on repeated cross-sections, that literature had to confine itself to the few observable characteristics that were available in the cross-section, such as gender, age, education, and, sometimes, industry. With longitudinal data, we are able to define groups of individuals based on their *history*, such as individuals with high versus low past average income and/or income growth rates.

An early literature on the sources of business cycle fluctuations debated the distinction between countercyclical dispersion versus a systematic factor structure. In a provocative

[3] They observed that if shocks are persistent and countercyclical, then, at a given age, cohorts that have lived through more recessions should have a larger cross-sectional dispersion of income than those who have not.

paper, Lilien (1982) showed that the dispersion of employment growth across sectors was time-varying in a way that was correlated with the unemployment rate. He interpreted this finding as evidence that sectoral shifts caused the cyclical fluctuations in the unemployment rate. Abraham and Katz (1986) challenged this conclusion by showing that a factor structure in which different sectors loaded differently onto an aggregate factor could generate the same correlation between dispersion and unemployment, even though the driving force was an aggregate shock.

Finally, in a related strand of literature, Bloom et al. (2011) fit an AR(1) process to firm-level total factor productivity (TFP) time series and allow a fixed aggregate shock and fixed firm effect. They find that the residual of the AR(1) has a larger cross-sectional dispersion during recessions. While the skewness also appears to be more negative, the difference is not statistically significant. In contrast to that paper, we do allow for a factor structure (loading factor on their aggregate shock) and allow the loading factor to vary with observables. Of course, we study individual labor income, whereas they focus on firm-level TFP, so the two sets of results are not necessarily inconsistent with each other.

2 The Data

We employ a unique, confidential, and very large panel dataset on earnings histories from the U.S. Social Security Administration records. For our baseline analysis, we draw a 10 percent random sample of US males—covering 1978 to 2010—directly from the Master Earnings File (MEF) of Social Security records.[4]

The Master Earnings File. The MEF is the main source of earnings data for the Social Security Administration and grows every year with the addition of new earnings information received directly from employers (Form W-2 for wage and salary workers).[5] The MEF includes data for every individual in the United States who has a Social Security number. The dataset contains basic demographic characteristics, such as date of birth, sex,

[4]Our focus on males is motivated by the fact that this group had a relatively stable employment rate and labor supply during this period. In contrast, female labor participation increased substantially during this period. Because our dataset contains only labor earnings but no hours information, including women in the analysis would have introduced an important confounding factor, which we wished to avoid.

[5]Although the MEF also contains income information for self-employed individuals, these data are top-coded at the taxable limit until 1994. Because of this, we do not use these data in this paper. In an earlier version, we conducted all the analysis using total labor income (and included self-employed individuals) and found no difference in our substantive conclusions.

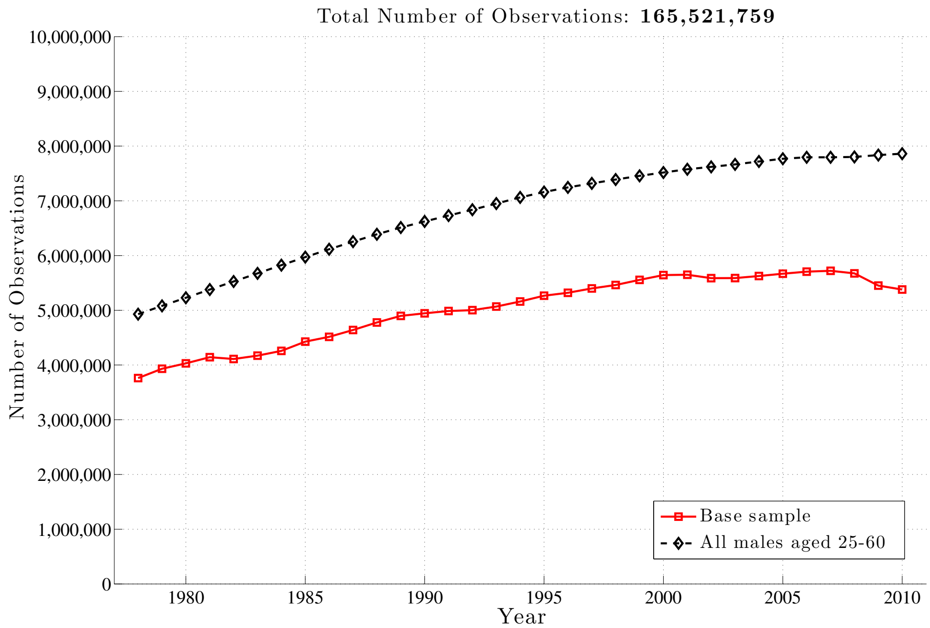


Figure 2: Number of Observations By Year

race, type of work (farm or non-farm, employment or self-employment), self-employment taxable income, and several other variables. Earnings data are uncapped (no top-coding) and include wages and salaries, bonuses, and exercised stock-options as reported on the W-2 form (Box 1).[6] For more information, see Panis et al. (2000) and Olsen and Hudson (2009). Finally, all nominal variables were converted into real ones using the PCE deflator with 2005 taken as the base year.

Creating the 10 Percent Sample. To construct a nationally representative panel of males, we proceed as follows. For 1978, a sample of 10% of US males are selected based on a fixed subset of digits of (a transformation of) the Social Security Number (SSN). Because these digits of the SSN are randomly assigned, this procedure easily allows randomization. For each subsequent year, new individuals are added to account for the newly issued SSNs in the United States; those individuals who are deceased are removed (from that year forward). This process yields a representative sample of 10% of US males every year.

Our *base sample* is determined as follows. For a statistic computed using data for

[6]Our earnings measure does not include deferred compensation, such as through 401(k), 403(b), and 457(b) plans, because information on these is not available consistently throughout the period. See Olsen and Hudson (2009) for details.

Table I: Summary Statistics of the Base Sample

Year	Median earnings ($)	Mean earnings	Change in log average earnings *per person* ×100	Change in log earnings, *Averaged over workers* ×100	Average age	Number of observations
	(in constant 2005 dollars)					
1978	39,488	47,938	—	—	39.4	3,640,877
1979	38,971	46,207	**–0.94**	1.10	39.3	3,797,417
1980	37,571	44,636	**–1.69**	**–3.11**	39.2	3,901,959
1981	37,908	44,785	0.08	2.00	39.2	4,011,200
1982	36,644	44,160	**–2.13**	**–3.26**	39.1	3,977,428
1983	36,431	44,276	**–0.35**	0.58	39.1	4,020,508
1984	36,847	45,760	1.21	6.53	38.9	4,090,461
1985	37,009	46,772	1.55	4.40	38.9	4,243,207
1986	37,100	48,062	1.06	3.68	38.9	4,311,235
1987	36,788	47,661	**–0.09**	1.78	38.9	4,423,615
1988	36,329	48,480	1.14	3.85	38.9	4,552,623
1989	35,614	46,572	**–1.50**	0.62	39.0	4,670,531
1990	35,207	46,262	**–0.66**	1.06	39.2	4,723,153
1991	34,451	45,766	**–0.78**	**–1.30**	39.3	4,768,475
1992	34,688	47,193	0.78	2.98	39.5	4,772,714
1993	34,660	47,471	0.11	3.33	39.7	4,829,933
1994	34,230	44,816	**–2.44**	2.00	39.8	4,904,776
1995	34,281	45,645	1.08	3.85	40.0	5,000,660
1996	34,863	46,730	0.94	3.88	40.2	5,045,831
1997	35,874	48,898	2.19	6.38	40.5	5,134,125
1998	37,351	51,348	2.20	7.06	40.7	5,198,954
1999	37,900	52,846	1.58	4.43	40.9	5,284,142
2000	38,525	55,030	2.06	4.35	41.1	5,366,942
2001	39,011	55,283	**–0.09**	1.93	41.3	5,376,439
2002	38,412	52,894	**–2.61**	**–2.36**	41.4	5,316,402
2003	38,187	53,145	0.01	0.55	41.6	5,303,052
2004	38,372	53,366	0.16	2.16	41.7	5,329,934
2005	38,196	53,586	0.23	2.12	41.8	5,359,877
2006	38,456	54,536	0.93	3.30	41.9	5,390,061
2007	38,526	55,322	0.67	2.44	41.9	5,405,122
2008	37,930	53,889	**–1.67**	**–1.03**	42.0	5,400,167
2009	36,984	51,946	**–3.86**	**–6.64**	42.1	5,238,303
2010	36,934	52,567	**–0.04**	1.25	42.1	5,161,313

Note: All statistics are computed for the base sample with the exception of column 3, which is computed as the change in log average earnings per (male) person.

(not necessarily consecutive) years $(t_1, t_2, .., t_n)$, an individual observation is included if the following three conditions are satisfied for all these years: the individual (i) is between the ages of 25 and 60, (ii) has annual wage/salary earnings that exceeds a time-varying minimum threshold, and (iii) is not self-employed (i.e., has self-employment income less than the same minimum threshold). This minimum, denoted $Y_{\min,t}$, is equal to one-half of the legal minimum wage times 520 hours (13 weeks at 40 hours per week). This condition is standard in the literature (see, e.g., Juhn et al. (1993) and Autor et al. (2008)) and allows us to focus on a subset of workers with a reasonably strong labor market attachment. Finally, the MEF dataset contains a few number of extremely high income observations each year. To avoid potential problems with these, we cap (top-code) observations above the 99.999th percentile.

Table I reports some key summary statistics for the base sample, such as average earnings, average age, change in measures of average earnings over time, and the number of observations in each year. Similarly, Figure 2 displays the number of individuals that satisfy these selection criteria, as well as the total number of individuals in each year. The sample starts with about 3.7 million individuals in 1978 and grows to about 5.4 million individuals by mid-2000s. Notice that the number of individuals in the sample does not follow population growth (black line marked with diamonds) one-for-one, because inclusion in the base sample also requires participating in the labor market in a given year (hence the slowdown in sample growth in the 2000s and the fall during the Great Recession).[7]

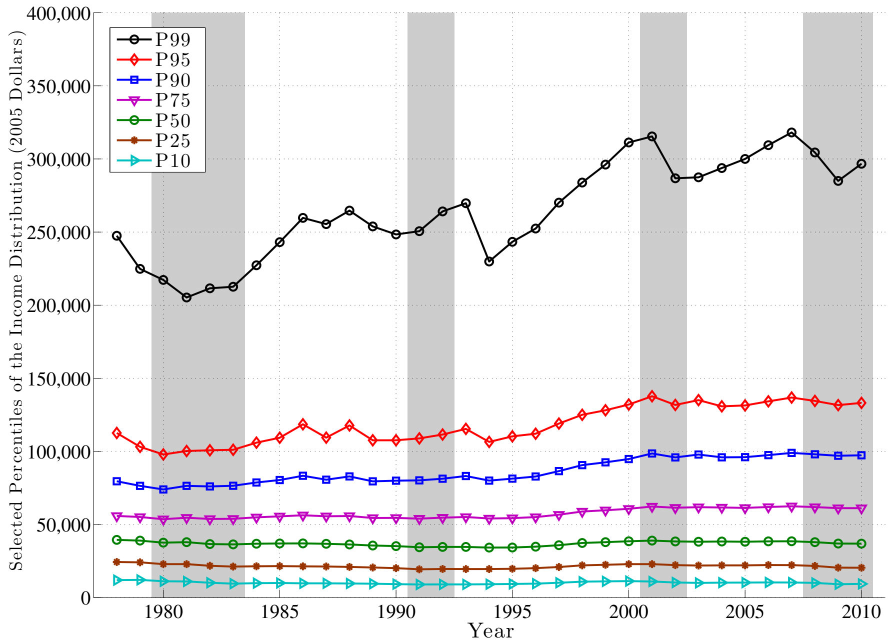
Figure 3 plots the levels of labor income that correspond to selected percentiles of the income distribution in each year. For example, the lowest income that qualifies a male worker in the top 10% (e.g., above the 90th percentile) has been steady at approximately $98,000 since year 2000. In 2011, a worker must be making more than $297,000 to be in the top 1%. This threshold was highest in 2007 when it reached $318,000.

Recessionary vs. Expansionary Episodes. Since the main focus of this analysis is on business cycle variation, we need to be clear about how a given year is classified as a recession or an expansion year. Because our labor earnings data are annual and recessions can start or end in any quarter during the year, this is not always straightforward.

The start date of a recession is determined as follows. If the National Bureau of Economic Research (NBER) peak of the previous expansion takes place in the first half of

[7] The appendix contains a more detailed comparison of inequality trends revealed by the base sample to those found in the Current Population Survey (CPS) data.

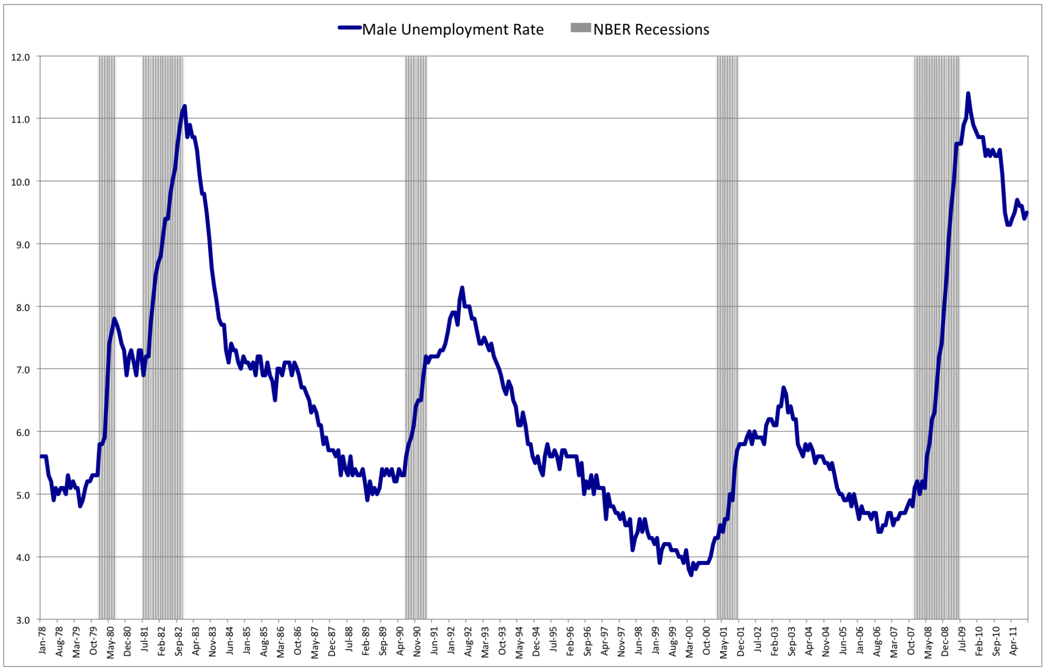
Figure 3: Selected Percentiles of Labor Earnings Distribution over Time



a given year, that year is classified as the first year of the new recession. If the peak is in the second half, the recession starts in the subsequent year.[8] The ending date of a recession is a bit more open to interpretation for our purposes, because the NBER "troughs" are often not followed by a rapid fall in unemployment rates and rise in individual wages. This can be seen in Figure 4. For example, whereas the NBER announced the start date of the expansion as March of 1991, the unemployment rate continued to rise and in fact peaked in the summer of 1992. Similarly, while the NBER trough was November 2001, the unemployment rate continued to rise and remained high until mid-2003. With these considerations in mind, we settled on the following dates for the last three recessions: 1991–92, 2001–02, and 2008–10. We opt to treat the 1980 to 1983 period as a single recession, given the extremely short duration of the intervening expansion, the anemic growth it brought, and the lack of a significant fall in the unemployment rate (Figure 4). Based on this classification, there are three expansions and four recessions during our sample period.

[8]In fact, two of the recessions we study start in the first quarter (1980 and 2001) and one starts in the fourth quarter (2007), so the classification of these is clear. Only one recession starts in the third quarter of 1990 and we shift the starting date to 1991 as per the rule described.

Figure 4: US Male Unemployment Rate, 1978–2011



3 Inequality Over the Business Cycle: First Look

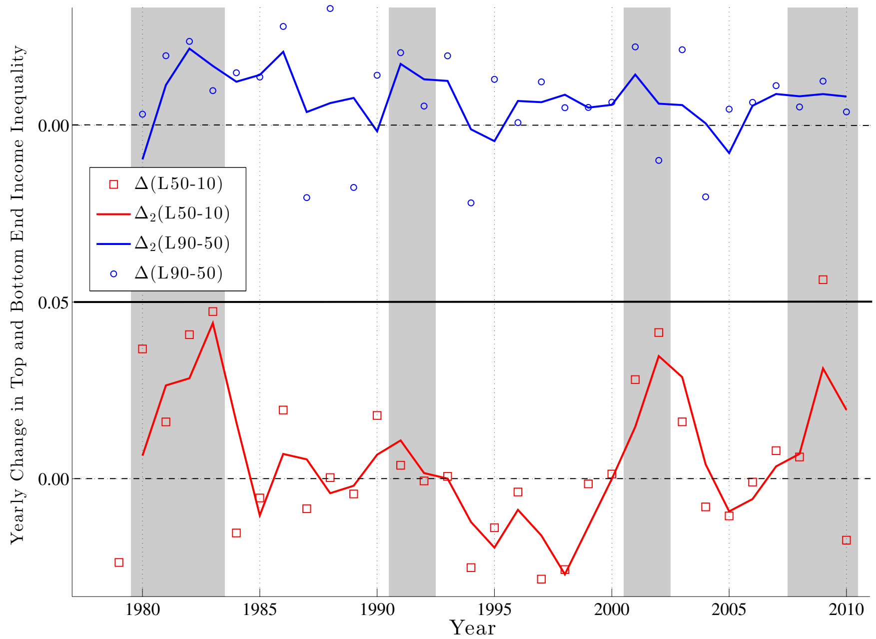
Before delving into the full-blown panel data analysis in the next section, we begin by studying some simple statistics to provide a bird's eye view of the business cycle patterns in income risk. The difference of the analysis in this section from previous work (e.g., Juhn et al. (1993), Autor et al. (2008), and others) is the panel dimension of our data. This allows us to document the moments of earnings *growth* at different horizons, which is informative about the short- and long-run dynamics of earnings.

3.1 Evolution of Inequality: Cross-Sectional Data

We begin by briefly establishing some facts on the evolution of cross-sectional inequality. This analysis does not require the panel dimension and it is presented here for completeness and comparison to the existing work.

It is useful to distinguish between the changes in top and bottom end inequality. To this end, we take the log differential between the 90th and 50th percentiles of the earnings distribution (hereafter L90-50), as well as the log differential between the 50th and 10th percentiles (L50-10), and plot the 1-year change in both series in Figure 5. To reduce

Figure 5: Change in Top and Bottom End Income Inequality



short-term mean reversion in inequality, the solid lines plot the 2-year difference in each inequality measure (divided by two), which is smoother. This differencing eliminates the secular trend and allows us to focus on the cyclical change in inequality.

First, notice the cyclical movement in the bottom-end inequality, rising in every one of the four recessions and falling (into the negative territory) subsequently. The increases in the 1980–83 and 2001–02 recessions are especially pronounced as is the fall during the 1990s. The change in the top-end inequality is also cyclical, rising during the 1980–83 and 1991–92 recessions. Compared with the bottom-end inequality though, L90-50 rises virtually throughout the period. Overall, the combination of these two pieces shows that overall inequality (L90-10) itself is countercyclical.

3.2 From Levels to Growth Rates

As noted earlier, the previous figure can be obtained using repeated cross-sections. Now, we turn to the properties of the income *growth* (or change) distribution, making use of the panel dimension of the MEF dataset.[9]

[9]Although some recent studies have also examined the time-series properties of income growth from panel data, th focused on secular trends rather than cyclical behavior (Dynan et al. (2007), Congressional

Properties of Income Shocks Over Time. It will be useful to distinguish between income growth over short and long horizons. To this end, in much of the following analysis, we examine 1-year and 5-year income growth rates and think of these as roughly corresponding to "transitory" and "persistent" income shocks. A more rigorous justification for this interpretation will be provided below.

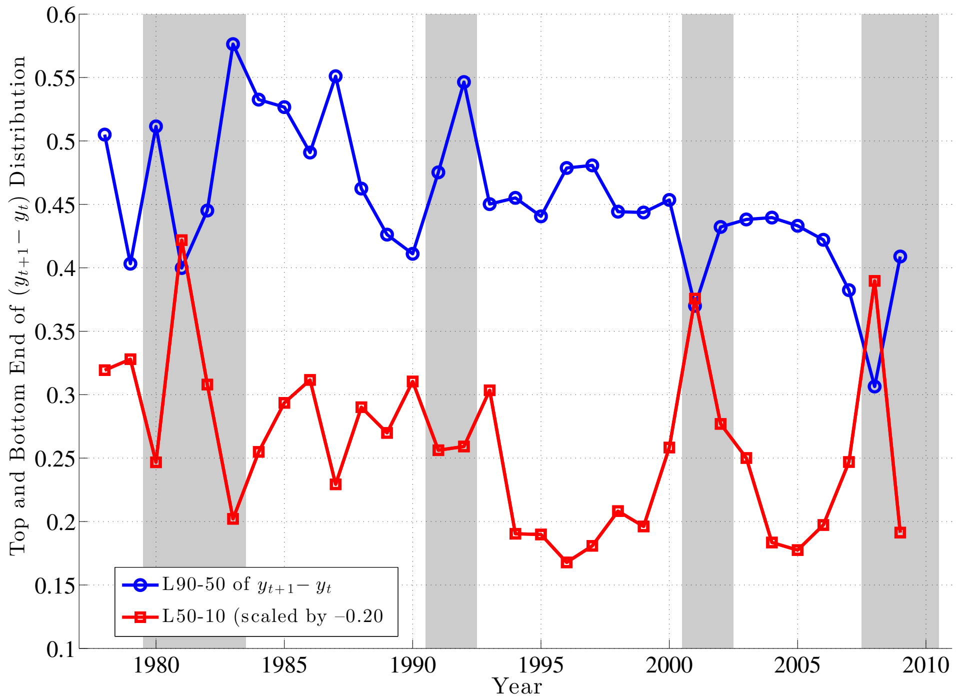
The top panel of Figure 6 plots the evolution of the top and bottom ends of the transitory income shock ($y_{t+1} - y_t$) distribution. The first obvious and important observation is that the top and bottom end of the shock distributions (L90-50 and L50-10, respectively) move in opposite directions over the cycle. In particular, L50-10 rises strongly during recessions, implying that there is an increased chance of *larger* downward movements during recessions. In contrast, the top end (L90-50) dips consistently in every recession, implying a smaller chance of large upward movements during recessions. In other words, relative to the median shock, the top end compresses, whereas the bottom end expands during recessions. Similarly, the bottom panel of Figure 6 plots the corresponding graph for persistent (5-year) income shocks. The striking co-movement of the L90-50 and L50-10 is clearly seen here (the cross-correlation of the two series is –0.67), arguably even more strongly than in the transitory shocks.

Several remarks are in order. First, the fact that L90-50 and L50-10 move in opposite directions implies that L90-10 (which measures overall dispersion of income shocks) changes little over the business cycle, because the fall in L90-50 partially cancels out the rise in L50-10. An alternative measure of shock dispersion—the standard deviation—is plotted in Figure 7 for both persistent and transitory shocks, which shows that dispersion does not increase much during recessions (notice the very small variation on the y-axis). Perhaps the only exception is the 2001–02 recession, during which time the transitory shock variance increases. In the coming sections, this point will be examined further and will be made more rigorously. This observation will provide one of the key conclusions of this paper, given how clearly it contradicts the commonly held belief that idiosyncratic income shock variances are strongly countercyclical (e.g., Storesletten et al. (2004)).

Second, looking at transitory shocks, L90-50 displays a clear downward trend during this time period. A fitted linear trend implies an 11 log points drop from 1979 to 2010. The interpretation is that the likelihood of large upward movements has become less likely

Budget Office (2008), Sabelhaus and Song (2009, 2010), Kopczuk et al. (2010), Solon and Shin (2011), and Moffitt and Gottschalk (2012)).

Transitory



Persistent

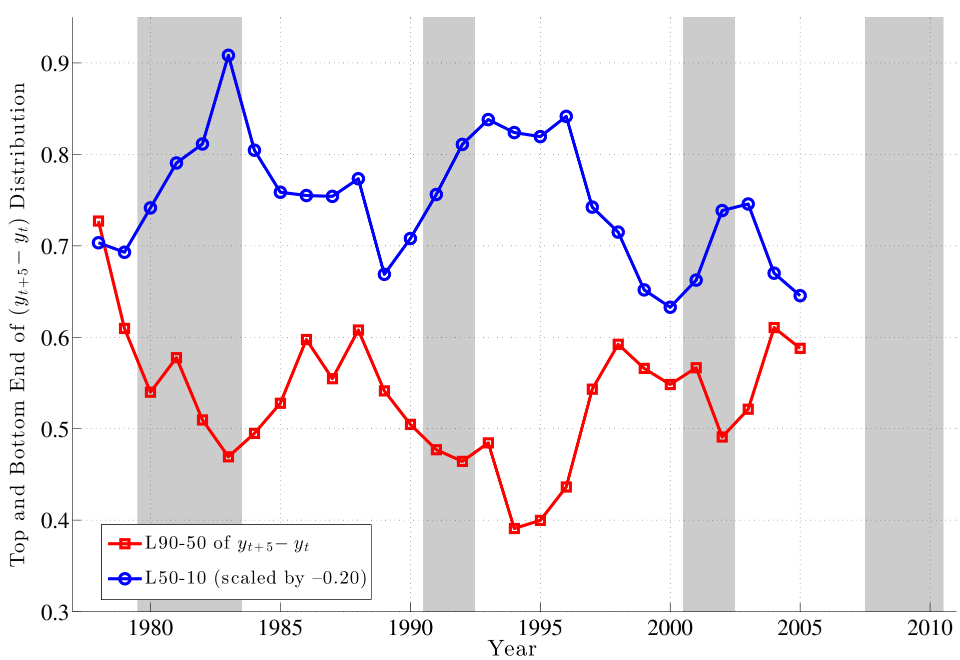


Figure 6: Top and Bottom Ends of Wage Income Change Distribution

during this period. We see a similar, but less pronounced, trend in the L50-10, which indicates that the likelihood of large falls has also become somewhat smaller. Overall though, both the L90-10 and the standard deviation of income growth (Figure 7) display a clear downward trend. Notice that this conclusion is in contrast to the conventional wisdom since the 1990s that income shock variances have generally risen since the 1980s (Moffitt and Gottschalk (1995)). However, it is consistent with a number of recent papers that use administrative data (e.g., Sabelhaus and Song (2010) and others).[10] In this paper, we will not dwell much on this trend, except when it is relevant for our analysis of the cyclical changes in income risk.

Third, and finally, the L90-50 of the *shock* distribution falls in recessions, even though the L90-50 of the *level* distribution rises, as we saw in Figure 5. Thus, top-end inequality rises in recessions, whereas the probability of individuals getting very positive income shocks becomes smaller. This is a distinction that will be examined further below.

The finding described above—that the top end of the shock distribution compresses during recessions, while at the same time the bottom end expands—suggests that one important cyclical change could be found in the skewness of shocks. Figure 8 plots the skewness of 1- and 5-year income growth distributions. As can be seen here, during recessions income growth becomes more left-skewed (negative skewness increases) and the magnitude is large. Below, we return to this point and sharpen it by conditioning income changes on narrowly defined groups of individuals.

4 Panel Analysis

The analysis so far intended to provide a general look at how income shocks vary over the business cycle. However, one can imagine that the properties of income shocks vary systematically with individual characteristics and heterogeneity: for example, young and old workers can face different income shock distributions than prime-age workers with more stable jobs. Similarly, workers at different parts of the income distribution could experience different types of income risks. The substantial number of individuals in our sample allows us to account for such variation without making any strong parametric assumptions.

[10] Moreover, Solon and Shin (2011) investigate the robustness of the finding from the PSID that the variance of 1-year income changes trends up over time. They show (figure 2 of their paper) that whether or not self-employment income is included in the measure of labor income makes a big difference to the trends, especially after 1990. In particular, focusing on wages and salaries implies no rise in variance, whereas including business income (farm, business, etc.) implies a 15 log point rise in the variance.

Figure 7: Standard Deviation of Transitory and Persistent Income Changes

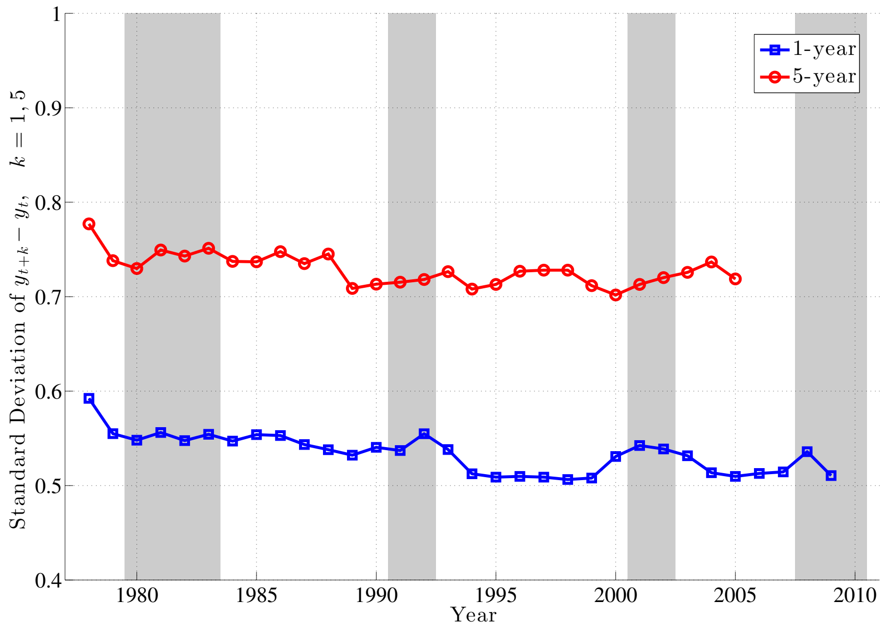
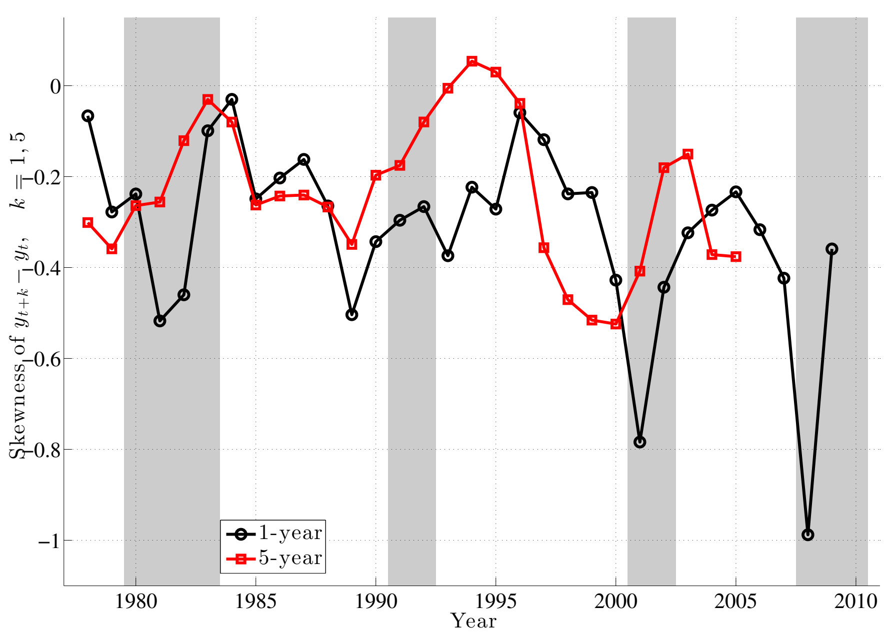


Figure 8: Skewness of Transitory and Persistent Income Changes



4.1 A Framework for Empirical Analysis

Let $\widetilde{y}_t^i$ denote individual i's log labor earnings in year t, and let $\mathbf{V}_{t-1}^i$ denote a vector of (possibly time-varying) individual characteristics that will be used to group individuals as of period $t-1$. Consider the following representation for the dynamics of individual income:

$$
\begin{aligned}
\widetilde{y}_t^i &= g(\theta, h_t) + \lambda_t + z_t^i + \varepsilon_t^i \\
z_t^i &= z_{t-1}^i + \eta_t^i,
\end{aligned}
\tag{1}
$$

where $g(\theta, h_t)$ is a flexible function of age (h) that captures lifecycle effects in log labor earnings, λ_t denotes the aggregate shock in the economy, and the transitory and persistent shocks are drawn from $\varepsilon_t^i \sim H(\varepsilon|\mathbf{V}_{t-1}^i, \lambda_t)$ and $\eta_t^i \sim G(\eta|\mathbf{V}_{t-1}^i, \lambda_t)$ with zero conditional mean. This specification allows the distributions of both types of shocks to vary across groups and over the business cycle.

Now define log labor earnings *net* of systematic lifecycle effects: $y_t^i \equiv \widetilde{y}_t^i - g(\theta, h_t)$. To study between-group and within-group variation over the business cycle, we take the kth difference of income in equation (1) and modify it to allow for a factor structure:

$$
\begin{aligned}
y_{t+k}^i - y_t^i = & \underbrace{f_1(\mathbf{V}_{t-1}^i)(\lambda_{t+k} - \lambda_t)}_{\text{factor structure}} \\
& + \underbrace{[(\eta_{t+k} + \eta_{t+k-1} + ... + \eta_{t+1})] + (\varepsilon_{t+k}^i - \varepsilon_t^i).}_{\text{stochastic component}}
\end{aligned}
\tag{2}
$$

The specification in (2) allows for three different types of business cycle effects. First, the factor structure—captured by the introduction of the function f_1—allows the conditional mean of income growth to vary systematically with the business cycle across different groups of workers. Second, *both* types of shocks have variances that can potentially vary with the business cycle in a way that is also different across groups of workers.

In our implementation, we will consider a vector $\mathbf{V}_{t-1}^i$ that includes three time-varying observable individual characteristics: age, past average income, and past income growth rate as of period $t-1$. An assumption that will be maintained in the analysis is that these characteristics vary slowly with time, so that $\mathbf{V}_t^i \approx \mathbf{V}_{t+k}^i$ for small k.

This formulation allows the effects of aggregate shocks to be transmitted differently to groups that differ in their labor market characteristics at the time a recession hits or an expansion gets underway. Of course, even individuals within these finely defined groups will likely experience different income growth rates during recessions and expansions, which will be captured by the permanent and transitory shocks above. These capture the within-group variation in shocks and we will also quantify the cyclical nature of such shocks.

Between-Group Variation in Shocks. It is useful to look at equation (2) again, this time taking the mean conditional on each group:

$$\mathbb{E}(y^i_{t+k} - y^i_t | \mathbf{V}^i_{t-1}) = f_1(\mathbf{V}^i_{t-1})(\lambda_{t+k} - \lambda_t) + \underbrace{\mathbb{E}(\eta_{t+k} + \eta_{t+k-1} + ... + \eta_{t+1} | \mathbf{V}^i_{t-1})}_{=0}$$
$$+ \underbrace{\mathbb{E}(\varepsilon^i_{t+k} - \varepsilon^i_t | \mathbf{V}^i_{t-1})}_{=0}.$$

Taking the means within each group eliminates both permanent and transitory shocks (since they average zero by assumption), yielding

$$\mathbb{E}(y^i_{t+k} - y^i_t | \mathbf{V}^i_{t-1}) = f_1(\mathbf{V}^i_{t-1})(\lambda_{t+k} - \lambda_t). \tag{3}$$

Equation (3) provides a simple expression for between-group variation in income growth. Between any two periods t and $t+k$, each group has a different loading factor $f_1(\mathbf{V}^i_{t-1})$ on the aggregate shock $(\lambda_{t+k} - \lambda_t)$. The key object of interest is f_1, whose shape will tell us about the factor structure of income changes over the business cycle.[11]

One drawback of this measure is that the left hand side of equation (3) can only be computed using individuals whose incomes are positive in year t and $t+k$ (so that log income is finite). This restriction could bias the results if the fraction of individuals who are dropped from the sample varies systematically with $\mathbf{V}^i_{t-1}$ and over the business cycle, which is quite possible. Later, in Section 5.2, we examine whether this bias could be important. As another way to address this concern, we also construct a slightly modified measure for the left hand side of (3). Basically, for a given group $\mathbf{V}^i_{t-1}$, we use all individuals to compute the average income of that group in t and $t+k$ and *then* take the logs of these

[11] To be more precise, f_1 should have a time subscript, since we will allow it to vary over time. However, to keep the notation clean, we will suppress the subscript in this paper.

averages to compute

$$f_2(\mathbf{V}^i_{t-1}) \equiv \log \mathbb{E}(Y^i_{t+k}|\mathbf{V}^i_{t-1}) - \log \mathbb{E}(Y^i_t|\mathbf{V}^i_{t-1}). \tag{4}$$

This measure now includes both the intensive margin and the extensive margin of income changes between two periods.[12] It will be our preferred measure in Section 6, although we will also compare it to the original measure f_1.

Within-Group Variation in Shocks. One focus of this analysis will be on simple measures of income shock volatility, conditional on individual characteristics. That is, fix a group of workers that have the same vector $\mathbf{V}^i_{t-1}$ at time t. Computing the within-group variance, we get:

$$\mathrm{var}(y^i_{t+k} - y^i_t|\mathbf{V}^i_{t-1}) = \underbrace{\left(\sum_{s=1}^{k} \mathrm{var}(\eta_{t+s}|\mathbf{V}^i_{t-1})\right)}_{k \text{ terms}} + \underbrace{(\mathrm{var}(\varepsilon_t|\mathbf{V}^i_{t-1}) + \mathrm{var}(\varepsilon_{t+k}|\mathbf{V}^i_{t-1}))}_{2 \text{ terms}}.$$

Two points should be noted from this formula. First, as we consider longer time differences (larger k), the variance reflects more of the permanent shocks as seen by the addition of the k innovation variances and given that there are always two variances from the transitory component regardless of k. For example, computing this variance over a five-year period that spans a recession (say 1979–84 or 1989–94) would allow us to measure how the variance of permanent shocks changes during recessions. It will also contain transitory variances, but for two years that are not part of a recession (1979 and 1984, for example). Second, looking at short-term variance, say $k = 1$, yields a formula that contains only one permanent shock variance and two transitory shock variances. So, as we increase the length of the period over which the variance is computed, the statistic shifts from being informative about transitory shock variances towards more persistent variation. In the analysis below, we will consider $k = 1$ and $k = 5$. And with some abuse of language, we will refer to them as transitory and persistent variances, respectively. Finally, each of these variances can also differ across groups of individuals. Below, we compute various group-specific statistics, including variances, to examine the nature of such variation.

[12] It is also more directly comparable to recent work that used cross-sectional data to quantify the cyclicality of incomes by constructing average incomes for synthetic groups, such as the top 1 percent, 10 percent, etc. (e.g., Parker and Vissing-Jørgensen (2010) and Saez (2012)).

4.2 Grouping Individuals into $\mathbf{V}^i_{t-1}$

Let t denote the generic time period that marks the beginning of a business cycle episode. We now describe how we group individuals based on their characteristics at time $t-1$. Each individual is identified by three characteristics that can be used to form groups. Not every characteristic will be used in the formation of groups in every experiment.[13]

1. Age. Individuals are divided into seven age groups. The first six groups are five-year wide (25–29, 30–34,..., 50–54) and the last one covers six years: 55–60.

2. Pre-Episode Average Income. A second dimension individuals differ along is their average income (and especially where they rank relative to others). For a given year t, we consider all individuals who were in the base sample (i) in year $t-1$ and (ii) in at least two more years between $t-5$ to $t-2$. For example, an individual who is 23 years old in $t-5$ (and hence is not in the base sample that year) will be included in the final sample for year t if he has earnings exceeding $Y_{\min}$ in every year between $t-3$ and $t-1$.

Furthermore, as noted above, we are interested in average earnings to see how a worker ranks in the income distribution relative to his peers. But even within the narrow age groups defined above, age variation can skew the rankings in favor of older workers. For example, between ages 25 and 29, average earnings grows by 35.4% in our sample, and between 30 and 34, they grow by 18.3%. So, unless this lifecycle component is accounted for, a 29-year-old worker in the first age group would appear in a higher income percentile than the same worker when he was 25. This would confound age and income differences, which ought to be avoided.

To adjust for this, we proceed as follows. First, using all income observations from our base sample from 1978 to 2010, we run a pooled regression of log raw earnings ($\widetilde{y}^i_{t,h}$) on age and cohort dummies (without a constant) to characterize the age profile of log earnings. We then scale the age dummies (denoted with d_h) so as to match the average log earnings

[13] One observable characteristic that has often been used in the literature on wage inequality is educational attainment. The MEF does not contain any information on education, so we cannot use it in our analysis. Having said that, papers that investigated the cyclicality of the skill premium (i.e., between-education-group differences) found only a modest correlation with the business cycle. For example, both Castro and Coen-Pirani (2008, Table 2) and Balleer and van Rens (2011, Table 1) report a correlation of skill premium with GDP and productivity close to zero (ranging from –0.15 to 0.20). Therefore, this omission is probably not an important shortcoming of our analysis.

of 25-year-old individuals used in the regression. Using these age dummies, we compute the average earnings between years $t-5$ and $t-1$ for the *average* worker of age h in year t . Then for a given worker i of age h in year t, we first average his earnings from $t-5$ to $t-1$ (and set earnings below $Y_{\min,t}$ equal to the threshold) and then normalize it by the population average computed using the age dummies. This 5-year average (normalized) earnings is denoted with $\overline{Y}_{t-1}^{i} \equiv (\sum_{s=1}^{5} e^{\widetilde{y}_{t-s}^{i}})/(\sum_{s=1}^{5} e^{d_{h-s}})$.[14]

3. Pre-Episode Income Growth. A third dimension is (recent) income growth. This could be an indicator of individuals whose careers are on the rise, as opposed to being stagnant, even after controlling for average income as done above. For this purpose, we compute $\Delta_5(y_{t-1}) \equiv (y_{t-1} - y_{t-s})\,/(s-1))$, where s is the earliest year after $t-6$ in which the individual has income above the threshold.[15]

5 Within-Group Shocks

We begin with the cyclicality of idiosyncratic shocks, as measured by within-group variation in income growth rates. An important question is whether or not idiosyncratic shocks have countercyclical variances. By idiosyncratic shocks, we mean income changes experienced by individuals who are ex ante (immediately prior to the shock) very similar based on their observable characteristics.

To study this question, we first condition on age and $\overline{Y}_{t-1}^{i}$. A graphical construct that will be used to study within-group variation is obtained by plotting the quantiles of $\overline{Y}_{t-1}^{i}$ for a given age group on the x-axis against the entire distribution of *future* income growth rates for that quantile on the y-axis. We denote these conditional distributions with $\mathbb{F}(y_{t+k} - y_t|\overline{Y}_{t-1})$.

Figure 9 is the first use of this graphical construct and contains a lot of information that will be referred to in the rest of this section. The top panel displays P90, P50 (median),

[14] We have also tried an alternative measure of average earnings that weighs each observation inversely with their distance from year $t-1$, to further group together individuals whose incomes were similar at more recent dates. To this end, for a given t, define the weight $w_{t,s}^{i} = (6-s)1\{\widetilde{Y}_{t-s}^{i} \geq Y_{\min,t}\}$, which is zero for ineligible observations and declines with s otherwise. Using these weights to construct the average earnings made almost no change to the results reported here.

[15] Information on detailed SIC codes is available, so, in principle, we could further classify individuals based on their 3- or 4-digit industry. Our preliminary results indicated that little was gained by this step, so we did not pursue this approach.

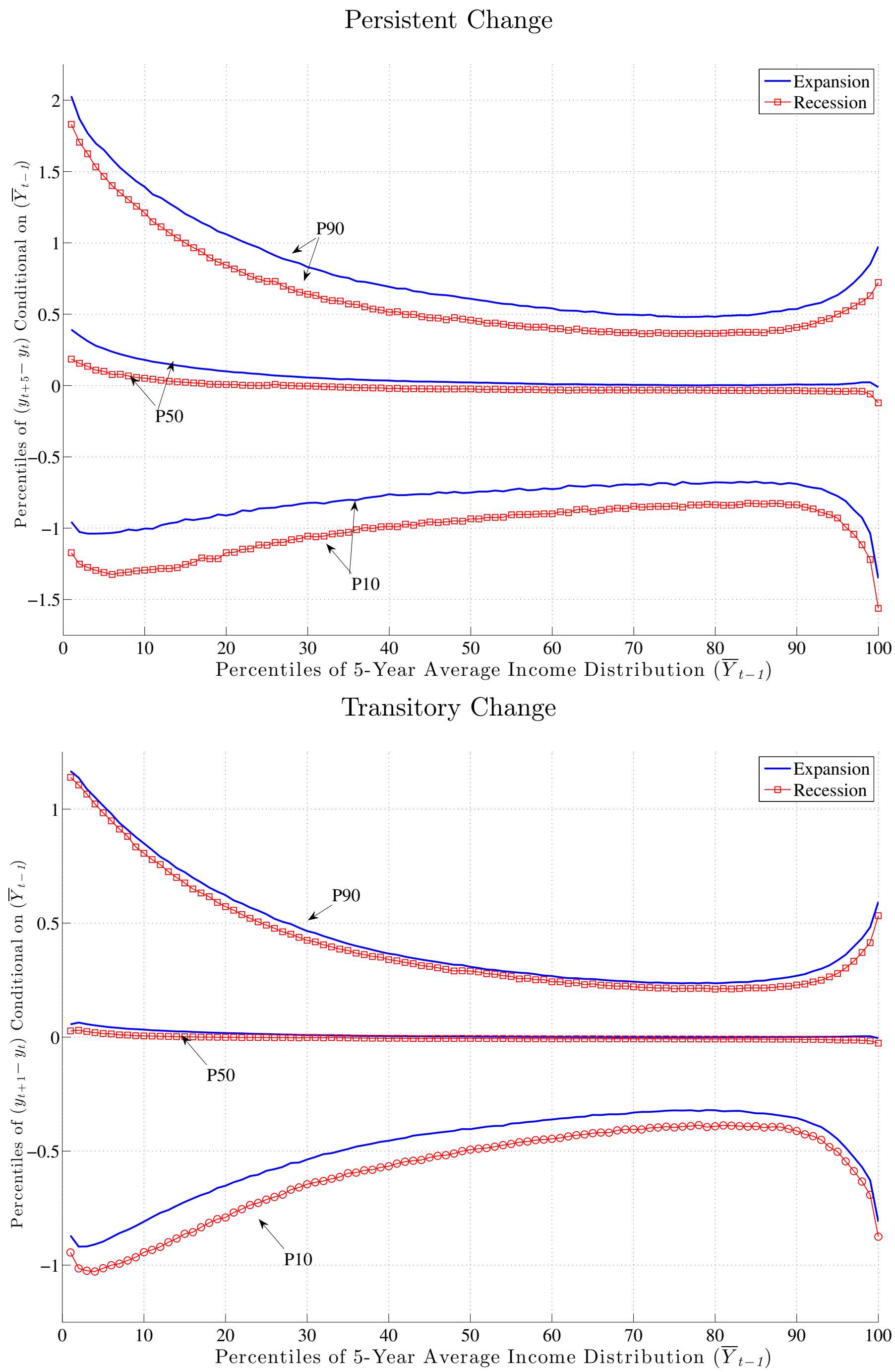


Figure 9: Percentiles of Income Growth Distribution: Recession vs Expansion

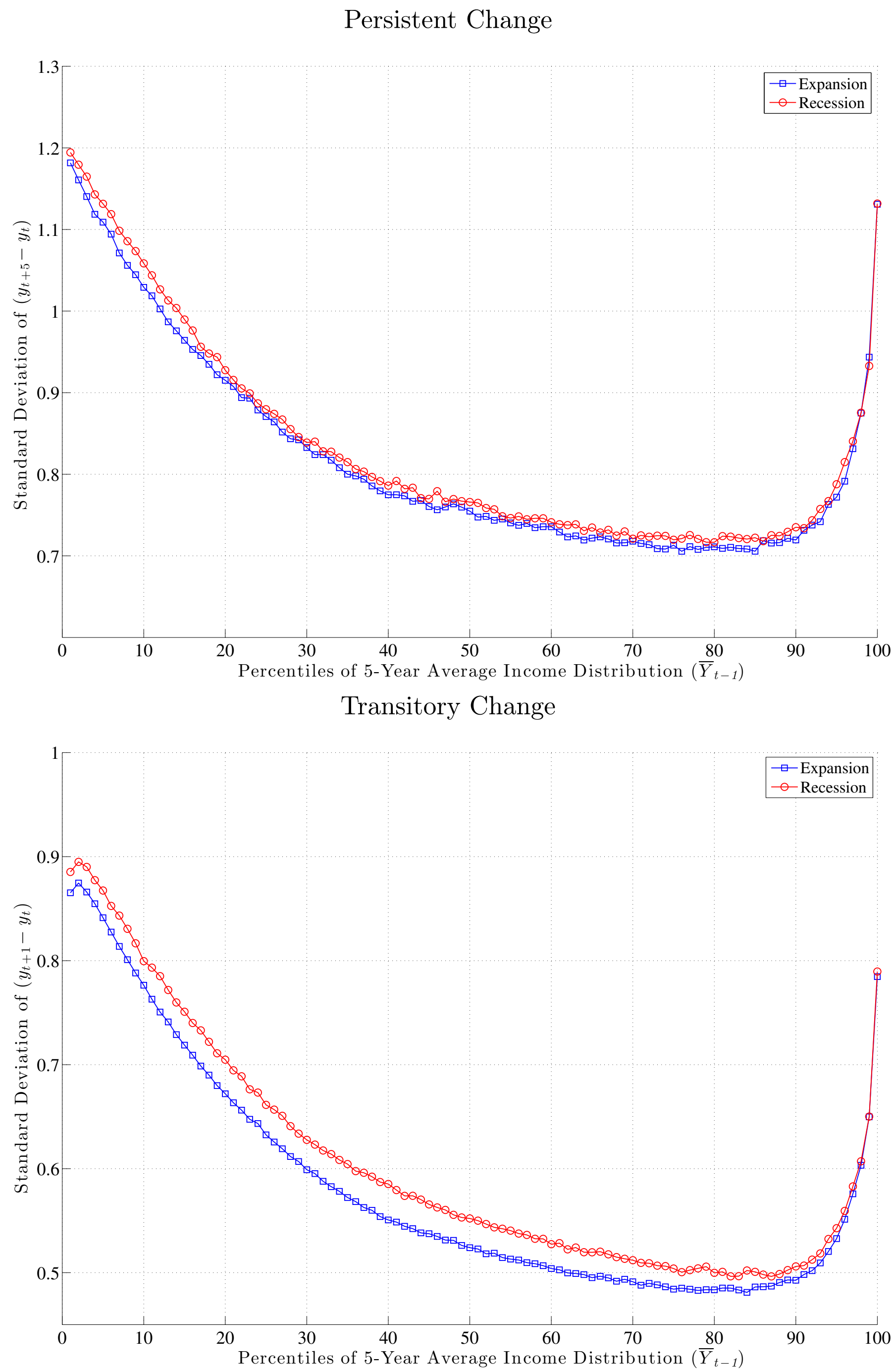


Figure 10: Dispersion of Transitory and Persistent Income Changes

and P10 of the distribution of long-run changes, $y^i_{t+5} - y^i_t$, (on the y-axis) for each percentile of $\overline{Y}^i_{t-1}$ (on the x-axis). To compare recessions and expansions, we averaged each one of these percentiles, separately over the four recessions (lines marked with "circles") and three expansions (solid blue lines) during our sample period. Similarly, because these figures look similar across age groups, we also averaged across the age groups to save space.[16]

We begin by commenting on the variation in these percentiles as we move to the right along the x-axis. Interestingly, the following pattern holds in both recessions and expansions: At any point in time, individuals with the lowest levels of past average income face the largest dispersion of income shocks ($y_{t+k} - y_t$) looking forward. That is, L90-10 is widest for these individuals and falls in a very smooth fashion moving to the right. Indeed, workers who are between the 70th and 90th percentiles of the $\overline{Y}^i_{t-1}$ distribution face the smallest dispersion of shocks looking ahead. As we continue moving to the right (into the top 10%), the shock distribution widens again and does so monotonically as a function of past income. Notice that the P10 and P90 of the $y^i_{t+5} - y^i_t$ distribution look like the mirror image of each other relative to the median, so the variation in L90-10 as we move to the right is driven by similar variations in P90 and P10 individually.

Turning to the bottom panel, the same graph is plotted now for $y_{t+1} - y_t$ (transitory shocks).[17] Precisely, the same qualitative features are seen here with low- and high-income individuals facing a wider dispersion of persistent shocks than those in the "safer" zones—between the 70th and 90th percentiles. Of course, the scales of both graphs are different: the overall dispersion of persistent shocks is much larger than that of transitory shocks, which is to be expected.

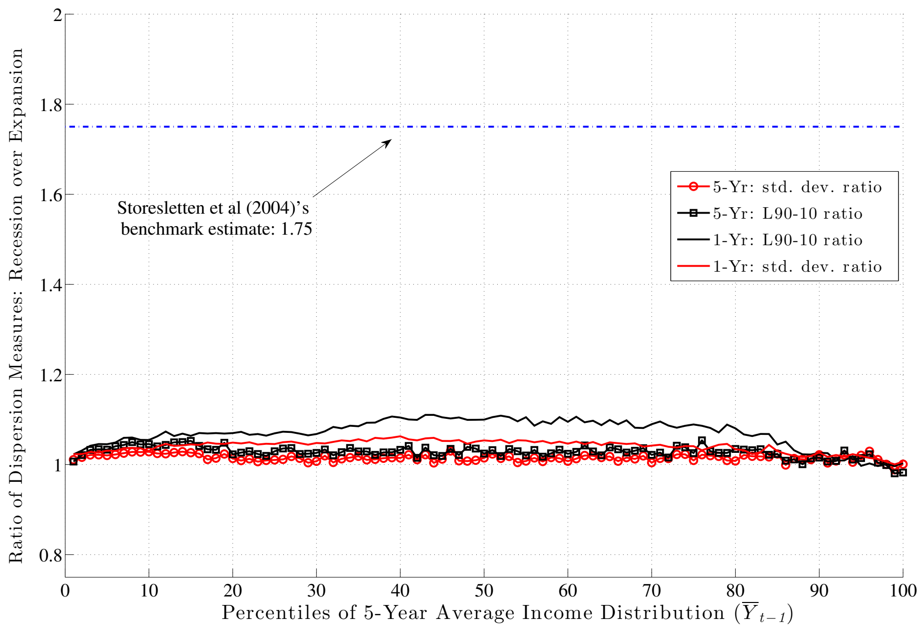
To sum up, both graphs reveal extremely strong and systematic variation in the dispersion of persistent and transitory income shocks across individuals with different past income levels.[18]

[16] For 5-year changes, recession years can be defined in a number of ways since many 5-year periods cover a given recession. We have experimented with different choices and found them to make little difference to the substantive conclusions drawn here. The reported results are for a simple definition that includes one 5-year change for each recession that starts one year before the recession begins. Specifically, the recession graph averages over four 5-year periods starting in $t = 1979$, 1989, 1999, and 2005 (since this is the latest possible 5-year change covering the Great Recession). Expansions average over all 5-year changes that do not coincide with a recession year. That is, periods starting in $t = 1983$, 1984, 1993, 1994, and 2002.

[17] For one-year changes, recession years are those with $t = 1980$, 1981, 1982, 1990, 1991, 2000, 2001, 2007, 2008, and 2009. The remaining years are considered as expansion years.

[18] This finding clearly contradicts one of the standard assumptions in the income dynamics literature—that the variance of income shocks does not depend on the current or past level of income. We explore these implications in a separate ongoing project.

Figure 11: Ratio of Shock Dispersion Measures: Recession over Expansion



Now we turn to two key questions of interest. First, what happens to idiosyncratic shocks in recessions? Are idiosyncratic shock variances countercyclical? And second, how does any potential *change* in the distribution of idiosyncratic shocks *vary* across income levels (i.e., the cross-partial derivative)? In other words, do we see the shock distribution of individuals in different income levels being affected differently by recessions?

Are Shock Variances Countercyclical? The existing literature has largely focused on the cyclicality of persistent shocks, so this is where we also start (top panel of Figure 9). First, both P90 and P10 shift downward by similar amounts from expansion to recession. (As can be anticipated from this, the gap (L90-10) changes very little over the business cycle, as we shall see momentarily.) Furthermore, following the same steps as the one used to construct these graphs, one can also compute the standard deviation of $y_{t+5} - y_t$ conditional on $\overline{Y}^i_{t-1}$ during recessions and expansions, which is plotted in the top panel of Figure 10. As seen here, the two graphs (for expansions and recessions) virtually overlap, over the entire range of pre-episode income levels. For transitory shocks (bottom panel), there is more of a gap, but the two lines are still quite close to each other.

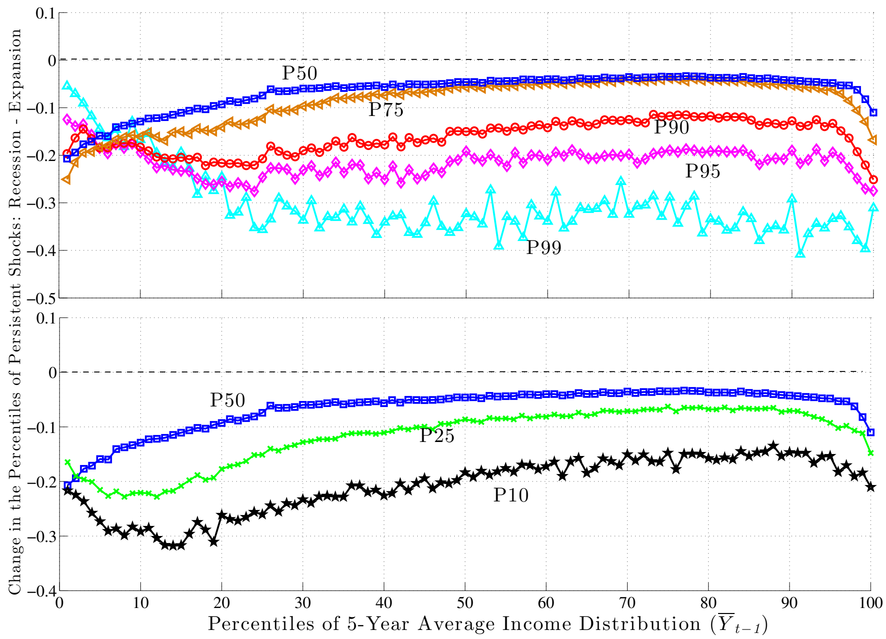
To make the measurement of countercyclicality more precise, Figure 11 plots the ratios of (i) standard deviations and (ii) L90-10s for recessions over expansions. For persistent shocks (lines marked with circles and squares), both the standard deviation and L90-10 measures are only about 2% higher in recessions than in expansions. In other words, while we find *some* evidence of counter-cyclicality, the magnitude is minuscule. For comparison, Storesletten et al. (2004) used indirect methods to estimate a standard deviation of 0.13 for innovations into a persistent AR(1) process during expansions and 0.21 for recessions. The ratio is 1.75 (marked on the figure for comparison) compared with 1.02 we find in this paper. The figure also plots the same two ratios (L90-10 and standard deviations) for transitory shocks. Here we see a bit more movement relative to persistent shocks: the standard deviation is higher by about 4 percent (averaged across the x-axis) and L90-10 is higher by about 6 percent. These findings suggest that to the extent that recessions involve larger dispersion of shocks, these are to be found in short-term shocks without much long-term effects. Having said that, these numbers are still very small compared with the values typically used in the literature.

A second question that was raised above was whether recessions affect the distribution of shocks differently in different parts of the income distribution. It is probably evident by now that the answer is, perhaps surprisingly, "no." Even though shocks, on average, have systematically very different dispersions in different parts of the distribution, the cyclical *change* in this dispersion is remarkably flat across the quantiles of $\overline{Y}_{t-1}^{i}$ (i.e., the cross-partial derivative is approximately zero). This is seen in the three figures just discussed, but is most apparent in Figure 11, where the ratios are quite flat, especially for persistent shocks. Therefore, we conclude that when it comes to the variance of persistent shocks, different income groups are affected similarly by business cycle fluctuations.

5.1 Countercyclical (Left-)Skewness

The obvious question now is: Do recessions have *any* effect on income shocks? The answer is yes, which could already be anticipated from Figure 9, by noting that while P90 and P10 move down together during recessions, P50 (the median of the shock distribution) remains extremely stable and moves down by only a little. This has important implications: L90-50 gets compressed during recessions, whereas L50-10 expands. In other words, for every income level $\overline{Y}_{t-1}^{i}$, when individuals look ahead during a recession, they see a much smaller chance of upward movements (relative to an expansion), but a much higher chance of large downward movements. In fact, this result is not specific to using P90 or P10, but is

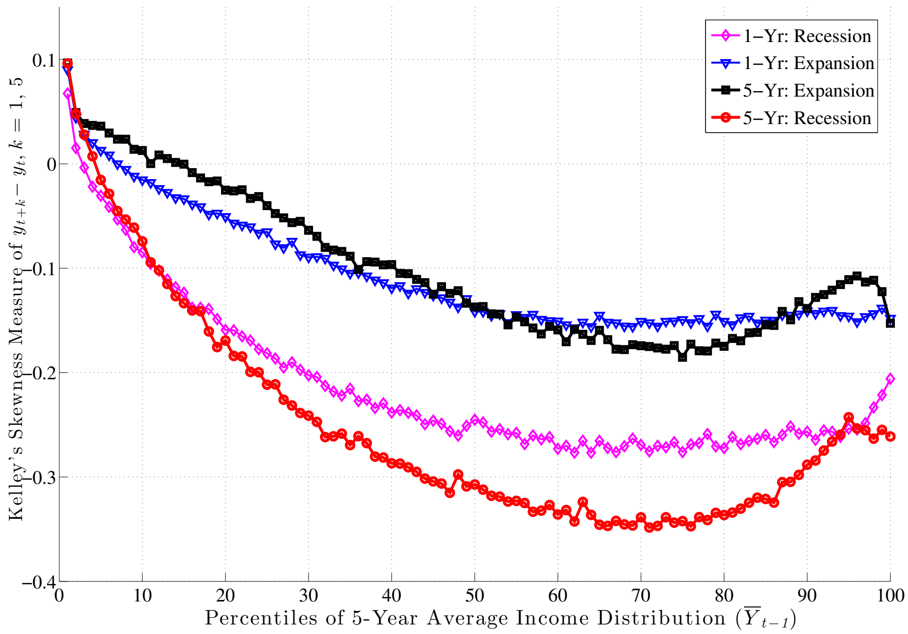
Figure 12: Change in the Percentiles of the Persistent Shock Distribution: Recessions Minus Expansions



pervasive across the entire distribution of future income growth rates. This can be seen in Figure 12, which plots the change in selected percentiles above (and including) the median from an expansion to a recession (top panel). The bottom panel shows selected percentiles below the median. Starting from the top, and focusing on the middle part of the x-axis, we see that P99 falls by about 30 log points from an expansion to a recession, wheres P95 falls by 20, P90 falls by 15, P75 falls by 6, and P50 falls by 5 log points, respectively. As a result, the entire upper half of the shock distribution gets squeezed towards the median. In other words, the half of the population who experience income change above the median now experience ever smaller upward moves during recessions. Turning to the bottom panel, we see the opposite pattern: P50 falls by 5 log points, whereas P25 falls by 9, and P10 falls by 20 log points respectively. Consequently, the bottom half of the shock distribution now expands, with "bad luck" meaning even "worse luck" during recessions.

From this analysis, a couple of conclusions can be drawn. First, idiosyncratic risk *is* countercyclical. However, this does not happen by a widening of the entire distribution (e.g., variance rising), but rather a shift towards a more left-skewed shock distribution. Although this is evident from the top end compressing and bottom end expanding, one

Figure 13: Skewness of Transitory and Persistent Income Changes



can compute measures of skewness to document this. With higher order moments, one has to be careful about extreme observations. These are not likely to be outliers as with survey data, but even if they are genuine observations, we may want to be careful that a few observations do not affect the overall skewness measure. For this purpose, we use "Kelley's measure" of skewness, which relies on the quantiles of the distribution and is robust to extreme observations (Figure 13). It is also very straightforward to interpret as we shall see in a moment. It is computed as the relative difference between the upper and lower tail inequalities: (L90-50 – L50-10)/L90-10. A negative number indicates that the lower tail is larger than the upper tail, and vice versa for a positive number.

Turning to Figure 13, first, notice that individuals in higher income percentiles face a more negatively skewed shock distribution, consistent with the idea that the higher an individual's income is, the more it has room to fall. Second, and more importantly, this negative skewness increases during recessions both for transitory and persistent shocks. For example, for individuals at the median of past income distribution, Kelley's measure for persistent shocks averages –0.14 during expansions. This number has a simple interpretation. It says that the dispersion of shocks above P50 accounts for 43% of overall L90-10 dispersion. Similarly, dispersion below P50 accounts for the remaining 57% (hence

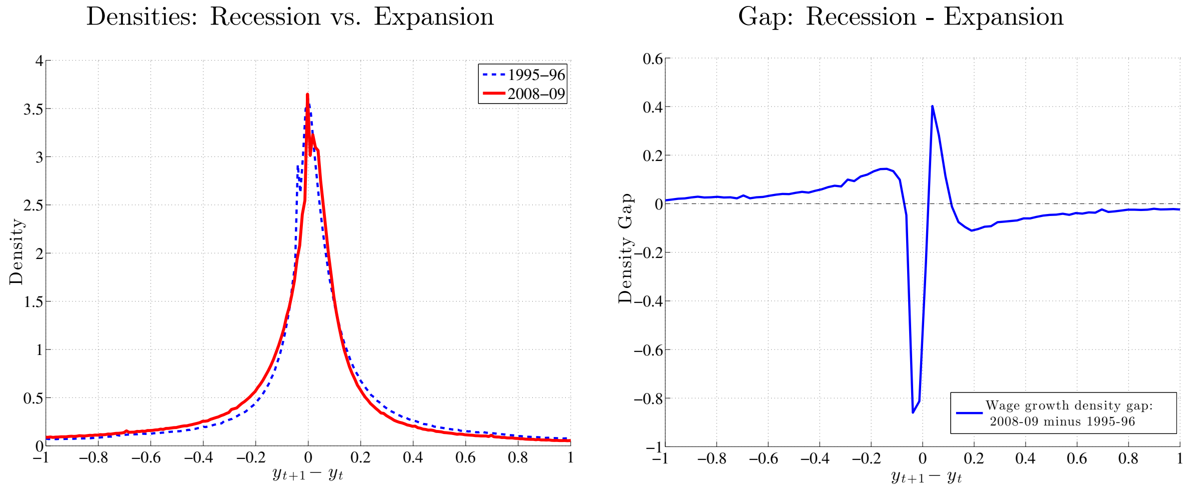
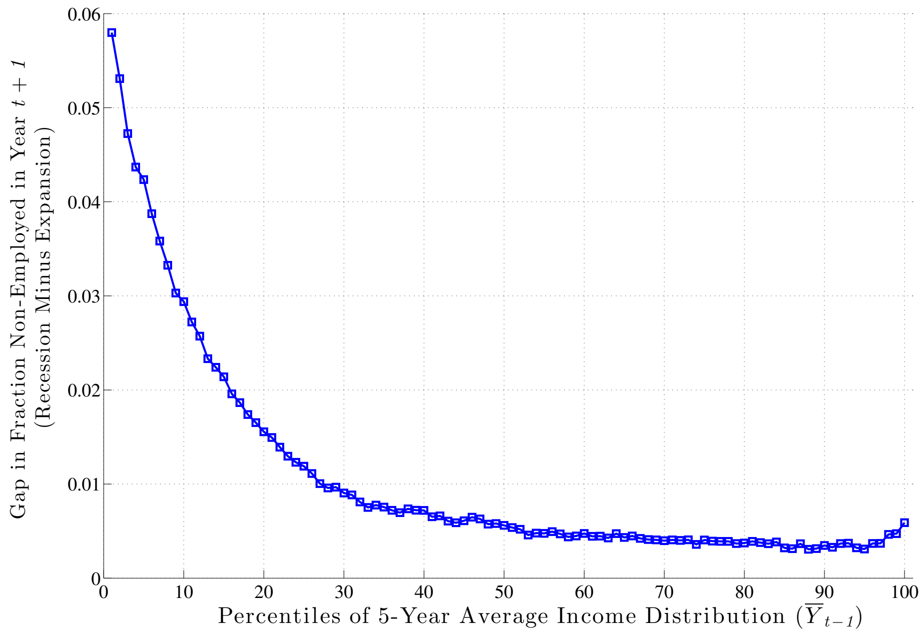


Figure 14: Histogram of Δy_t: US Data 1995–96 vs. 2008–09

$(43\% - 57\%)/100\% = -0.14)$ of L90-10. In recessions, however, this figure falls to –0.30, indicating that L90-50 accounts for 35% of L90-10 and the remaining 65% is due to L50-10. This is a substantial shift in the shape of the persistent shock distribution over the business cycle. The change in the skewness of transitory shocks is similar, if somewhat less pronounced. It goes from –0.14 down to –0.25 at the median. As seen in the figure, the increased left-skewness during recessions is pervasive—it takes place across the entire income distribution with similar magnitudes (with the exception of very low-income individuals).

To understand how different this conclusion is from a simple countercyclical variance formulation, recall Figure 1, which plots the densities of two Normal random variables: one with zero mean and a standard deviation of 0.13 (expansion) and a second one with a mean of –0.03 and a standard deviation of 0.21 (recession; both numbers from Storesletten et al. (2004)). As seen here, the substantial increase in variance and small fall in the mean implies that many individuals will receive larger positive shocks in recessions than in expansions under this formulation. For comparison, the left panel of Figure 14 plots the empirical densities of income growth from the US data, comparing the 1995–96 period to the worst year of the Great Recession (2008–09). To highlight how the density changes, the right panel plots the difference between the two densities. As seen here, the probability mass on the right side shifts from large positive shocks to more modest ones; on the left side, it shifts from small negative shocks to even larger negative ones. Thus, recessions are times when it becomes less likely for anybody to experience large upward income changes, whereas the risk of falling off the income ladder becomes significantly higher.

Figure 15: Probability of Full Year Non-Employment: Recessions Minus Expansions



Interestingly, in one of the earliest papers on cyclical changes in income risk, Mankiw (1986) postulated that in recessions, a fraction λ of individuals all draw the same negative shock which adds up to $-\mu$. So, ex ante, each person views a recession as a state where, with probability λ, their income will drop by $-\mu/\lambda$. Thus, negative shocks are concentrated among a subset of individuals in recessions. This structure induces a left-skewness of the same sort discovered in our analysis here, unlike the countercyclical variance structure proposed by Constantinides and Duffie (1996) and others.[19]

5.2 Extensive Margin of Business Cycle Risk

The analysis so far in this section has been based on changes in *log* income, which required us to drop individuals with zero income in a given year (t or $t+k$). In this sense, these results characterized the income risk of individuals who spend at least part of both years in the labor market (i.e., excludes the long-term unemployed). On average, an excluded

[19]An interesting question is whether there are cyclical changes in moments beyond the third (skewness), for example, in the fourth moment—the kurtosis. Although the answer is yes—the kurtosis is lower in recessions compared with expansions—the differences are quite modest. We omit those results for brevity. They are available upon request from the authors.

individual will have more than 18 months of zero income (so that he appears with zero income during an entire *calendar* year). Given that long-term unemployment is not very common in the United States, this probability is not very high. However, it is still a real risk whose likelihood probably changes over the business cycle.

Here, we provide a measure of this risk using the same graphical construct as above. For each percentile of past average income, Figure 15 plots the *rise* in the risk of full year *non*-employment (i.e., having income less than $Y_{\min}$) during recessions relative to expansions as a function of past average income. Notice that, except for the lowest 30 percentiles, the rise is less than 1% of workers in that quantile and in fact averages close to 0.5%. It is safe to conjecture that such a small change in the composition of individuals over the business cycle is not likely to materially affect the results presented so far, which largely relied on the 90th and 10th percentiles of the $y_{t+k} - y_t$ distributions.

Second, to the extent that this compositional change does affect the results (for the lower 30 percentiles for example), which direction would the bias go? It is clear from Figure 15 that more individuals with zero income are dropped during recessions. If these individuals were somehow included, these observations would register as large negative income changes, further amplifying the countercyclical skewness documented above. Having said that, our conjecture is that this effect would be modest as long as one uses a robust statistic to measure dispersion and skewness.

6 Between-Group (Systematic) Business Cycle Risk

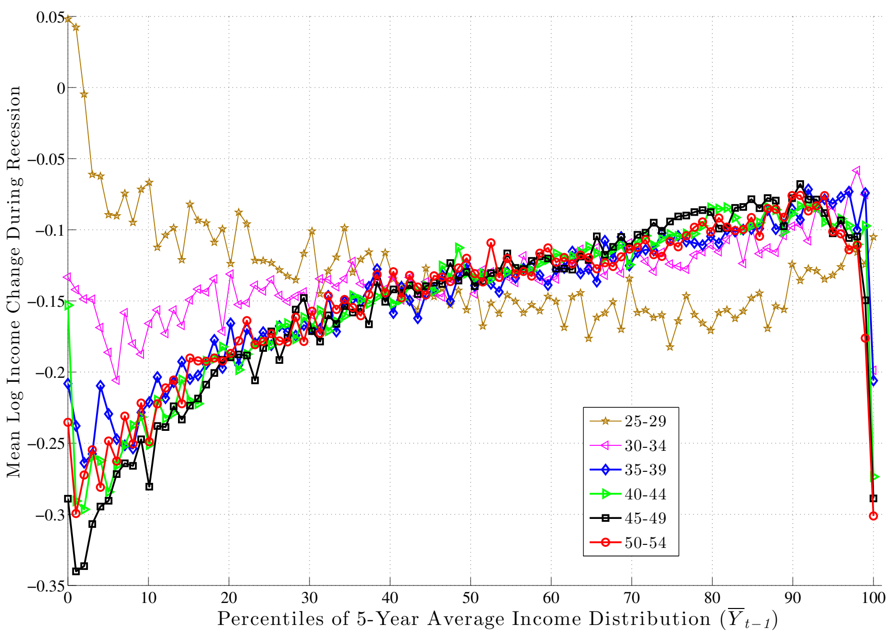
We now turn to the factor, or between-group, component of income risk. The goal here is to understand the extent to which income growth during a business cycle episode can be predicted by available observable characteristics prior to the episode.

6.1 Variation Across $\overline{Y}^i_{t-1}$ Quantiles

6.1.1 Recessions

We begin with a simple and useful benchmark. Consider a specific episode such as the Great Recession, with an aggregate shock of $(\lambda_{2010} - \lambda_{2007})$, which is a constant scalar. Suppose further that $\mathbf{V}^i_{t-1}$ is a grouping based on age and pre-episode average income: so

Figure 16: Growth in Log Average Income during the Great Recession (2007–10)



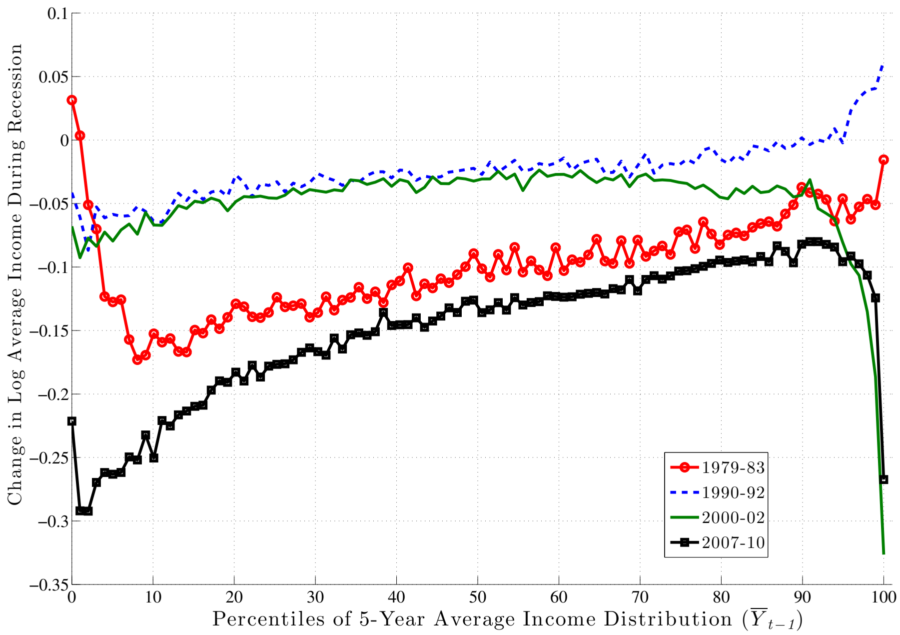
for each age group, we rank individuals by $\overline{Y}^i_{2006}$ and group them into 100 percentile bins. For each percentile group, we compute the log mean income in 2007 and 2010 (including zero incomes when applicable) and compute f_2 as defined in equation (4). Now, if we estimate f_2 to be a constant (flat) function, this would tell us that, during the Great Recession, there was no systematic variation in income growth across groups of individuals with different $\overline{Y}^i_{t-1}$. If f_2 is upward sloping, that would indicate a factor structure in favor of individuals with high pre-episode income.[20]

Figure 16 plots the estimated function f_2 for the 2007–10 period, for each of the six age groups defined above.[21] Several results can be seen here. First, f_2 looks quite different for the youngest two groups compared to the rest—but is virtually identical for all four age groups between ages 35 and 54. Motivated by this finding (which is also true in other periods we study), from now on we combine individuals aged 35 to 54 into one group and refer to them as "prime-age males." For brevity, we also combine the first two age groups into one and refer to them as "young workers" (ages 25 to 34).

[20]These statements implicitly assume that shocks *do not* exhibiting mean reversion. If z_t is a mean-reverting process, f_2 will be downward sloping in $\overline{Y}_{2006}$ in the absence of any factor structure. So any sign of upward slope (overcoming this potential downward bias) would be an indication of a factor structure.

[21]The 55–60 year-old age group is excluded because many of these individuals are older than 60 by 2010.

Figure 17: Growth in Log Average Income during Recessions, Prime-age (35–54) Males



Prime-age Males. Second, and substantively more important, for prime-age males the f_2 function is upward sloping and rises almost linearly up to about the 90th percentile (with a mild concavity in the lowest 20 percentiles), subsequently tapers off, and then falls of a cliff at the very top income percentile. To see this pattern more clearly, Figure 17 plots the aggregated f_2 function for prime-age males for all four recessions during our sample period. For the Great Recession (black line with squares), f_2 is upward sloping in an almost linear fashion and rises by about 17 log points between the 10th and 90th percentiles. So, workers with a pre-recession average income in the 10th percentile saw their income decline by about 25 log points during the recession, compared with a decline of only 8 log points for workers in the 90th percentile.[22] Clearly, this factor structure leads to a significant widening of income inequality over much of the distribution. However, this good fortune of high-income individuals does not extend to the very top: f_2 first flattens beyond the 90th percentile and then for the top 1%, it actually falls very steeply. Specifically, those in the top 1% experienced an average loss of 27 log points compared with 12.5 log points for those in the second highest percentile. One conclusion we draw from this analysis is

[22]Recall that the income measure used in these computations, y_t, is net of income growth due to lifecycle effects as explained in Section 4. This adjustment shifts the intercept of the f_2 function downward, which should be considered when interpreting the reported income growth figures.

that individuals near the 90th percentile of the average income distribution (making about $100,000 per year) as of 2006 have suffered the smallest income loss of any income group during the recession. In this sense, these workers have been more resilient than both higher and lower income groups.

Turning to the other major recession in our sample—the 1979–83 episode—f_2 looks very similar to the Great Recession period between the 10th percentile and about the 95th percentile, with the same linear shape and a slightly smaller slope. However, for individuals with very low average income (below the 10th percentile), the graph is downward sloping, indicating some mean reversion during the recession.[23] Also, and perhaps surprisingly, there is no steep fall in income for the top 1% during this recession—in fact, these individuals experienced the highest income growth of all income groups during this recession. Overall, however, for the majority of workers, the 1979–83 recession was very similar to—slightly milder than—the Great Recession, both in terms of its between-group implications and its average effect. Of course, the former contains two actual recessions and lasts one extra year, which goes to show the severity of the latter.

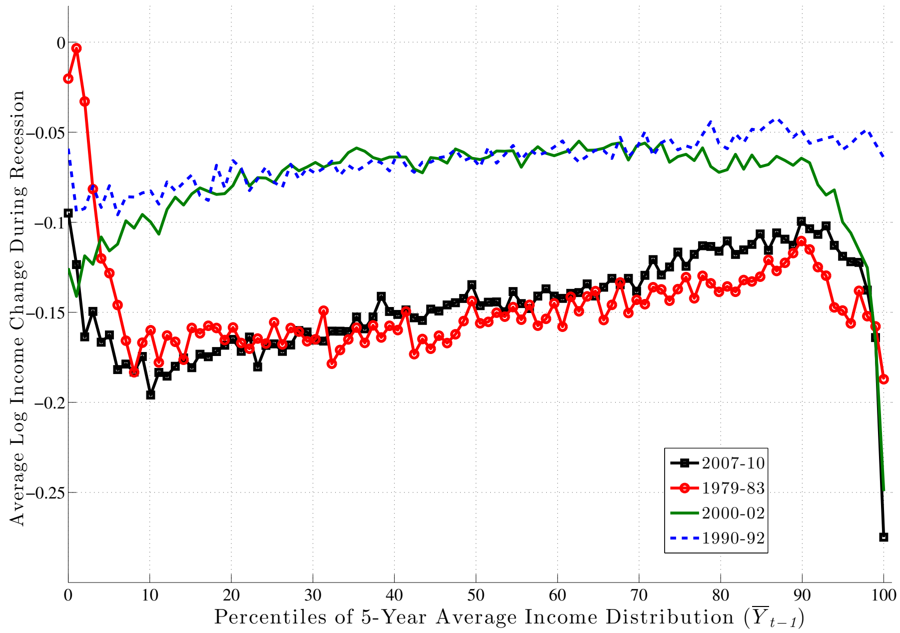
As for the other two recessions during this period, both of them feature modest falls in average income—about 3 log points for the median individual in these graphs. The 1990–92 recession also features mild but clear between-group differences, with f_2 rising linearly by about 7 log points between the 10th and 90th percentiles.[24] The 2000–02 recession overlaps remarkably well with the former up to about the 70th percentiles and then starts to diverge downward. In particular, there is a sharp drop after the 90th percentile. In fact, for the top 1%, this recession turns out to have the worse outcomes of all recessions—an average drop of 33 log points in two years!

Inspecting the behavior of f_2 above the 90th percentile reveals an interesting pattern. For the first two recessions in our sample period, very-high-income individuals fared better than anybody else in the population, whereas for the latest two recessions, there has been a remarkable reversal of these fortunes and the highest-income workers suffered the most.

[23] We conjecture that this has more to do with the fact that for the 1979–83 recession, we were limited to using only income in 1978 to form groups (rather than taking 5-year averages in other periods), which led to a higher degree of mean reversion than would otherwise have been the case.

[24] These two recessions last half as long as the other two longer recessions, so the slope of these graphs should be interpreted in this context. However, normalizing total income growth (the vertical axis in these graphs) by the duration of each recession is not necessarily a satisfactory solution, because even during longer recessions, the largest income falls have been concentrated within one- or two-year periods (2008–09, for example).

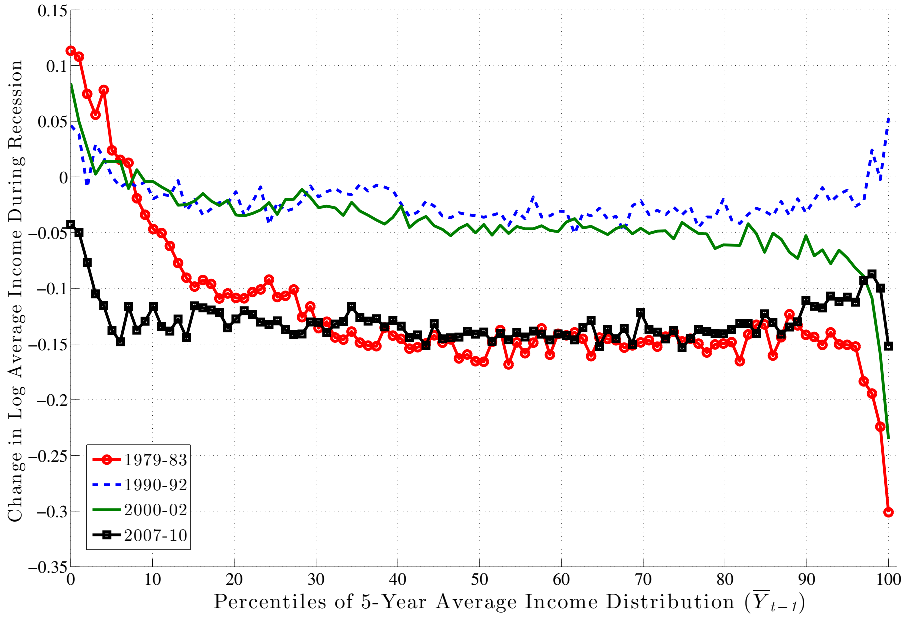
Figure 18: Average Growth in Log Income during Recessions, Prime-age Males



Finally, we construct the alternative measure of average income growth, f_1, described above. Recall that f_1 differs from f_2 in two important ways. First, f_1 excludes individuals with zero income either in year t or $t + k$. Because the probability of full-year non-employment rises in recessions most strongly for low-income individuals (Figure 15), dropping them will tend to increase f_1 below the median relative to f_2. Second, because f_1 is based on the average of log earnings, whereas f_2 is based on the log of average earnings, the latter will tend to be higher within quantiles that have a wider dispersion of income growth rates (due to Jensen's inequality). So, we would expect this force to raise f_2 relative to f_1 below the median level of $\overline{Y}_{t-1}$ where the variance of shocks is higher, as well as at the very top end for the same reason.

Figure 18 plots f_1 for each of the four recessions. A quick glance at the previous figure shows that the two measures reveal the same qualitative patterns. The clear upward-sloping factor structure is there for all recessions. Quantitatively, the slope is somewhat smaller—10 log points difference between 90th and 10th percentiles during the Great Recession vs 17 log points under f_2. Inspecting the two graphs shows that the difference mainly comes from the steeper drop in f_2 between the 20th and 1st percentiles, probably due to the increased chance of unemployment in this range mentioned above. Between

Figure 19: Growth in Log Average Income during Recessions, Young (25–34) Males



the 20th and 90th percentiles the two graphs look very similar. The other recessions show slopes that are also slightly lower than before. Another difference to note is that under f_1, the 1980–83 recession looks less favorable to individuals in the top 10 percent—their income growth pattern resembles the recent recessions more closely. This suggests that the strong performance of this group revealed by f_2 was affected by some large gains at the right tail which dominated the mean income measure for these groups in 1983.

Overall, the two measures are quite comparable. In the rest of the paper, we will continue to mainly focus on f_2 (but comment on differences from f_1 when applicable) so as to capture the total income risk, which includes the risk of long-term unemployment rising during recessions.

Young Males. Before concluding, it is interesting to compare these patterns with those for younger workers: during the 2007–10 recession, f_2 is virtually flat for this group, indicating no significant between-group differences (and even a small rise for those in the 90th percentile and above). In other words, for young workers, the aggregate shock seems to have hit all income groups more or less with the same force (no clear factor structure). The same is true for the milder recessions of 1990–92 and 2000–02, which also show a generally flat shape between the 10th and 90th percentiles. The only deviation is for the lowest incomes during the 1979–83 recession—although the caveat noted above applies here as well. The top 1% also show a strong drop for three out of four recessions. Overall, however, for younger individuals, pre-recession income is a far less useful predictor of fortunes during a recession.

Taking Stock. To sum up our findings for prime-age males, there is a very clear systematic pattern to average income growth during recessions. For the substantial majority of individuals below the 90th percentile, income loss during a recession varies (specifically, decreases) almost linearly with pre-recession average income level. The slope of this relationship also varies with the severity of the recession: the severe recessions of 1979–83 and 2007–10 saw a gap between the 90th and 10th percentiles in the range of 15 log points, whereas the milder recessions of 1990–92 and 2000–02 saw a gap of 4–7 log points. Second, the fortunes of very high income individuals require a different classification, one that varies over time: more recent recessions have seen substantial income losses for high-income individuals, unlike anything seen in previous ones. Below we will further explore the behavior of the top 1 percent over the business cycle.

6.1.2 Expansions

The next two figures (20 and 21) plot the counterparts of the f_2 function during expansions for prime-age and young males, respectively. Broadly speaking, during expansions f_2 displays either a U-shape or a hockey stick shape, which is in stark contrast to the pervasive upward-sloping figure that emerges during recessions.

For prime-age males, there is a clear pattern for workers that enter an expansion with an average income above roughly the 70th percentile: the f_2 function is upward sloping, indicating further spreading out of the income distribution at the top during expansions. For workers below the median, income behavior has varied across expansions. The 1990s expansion has been the most favorable, with a strong mean reversion raising the incomes of workers at the lower end relative to the median. The other two expansions show little factor structure in favor of low income workers—the function is quite flat, indicating that income changes have been relatively unrelated to past income. The pattern is somewhat different for younger workers: there is a pronounced U-shape in all expansions, showing a

Figure 20: Growth in Log Average Income during Expansions, Prime-age Males

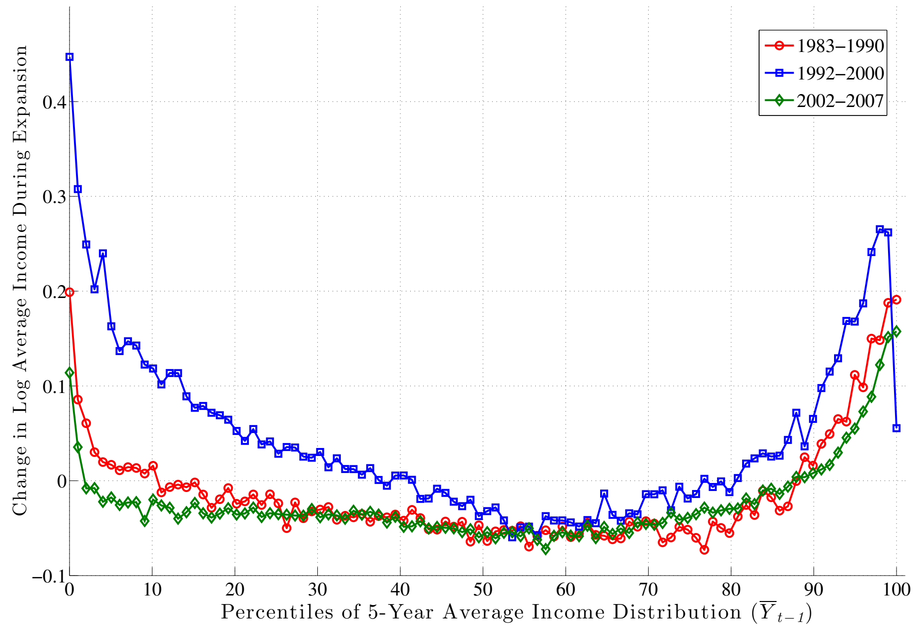


Figure 21: Growth in Log Average Income during Expansions, Young Males

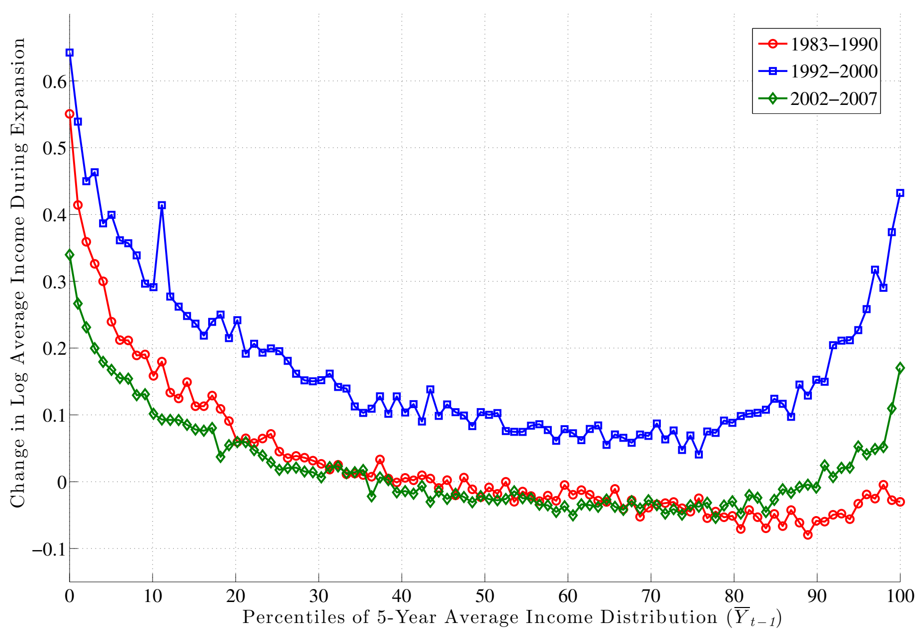
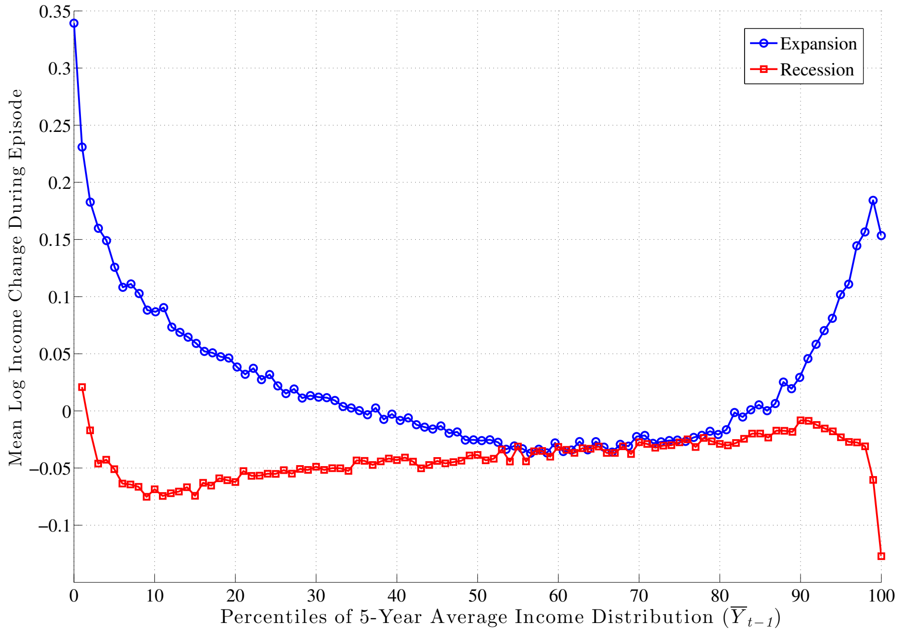


Figure 22: Growth in Log Average Income: Expansions vs. Recessions, All Workers

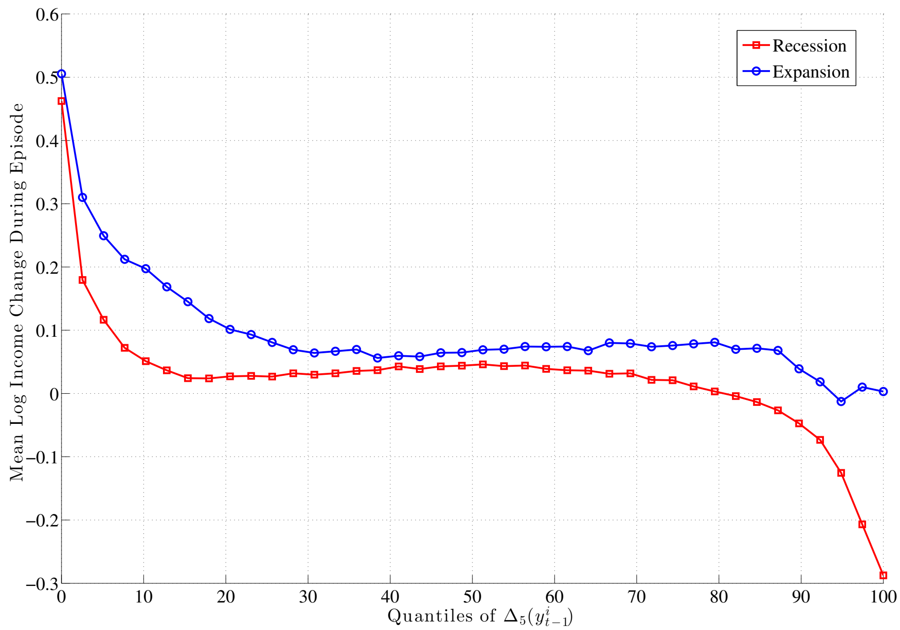


Note: Recession graph has been scaled upward by 6.5 log points to be tangent to the Expansion graph.

catching up for all workers below the median.[25]

To summarize these patterns, Figure 22 aggregates f_2 across all age groups and combines separate recessions and expansions. As seen here, there is a U-shape emerging during expansions—indicating a compression of the income distribution at the bottom and expansion at the top. In contrast, recessions reveal an upward sloping figure, indicating a widening of the entire distribution except at the very top (above the 95th percentile). Thus, the main cyclical impact of business cycles is felt below the median, which expands during recessions and compresses during expansions. The same pattern also emerges at the top—inside the top 10 percent of the income distribution.

Figure 23: Growth in Log Average Income by Quantiles of Recent Growth Rate



6.2 Rising Stars versus Stagnant Careers

We now control for three characteristics—age, $\overline{Y}^i_{t-1}$, and $\Delta_5(y^i_{t-1})$—simultaneously. To this end, we proceed as follows. For each episode under study, we first sort individuals within an age group according to their $\overline{Y}^i_{t-1}$ and $\Delta_5(y^i_{t-1})$ (independently in each dimension) and compute 50– and 40–quantile thresholds, respectively. We use these thresholds to assign each individual into groups formed by the intersection of age, pre-episode average income (indexed by j), and income growth (indexed by p) categories. To give an idea about the bounds of a typical group, for the analysis of the Great Recession, one such group will consist of individuals who (i) were between the ages of 35 and 39 in year 2006, (iii) earned an average annual income ($\overline{Y}^i_{t-1}$) between \$32,033 and \$33,455 from 2002 to 2006, and (iii) experienced an annual income growth rate between 1.30% to 1.49% per year from 2002 to 2006. It is clear that this is a very finely defined group of individuals.

[25]In light of this U-shape pattern, the effect of recessions on young and prime-age males looks more similar, at least below the median of $\overline{Y}_{t-1}$. This is because prime-age males' relatively flat profile during expansions turns into an upward sloping one (indicating falling behind), whereas for the young the downward-sloping profile (indicating catching up) during expansions turns into a flat one during recessions (Figure 19). Thus, for both groups recessions rotate the function f_2 counter-clockwise around the median point.

For each of these 2000 cells, we compute the average labor earnings: $y_t^{j,p}$ and $y_{t+k}^{j,p}$.[26] We then regress

$$y_{t+k}^{j,p} - y_t^{j,p} = \sum_{j=1}^{50} \alpha_j d_{\overline{Y}}^j + \sum_{p=1}^{40} \gamma_p d_{\Delta y^i}^p + u_t^{j,p}, \tag{5}$$

where $d_{\overline{Y}}^j$ is a dummy variable that equals one if the group on the left hand side belongs in the j^{th} quantile of the $\overline{Y}_{t-1}$ distribution and zero otherwise. The dummy $d_{\Delta y^i}^p$ is defined analogously for the quantiles of $\Delta_5(y_{t-1}^i)$. The 90 dummies are estimated via ordinary least squares.

The main finding from this analysis is that pre-episode income growth has a significant effect on future income growth. This is shown in Figure 23, which plots average income growth during expansions (blue line with circle markers) and recessions (red line with square markers). While mean reversion is apparent in both cases, the gap between the two graphs is smallest in the middle and expands at both ends. The implication is that workers with the highest and lowest income growth rates prior to an episode do better during expansions than recessions. This is related to the fact documented earlier that the top of the income shock distribution collapses during recessions. Consequently, those individuals whose income would have grown relatively faster actually slow down during a recession.[27]

6.3 The Top 1 Percent

In this section, we take a closer look at the business cycle variation in the incomes of high earners. As before, we define groups of individuals based on their pre-episode average earnings and examine how they fare during that episode. This allows us to control for compositional changes (within income groups) that are likely to occur during severe recessions or over long time periods (which we will explore below). In the rest of this section, we focus on prime-age males.

[26] Because the two variables can be correlated, there is no presumption that every cell will contain the same number of observations (unlike the previous experiment with a single characteristic). Therefore, we drop cells that have less than 30% of the maximum number of observations.

[27] Incidentally, controlling for past income *growth* has virtually no effect on the relationship between the quantiles of *average* income and future income growth documented above. Thus, further conditioning does not alter the relationship documented so far. Available upon request.

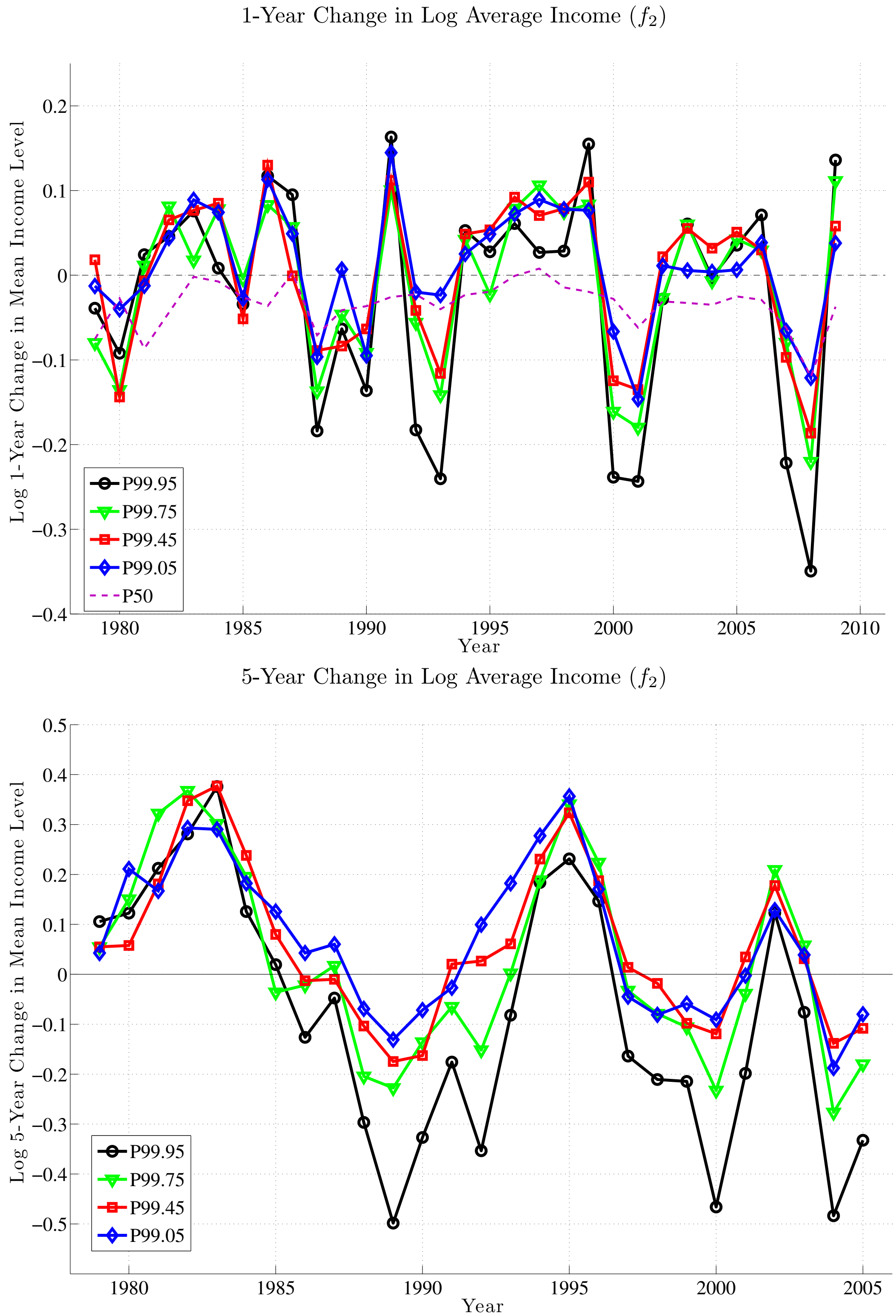


Figure 24: 1- and 5-Year Income Growth, Top 1% of Prime-age Males

To understand the differences and similarities within the top 1 percent, we divide this group into 10 quantiles and focus on the 1st, 5th, 8th, and 10th quantiles. We refer to each quantile by the middle point: P99.05, P99.45, P99.75, and P99.95. The top panel in Figure 24 plots the annual change in log average income (the f_2 measure) over time for each of these quantiles.

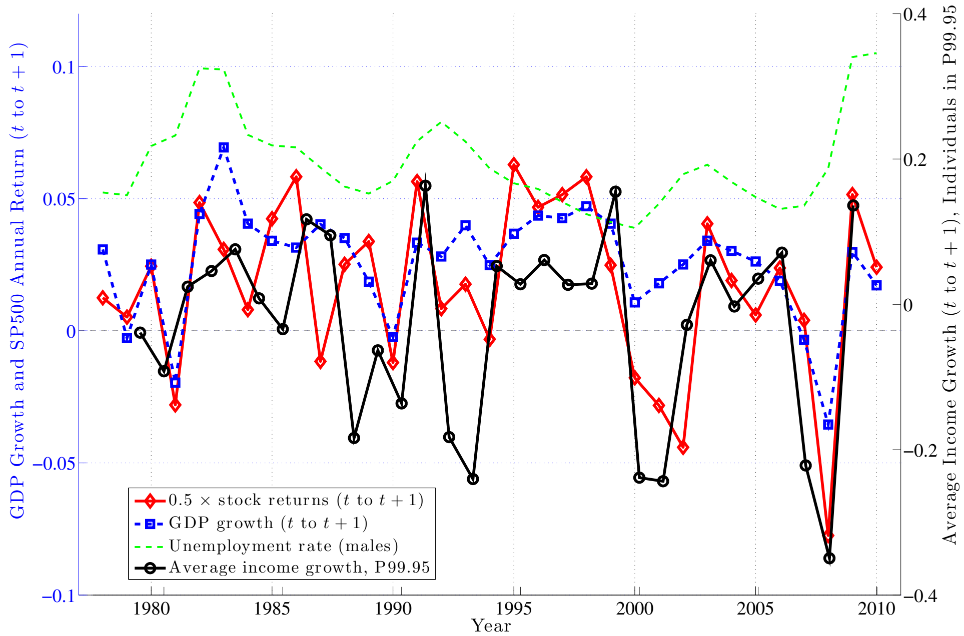
First, notice that the four groups move quite closely to each other until the late 1980s, after which point a clear ranking emerges: higher income quantiles become more cyclical than lower ones. In particular, in the past 20+ years, individuals in higher quantiles have seen their income plummet in recessions relative to lower quantiles, but did not see their income bounce back more in the subsequent expansion, which would have allowed them to catch up. In fact, during every expansion in our sample, the average income for individuals in each group grew by similar amounts. The implication is that these "differential losses" during recessions across income quantiles are very persistent. This can be seen in the lower panel of Figure 24, which plots the 5-year change in log average income.

As seen here, individuals who were in P99.95 as of year 1999 saw their income fall by an average of 50 log points between 2000 and 2005! Similarly large losses were experienced by individuals in the same quantile as of 1988 (during the 1989 to 1994 period) and 2003 (from 2004 to 2009). By comparison, the 5-year income loss for individuals in P99.05 ranges from 10 to 20 log points in the same three recessions. Thus, the cyclicality of incomes and persistent income losses in recessions increase with the level of income. This conclusion is consistent with the findings in Parker and Vissing-Jørgensen (2010) and Saez (2012) who used repeated cross-sections to study the same question. The former paper also finds that the differential cyclicality of incomes at the top has increased after the mid-1980s as we find in this paper.[28]

While the recession/expansion classification used so far is simple and has proved useful, it is also true that there is a lot variation in aggregate outcomes from year to year during a given expansion (and same for a recession). Thus, it is informative to compare the average income changes directly to some key variables that are well-known to co-move with the business cycle. In Figure 25, we plot three such variables—the annual growth in

[28]Notice that the measure used here, f_2, averages each group's income before taking the logs, which could, in principle, be affected by few very large income levels. Alternatively, we can construct f_1 which is the mean of log income changes, which is less sensitive to this problem. We include the resulting graph in the appendix, which shows the same qualitative patterns documented here, and reveals even larger losses for the top income groups—about 60 log points loss for individuals in P99.95 in each of the last three recessions.

Figure 25: Average Income Growth for P99.9 vs. Business Cycle Variables



real GDP per capita, the annual real aggregate return on the US stock market, and the annual male unemployment rate—against 1-year average income growth for individuals in P99.95 (reproduced from the previous figure).[29] The co-movement of income growth for top earners with GDP growth and stock returns is quite striking. In fact, the correlation with all three series exceed 0.6 in absolute value. Furthermore, regressing average income growth separately on each one of these series also reveals significant cyclical sensitivity. For example, post-1985 a 1 percentage point rise in the male unemployment rate has been accompanied with an average income decline of 6.87 percent for individuals that were in P99.9 before the shock. Similarly, a 1 percent slowdown in per capita GDP growth implies a 4.55% decline in the income of the same individuals.[30] For comparison, the corresponding numbers for individuals at the median of the income distribution is 1.08 and –1.77.

[29] The aggregate stock market return is from Robert Shiller's dataset, publicly available on his web site at Yale University.

[30] The corresponding figures for the whole sample period are 4.76 percent and 3.07 percent, respectively.

7 Conclusions

This paper has documented some new facts regarding between- and within-group variation in income growth rates over the business cycle. Using a very large and confidential panel dataset with little measurement error, we document three sets of empirical facts.

Our first set of findings concerns the cyclical nature of idiosyncratic shocks. Our empirical findings strongly support the view that recessions have little effect on the variance of income shocks, but cause a significant increase in the left-skewness of shocks. In other words, during recessions, the upper end of the shock distribution collapses—i.e., large upward wage movements become less likely—whereas the bottom end expands—i.e., large drops in income become more likely. Moreover, the center of the shock distribution (i.e., the median) is very stable and moves very little compared to either tail.[31] What does change (more significantly) is the behavior of the tails, which swing back and forth in unison over the business cycle. These swings lead to cyclical changes in skewness, but not so much in overall dispersion. We conclude that recessions are best viewed as a small negative shock to the median and a large negative shock to the skewness of the idiosyncratic income shock distribution, with little change in the variance.

Second, we examined the systematic component of business cycle risk. First, we found important age variation between young (25–34) and prime-age (35–54) male workers, but not much variation within each group. Second, and more importantly, the pre-episode average income level turns out to be an excellent predictor of a worker's income growth rate during business cycle episodes. The magnitudes are large and the documented patterns are simple (straight lines, or U-shapes, etc). Third, the one deviation we find from these simple patterns is a remarkable non-linearity for individuals who enter a recession with very high incomes—those in the top 1%. During the last two recessions, these individuals have experienced enormous income losses (about 30 log points), which dwarfs the losses of individuals even with slightly lower incomes. In fact, individuals who entered the last three recessions in the top 99.9th percentile of the income distribution had incomes five years later that was at least 50 log points lower than their pre-recession levels.

The analysis in this paper was intended to be as non-parametric as possible, which

[31] This is also related to another fact not studied in detail in this paper: earnings changes are extremely leptokurtic. That is, most income changes from year to year are very small, but once in a while there is a substantial change in earnings. Consequently, in a given year, most observed income changes are small and change little from recession to expansions, which gives the distribution its very stable center.

was made feasible with the very large sample size. However, an alternative approach that has been used in the extant literature relies on fitting a separate AR(1) process to each individual's time series of earnings (see, e.g., Heaton and Lucas (1996) and Bloom et al. (2011)). We have implemented different versions of this method as well. In the most general case, we started by first running the following regression:

$$\tilde{y}^i_{t,h} = \beta^i \times (\overline{y}^A_t - \overline{y}^A) + \left[a^i + b^i h + c^i h^2\right] + \xi^i_{t,h}, \tag{6}$$

to account for a factor structure (first term; where $\overline{y}^A_t$ is average log earnings) and an individual-specific lifecycle component (terms in square brackets). We then fit an AR(1) to the residual income, $\xi^i_{t,h}$, and studied the cyclical properties of the estimated innovation series. This analysis yielded the same substantive conclusions as those reported in this paper. For example, the behavior of the skewness of these innovations mirrored those reported in Figure 8 almost exactly. One drawback of this analysis is that we can only track a handful of cohorts with very similar ages over time (since we need to run the regression in (6) with a sufficiently long time series). Thus, age effects are confounded with time effects, which makes us less comfortable about drawing strong conclusions. Partly due to this concern and for sake of brevity, we did not include these results in the paper. They are available upon request.

A Data Appendix

Panis et al. (2000) and Olsen and Hudson (2009) contain more detailed descriptions of the MEF dataset.

A.1 Comparison to CPS Data

Figure 26 plots the log differential between the 90th and 50th percentiles of the labor income distribution, as well as the log differential between the 50th and 10th percentiles (hereafter abbreviated as L90-50 and L50-10, respectively). A couple of remarks are in order. First, it is useful to compare this figure to the Current Population Survey (CPS) data, which has been used extensively in the previous literature to document wage inequality trends. An important point to keep in mind is that studies that used the CPS have typically focused on *hourly* wage inequality, whereas our dataset only contains information on *annual* (wage and labor) earnings. With this difference in mind, note that Autor et al. (2008, figure 3) report a level of L90-50 of 55 log points in 1978, which rises by about 30 log points until 2005. In this paper, the level of L90-50 is 72 log points (most likely higher because of the dispersion in labor supply) and rises by about 28 log points until 2005, which is very similar to Autor et al. (2008)'s numbers. In both datasets, the rise in L90-50 is secular and is remarkably stable over three decades.[32] Thus, even though the difference between hourly wage and annual income matters for the levels, it has little effect on the secular trend during this period.

Second, turning to the bottom end, the CPS data shows slightly different patterns, depending on whether one uses CPS March weekly wages or May/ORG hourly data. But the general pattern is a rapidly widening L50-10 gap from 1978 to 1987, which then stays flat or declines, depending on the dataset. In our case, the rise in L50-10 happens between 1979 and 1983, which then stays relatively flat until 2000, after which time it starts rising again. It seems safe to conjecture that labor supply heterogeneity could be more important at the bottom end and could account for some of the gap between the two datasets. Another source of difference could be underreporting of income in our administrative dataset, or over-reporting in the CPS. Some papers on measurement error adopt this latter interpretation (e.g., Gottschalk and Huynh (2010)). Notice also that the level of L50-10 is much

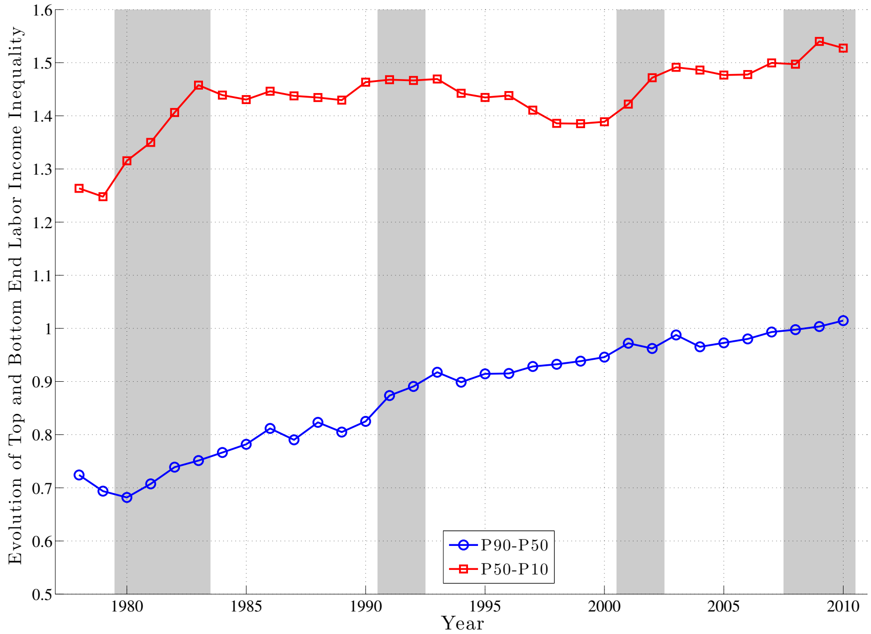
[32] Fitting a quadratic polynomial to the L90-50 reveals a very small negative curvature, indicating an ever so slight slowdown in the rate of increase of inequality at the top.

Table II: Further Statistics of the 10 Percent Sample

	Annual Wage and Salary Income				
Year:	Mean (log)	Std. Dev (log)	Skewness (log)	Median	Max Income†
1978	10.431	0.861	–0.762	38,823	5,629,944
1979	10.420	0.834	–0.798	38,451	3,043,686
1980	10.378	0.843	–0.785	37,124	3,900,245
1981	10.386	0.850	–0.832	37,465	3,191,016
1982	10.353	0.869	–0.756	36,226	3,164,862
1983	10.337	0.891	–0.747	35,966	3,350,164
1984	10.354	0.899	–0.709	36,336	5,649,401
1985	10.363	0.907	–0.677	36,491	5,997,900
1986	10.368	0.929	–0.630	36,535	5,518,408
1987	10.355	0.922	–0.644	36,227	8,836,576
1988	10.352	0.937	–0.599	35,775	10,323,465
1989	10.328	0.929	–0.647	35,098	7,963,985
1990	10.316	0.939	–0.659	34,727	8,436,263
1991	10.303	0.941	–0.585	34,017	7,671,863
1992	10.322	0.943	–0.513	34,403	11,382,868
1993	10.328	0.951	–0.507	34,413	9,824,403
1994	10.317	0.934	–0.566	34,074	7,380,117
1995	10.326	0.939	–0.536	34,128	7,761,374
1996	10.344	0.941	–0.532	34,700	10,145,898
1997	10.384	0.934	–0.465	35,715	11,928,487
1998	10.433	0.925	–0.434	37,191	14,686,511
1999	10.450	0.929	–0.420	37,741	18,190,317
2000	10.468	0.936	–0.400	38,379	32,008,754
2001	10.479	0.951	–0.429	38,855	17,144,706
2002	10.449	0.956	–0.504	38,260	13,885,282
2003	10.443	0.968	–0.499	38,014	14,023,289
2004	10.443	0.965	–0.502	38,159	15,811,530
2005	10.442	0.967	–0.492	37,983	16,138,366
2006	10.452	0.972	–0.484	38,242	18,897,685
2007	10.455	0.982	–0.476	38,330	20,177,930
2008	10.443	0.973	–0.452	37,801	16,907,029
2009	10.407	0.979	–0.460	36,846	12,541,077
2010	10.414	0.974	–0.384	36,787	13,983,100

Note: The sample is truncated at the 99.999th percentile. This condition eliminates about 40 to 60 individuals per year (corresponding to 400 to 600 males in the US economy). †The maximum income reported in the last column corresponds to the truncation point.

Figure 26: Top and Bottom End of Labor Income Distribution



higher in our sample—about 125 log points in 1978 compared with 65 log points in the CPS, which again can be explained by a combination of labor supply heterogeneity and under- or over-reporting.[33] Overall, the two datasets reveal the same pattern at the top end, while having similar but slightly different behavior at the bottom.

B Additional Figure

Figure 27 plots the counterpart of Figure 24 using a different measure of income growth (f_1). The same pattern is visible here with an even larger 5-year loss for all individuals in the top 1 percent.

[33] In our sample, the average wage income at the 10th percentile is $8,520 per year. If an individual works 52 weeks a year at a wage of $5.85 per hour (legal minimum wage in 2007), he has to work 28 hours per week, which does not appear to be an unreasonable figure.

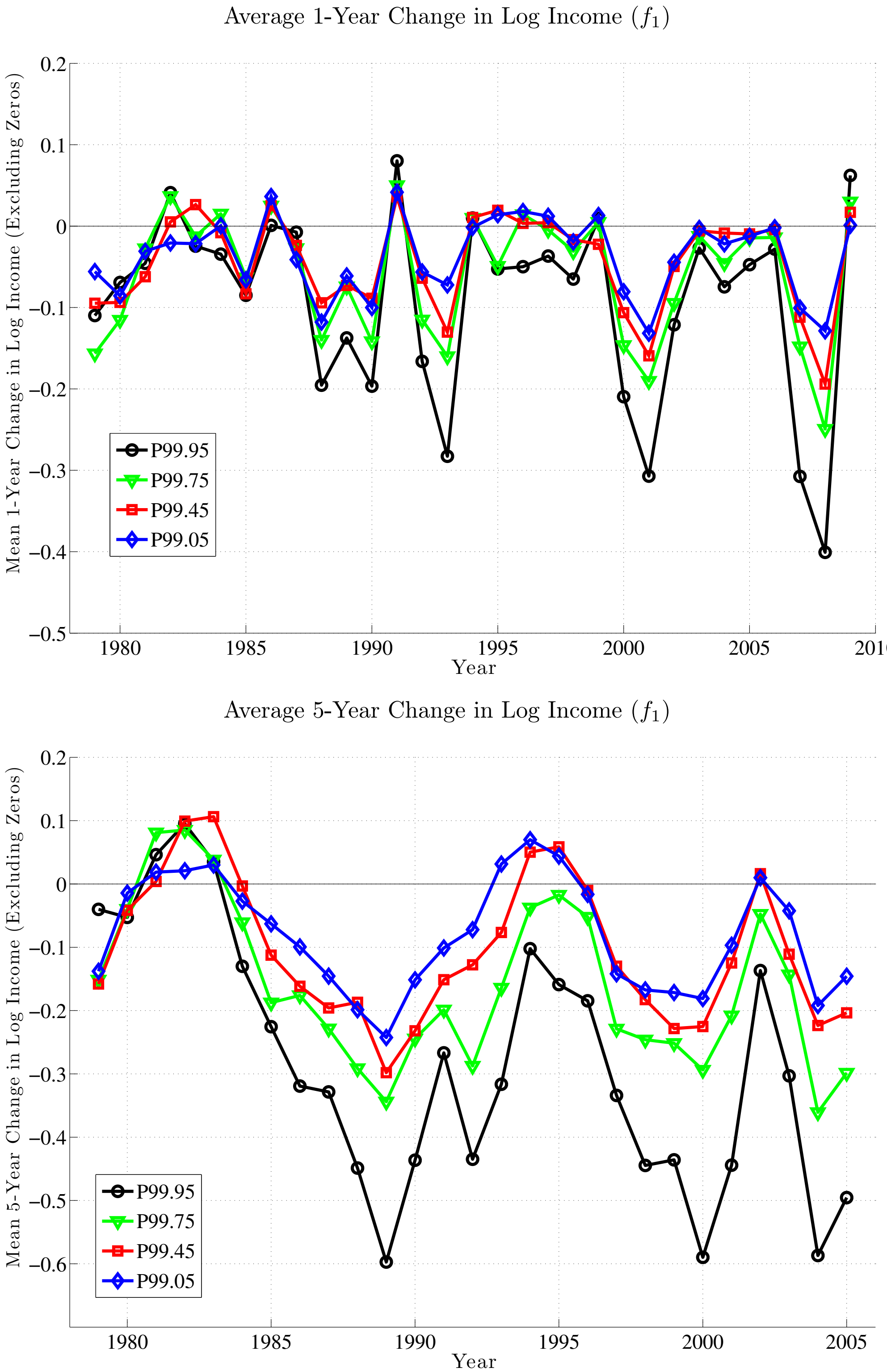


Figure 27: 5-Year Income Growth, Top 1% of Individuals

References

Abraham, Katharine G. and Lawrence F. Katz, "Cyclical Unemployment: Sectoral Shifts or Aggregate Disturbances?," *Journal of Political Economy*, 1986, *94* (3), 507–22.

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney, "Trends in U.S. Wage Inequality: Revising the Revisionists," *The Review of Economics and Statistics*, 2008, *90* (2), 300–323.

Balleer, Almut and Thijs van Rens, "Skill-Biased Technological Change and the Business Cycle," CEPR Discussion Papers No 8410 2011.

Bloom, Nicholas, Max Floetotto, Nir Jaimovich, Itay Saporta-Eksten, and Stephen J. Terry, "Really Uncertain Business Cycles," Working Paper, Stanford University 2011.

Bonhomme, Stephane and Laura Hospido, "Earnings Inequality in Spain," Working Paper, CEMFI 2012.

Bönke, Timm, Matthias Giesecke, and Holger Lüthen, "The Dynamics of Earnings in Germany: Evidence from Social Security Records," Working Paper, Freie Universtat Berlin 2011.

Brav, Alon, George Constantinides, and Christopher Geczy, "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence," *Journal of Political Economy*, 2002, *110* (4), 793–824.

Castro, Rui and Daniele Coen-Pirani, "Why Have Aggregate Skilled Hours Become So Cyclical Since the Mid-1980s?," *International Economic Review*, 2008, *49* (1), 135–184.

Congressional Budget Office, "Recent Trends in the Variability of Individual Earnings and Family Income," Washington, DC 2008.

Constantinides, George M. and Darrell Duffie, "Asset Pricing with Heterogeneous Consumers," *The Journal of Political Economy*, 1996, *104* (2), 219–240.

Dynan, Karen, Douglas W. Elmendorf, and Daniel E. Sichel, "The Evolution of Household Income Volatility," Technical Report, Federal Reserve Board 2007.

Gottschalk, Peter and Minh Huynh, "Are Earnings Inequality and Mobility Overstated? The Impact of Non-Classical Measurement Error," *Review of Economics and Statistics*, 2010, *92* (2), 302–315.

Haider, Steven J. and Gary Solon, ""Life-Cycle Variation in the Association between Current and Lifetime Earnings.," *American Economic Review*, 2006, *96* (4), 1308–1320.

Heaton, John and Deborah J Lucas, "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, June 1996, *104* (3), 443–87.

Juhn, Chinhui, Kevin M Murphy, and Brooks Pierce, "Wage Inequality and the Rise in Returns to Skill," *Journal of Political Economy*, June 1993, *101* (3), 410–42.

Kopczuk, Wojciech, Emmanuel Saez, and Jae Song, "Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937," *Quarterly Journal of Economics*, 2010, *125* (1).

Lemieux, T., "Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?," *American Economic Review*, 2006, *96* (3), 461–498.

Lilien, David M., "Sectoral Shifts and Cyclical Unemployment," *Journal of Political Economy*, 1982, *90* (4), 777–793.

Mankiw, N. Gregory, "The equity premium and the concentration of aggregate shocks," *Journal of Financial Economics*, September 1986, *17* (1), 211–219.

Mehra, Rajnish and Edward C. Prescott, "The equity premium: A puzzle," *Journal of Monetary Economics*, March 1985, *15* (2), 145–161.

Moffitt, Robert A. and Peter Gottschalk, "Trends in the Variances of Permanent and Transitory Earnings in the U.S. and Their Relation to Earnings Mobility," Boston College Working Papers in Economics 444, Boston College July 1995.

Moffitt, Robert and Peter Gottschalk, "Trends in the Transitory Variance of Male Earnings: Methods and Evidence," Winter 2012, *47* (2), 204–236.

Olsen, Anya and Russell Hudson, "Social Security Administration's Master Earnings File: Background Information," *Social Security Bulletin*, 2009, *69* (3), 29–46.

Panis, Constantinjn, Ronald Euller, Cynthia Grant, Melissa Bradley, Christine E. Peterson, Randall Hirscher, and Paul Stinberg, *SSA Program Data User's Manual*, Baltimore, MD: Social Security Administration, 2000.

Parker, Jonathan A. and Annette Vissing-Jørgensen, "The Increase in Income Cyclicality of High-Income Households and Its Relation to the Rise in Top Income Shares," *Brookings Papers on Economic Activity*, 2010, *41* (2), 1–70.

Piketty, Thomas and Emmanuel Saez, "Income Inequality in the United States: 1913-1998," *Quarterly Journal of Economics*, 2003, *118* (1), 1–39.

Sabelhaus, John and Jae Song, "Earnings Volatility Across Groups and Time," *National Tax Journal*, 2009, *62* (2), 347–364.

_ **and** _ , "The Great Moderation in Micro Labor Earnings," *Journal of Monetary Economics*, 2010, *57*, 391–403.

Saez, Emmanuel, "Striking it Richer: The Evolution of Top Incomes in the United States," Working Paper, University of California at Berkeley March 2012.

Schulhofer-Wohl, Sam, "Heterogeneity and Tests of Risk Sharing," Technical Report, Federal Reserve Bank of Minneapolis 2011.

Solon, Gary and Donggyun Shin, "Trends in Men's Earnings Volatility: What Does the Panel Study of Income Dynamics Show?," *Journal of Public Economics*, 2011, *95*, 973–982.

Storesletten, Kjetil, Chris I. Telmer, and Amir Yaron, "Cyclical Dynamics in Idiosyncratic Labor Market Risk," *Journal of Political Economy*, June 2004, *112* (3), 695–717.