NBER WORKING PAPER SERIES

THE NATURE OF COUNTERCYCLICAL INCOME RISK

Fatih Guvenen
Serdar Ozkan
Jae Song

Working Paper 18035
http://www.nber.org/papers/w18035

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2012

A substantial part of this paper was completed while Guvenen was a visiting economist at the Federal Reserve Bank of Chicago, whose hospitality is gratefully acknowledged. The views expressed herein are those of the authors and not necessarily those of the Social Security Administration, the Federal Reserve Bank of Chicago, the Federal Reserve System, or the National Bureau of Economic Research.

The Nature of Countercyclical Income Risk
Fatih Guvenen, Serdar Ozkan, and Jae Song
NBER Working Paper No. 18035
May 2012
JEL No. E24,E32,E44,J21,J31

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. To study the former, we group individuals along several observable characteristics at the time a recession hits. We find two variables to be excellent predictors of fortunes during a recession. First, prime-age workers that enter a recession with high average earnings suffer substantially less compared with those who enter with low average earnings. Second, we estimate "individual betas" (analogous to "stock betas" in finance) and examine their out-of-sample predictive power. We find that the earnings of high-beta individuals (those that exhibited higher sensitivity to prior recessions and expansions) fall significantly more during subsequent recessions. Next, we turn to within-group differences. 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. 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.

Fatih Guvenen
Department of Economics
University of Minnesota
4-151 Hanson Hall
1925 Fourth Street South
Minneapolis, MN, 55455
and NBER
guvenen@umn.edu

Serdar Ozkan
Federal Reserve Board
20th and C Streets, NW
Washington, DC 20551
serdar.ozkan@frb.gov

Jae Song
Social Security Administration
Office of Quality Performance
2121 Crystal Drive, Suite 825
Arlington, VA 22202
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 decline in mean earnings just noted, the *median* worker actually experienced a slight rise in earnings—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% 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 on average 5 million individuals in each year for a total of more than 175 million individual-year observations during this period. The smallest subsample we use (in one experiment) contains more than 1.5 million individuals that are observed for at least 30 years (between ages 25 and 60). 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 (due to, e.g., cash earnings), which can be a concern at the lower end of the earnings distribution.

[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.

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 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 four observable dimensions at the time a business cycle episode begins: (i) age, (ii) pre-episode average income, (iii) pre-episode income growth rate, and (iv) individual-specific inherent sensitivity to business cycles as measured by the comovement of a worker's labor income during other expansions and recessions (denoted by β^i).

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. Therefore, 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 and/or in their β^is. 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

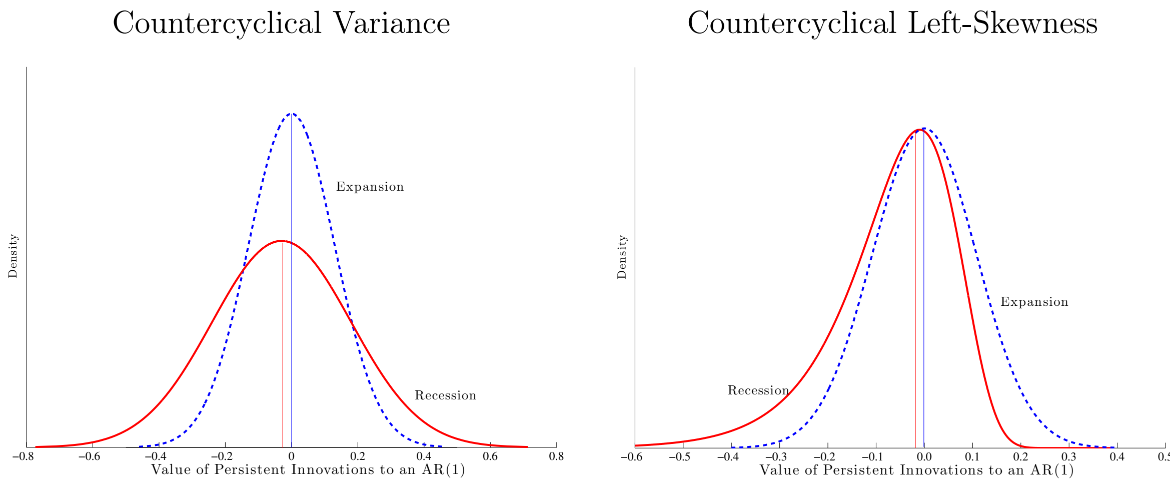


Figure 1: Countercyclical Variance or Countercyclical Left-Skewness?

average income all the way up to the 95th percentile (Figure 19). 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 (about 35 percent).

A second variable that is an excellent predictor of business cycle risk is β^i. To show this, we estimate individual-specific β^is for each worker by excluding a given episode, and then ask how much that β^i allows us to predict income loss during the excluded episode. For example, for the 1980–83 recession, individuals who are in the 90th percentile of the β^i distribution—those with very procyclical incomes—experienced an income loss that was 16 percent worse than that of those in the 10th percentile. The opposite pattern emerges during expansions: For example, from 1994 to 1996, those with high β^i experienced an income growth that was, on average, 15 percent *higher* (in two years) than that of those with low β^i.

1.1 Literature Discussion

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, 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 age, education, and, sometimes, sector. 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 and those whose incomes are procyclical versus countercyclical, among others. Finally, we focus on business cycle variation whereas that literature examined secular trends.[2]

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 rate*," which involves a difference-in-difference-in-difference! It is extremely challenging to answer such a question without a very large and clean dataset.

[2]This paper is also similar in spirit to the asset pricing literature that has paid close attention to a factor structure in stock returns. In a series of influential papers, Fama and French (1992, 1993) built a three-factor model of stock returns that includes a market factor, a "high book-to-market value" factor, and a "small cap" factor. Campbell et al. (2001) investigated the behavior of idiosyncratic volatility in stocks and found it to have increased over time. Their notion of idiosyncratic risk is quite similar to ours and allows for (and removes) components of stock returns that have a cyclical factor structure, i.e., the part that comoves with the market as well as with industry returns. Our formulation is more flexible in certain ways in that we are less parametric (they allow for a linear term as in our β^i). But we allow full non-linearity in lifetime income and find that to be important (especially above the 90th and below the 10th percentile).

[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.

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.

An early literature on the sources of business cycle fluctuations debated the distinction between countercyclical dispersion versus a systematic factor structure. In a provocative paper, Lilien (1982) provided some evidence showing 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 (for 25 years) 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 a full 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 do not necessarily contradict 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]

[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.

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), as well as from the Internal Revenue Service (for self-employed workers, from Schedule-SE and the unreported wages and tips line item on Form 1040). 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 year of birth, sex, race, type of work (farm or non-farm, employment or self-employment), self-employment taxable income, and several other variables. Wage 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).[5] The self-employment earnings variable is capped at the Medicare taxable limit until 1993 and is uncapped starting in 1994 (with the removal of the limit). For more information, see Panis et al. (2000) and Olsen and Hudson (2009). Finally, the labor earnings of a worker are computed as the sum of wage earnings and two-thirds of self-employment income when applicable. The two-thirds factor arises to account for the fact that part of self-employment income is the return on capital invested in the business.[6] 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.

Sample Selection Criteria. We include in our *base sample* an individual-year observation in which the individual (i) is between the ages of 25 and 60 and (ii) has annual labor

[5]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.

[6]This is an admittedly rough adjustment as the capital share of self-employment income is likely to vary not only across sectors but also between any two establishments. However, we found the substantive findings in this paper to be robust even to using wage/salary income alone, so further tweaking this adjustment is unlikely to make a material difference.

Table I: Summary Statistics of the 10 Percent Sample

Year	Mean labor earnings ($)	Mean wage earnings	Change in log average earnings per person ×100	Change in average log earnings per worker ×100	Average age	Number of observations	
	(in constant 2005 dollars)					Base sample	Wage sample
1978	45,864	47,189	—	—	39.5	3,980,599	3,761,356
1979	44,366	45,538	**–0.80**	1.13	39.4	4,162,909	3,931,435
1980	42,872	44,043	**–1.68**	**–3.53**	39.3	4,264,881	4,029,522
1981	43,000	44,201	0.05	1.37	39.3	4,379,317	4,142,328
1982	42,295	43,588	**–2.12**	**–3.58**	39.2	4,356,208	4,110,124
1983	42,308	43,651	**–3.49**	0.54	39.2	4,436,667	4,171,218
1984	43,627	45,045	1.17	6.30	39.1	4,536,586	4,258,346
1985	44,582	46,006	1.51	4.34	39.1	4,712,759	4,426,662
1986	45,681	47,194	1.00	3.62	39.1	4,821,170	4,515,245
1987	45,258	46,769	**–0.12**	1.97	39.1	4,962,472	4,639,916
1988	45,970	47,538	1.06	3.78	39.1	5,113,613	4,777,804
1989	44,291	45,717	**–1.42**	0.59	39.2	5,237,707	4,898,030
1990	44,003	45,473	**–0.66**	0.28	39.3	5,280,612	4,945,056
1991	43,880	45,052	**–0.42**	**–0.93**	39.5	5,330,263	4,986,447
1992	45,494	46,731	1.08	3.00	39.7	5,353,365	5,002,691
1993	45,850	47,126	0.23	3.23	39.9	5,427,052	5,068,495
1994	44,719	45,482	**–1.00**	2.29	40.0	5,526,103	5,159,753
1995	45,536	46,343	1.06	3.68	40.2	5,638,754	5,266,294
1996	46,642	47,507	0.98	3.78	40.4	5,697,737	5,319,693
1997	48,803	49,698	2.13	6.22	40.7	5,776,734	5,399,193
1998	51,236	52,168	2.17	6.96	40.9	5,839,855	5,461,915
1999	52,762	53,682	1.58	4.27	41.1	5,935,054	5,555,371
2000	55,035	55,985	2.09	4.01	41.3	6,019,909	5,643,566
2001	55,201	56,208	**–1.29**	1.69	41.5	6,031,253	5,650,159
2002	52,732	53,710	**–2.59**	**–2.68**	41.6	5,981,791	5,587,682
2003	52,815	53,879	**–0.06**	0.41	41.8	5,998,431	5,588,487
2004	53,211	54,179	0.33	2.13	41.9	6,041,125	5,625,467
2005	53,559	54,462	0.36	2.09	42.1	6,091,447	5,668,493
2006	54,482	55,413	0.91	3.10	42.1	6,134,682	5,706,202
2007	55,283	56,273	0.67	2.00	42.1	6,140,373	5,720,943
2008	53,982	54,833	**–1.63**	**–1.27**	42.2	6,061,807	5,674,103
2009	51,149	52,328	**–4.62**	**–6.76**	42.3	5,775,199	5,450,854
2010	52,206	53,222	0.46	1.30	42.3	5,736,414	5,378,909

Note: Mean labor earnings and wage income are computed for individuals that exceed the minimum income threshold in the corresponding year. The two samples have similar age structures with a maximum age gap of 0.2. To save space, we only report the average age for the base (labor earnings) sample.

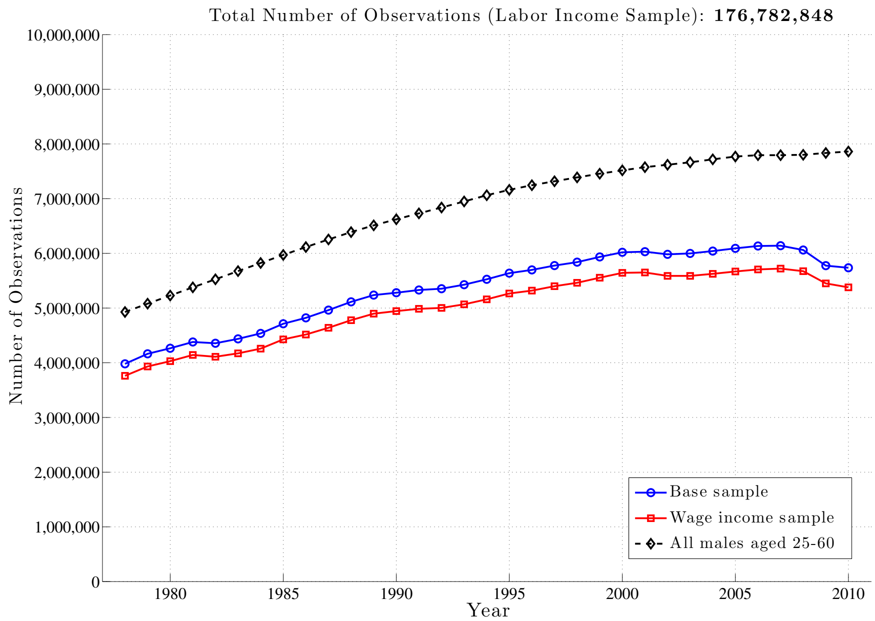


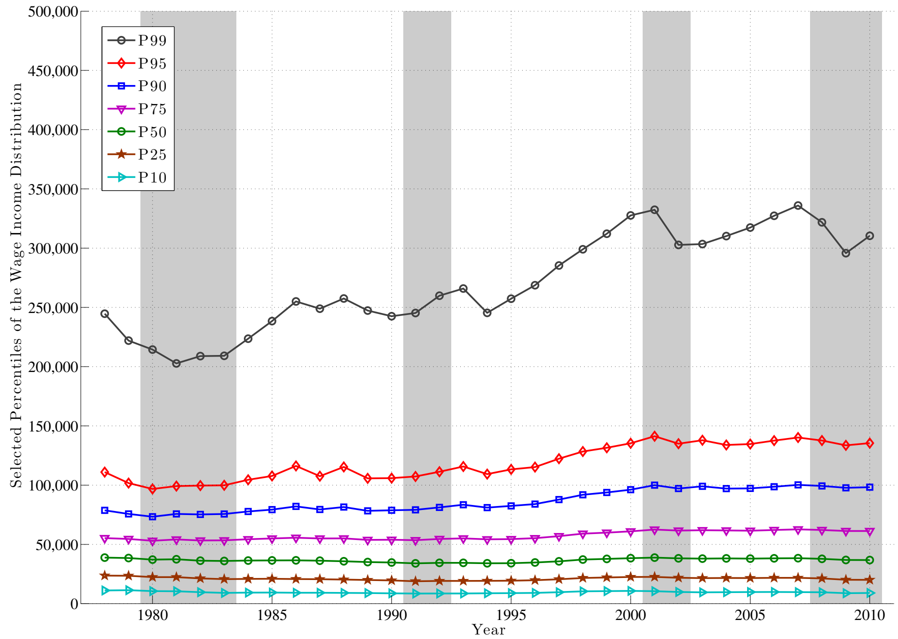
Figure 2: Number of Observations By Year

earnings (i.e., wage/salary *plus* self-employment earnings) that exceeds a time-varying minimum threshold. This minimum, denoted $Y_{\min,t}$, is equal to one-half of the legal minimum wage times 13 weeks of full-time work (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. To abstract from compositional changes, we focus exclusively on males (for whom the average employment rate has been relatively stable) and exclude females, given the large transitional dynamics involving their labor supply behavior during this period. Below, we will impose further conditions for each individual to be included in different experiments.

Table I reports some key summary statistics for the base sample, as well as a secondary sample that excludes self-employed workers—hence called the "wage income sample."[7] The table reports key statistics on 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 four million individuals in 1978 and grows to about six million individuals in 2010. Notice that the number

[7]More concretely, those with self-employment income exceeding $Y_{\min,t}$ are excluded from the base sample to obtain this secondary sample, for reasons explained in Section 6.

Figure 3: Selected Percentiles of Labor Earnings Distribution over Time



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).

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 \$310,400 to be in the top 1%. This threshold was highest in 2007 when it reached \$336,000.

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 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 (i.e., essentially zero growth for our average wage income measure; see column 4 of Table I), and the lack of a significant fall in the unemployment rate (Figure 4). Based on this classification, there are three full expansions and four recessions during our sample period.

3 Inequality Over the Business Cycle: First Look

In this section, we study some statistics of the income distribution to gain some insights into the basic facts about the evolution over time. 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 the earnings *growth* at different horizons, which is informative about the short-run and long-run dynamics of earnings.

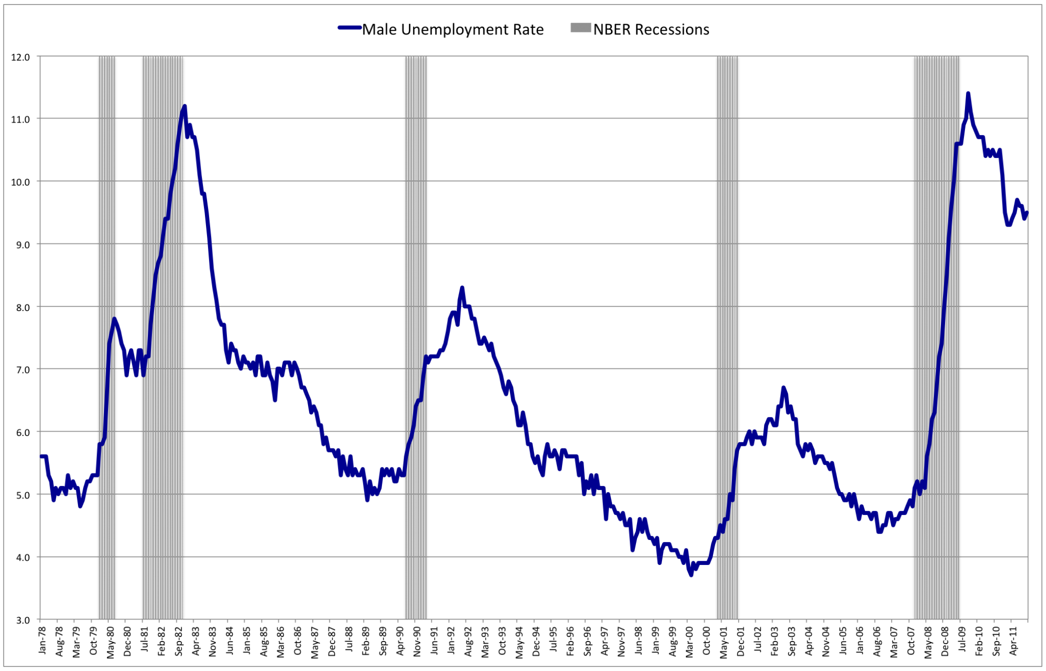
3.1 Evolution of Inequality: Cross-Sectional Data

We begin by briefly establishing some facts on the evolution of cross-sectional inequality. This analysis does not require the panel dimension and provides a comparison to the existing work.

Figure 5 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

[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

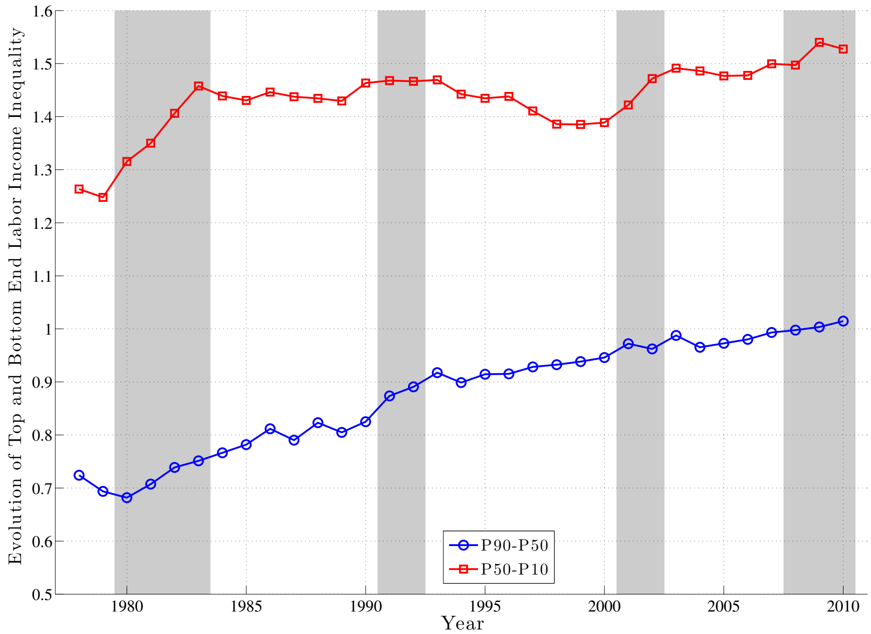


(hereafter abbreviated as L90-50 and L50-10, respectively). A couple of remarks are in order. First, it is useful to compare this figure to the Current Population Survey (CPS) data, which has been used extensively in the previous literature to document wage inequality trends. An important point to keep 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.[9] 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

[9]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.

Figure 5: Top and Bottom End of Labor Income Distribution

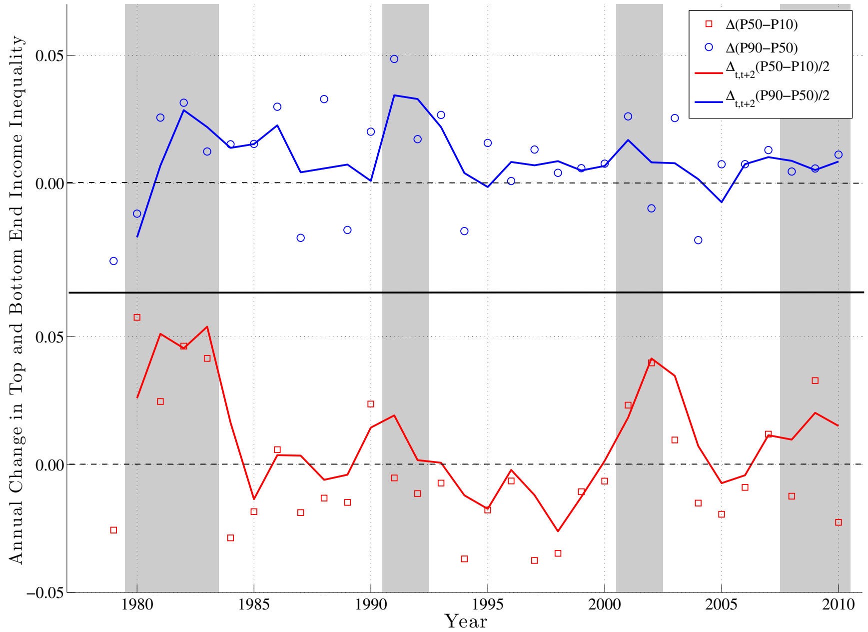


general pattern is a rapidly widening L50-10 gap from 1978 to 1987, which then stays flat or declines, depending on the dataset. In our case, the rise in L50-10 happens between 1979 and 1983, which then stays relatively flat until 2000, after which time it starts rising again. It seems safe to conjecture 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 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.[10] Overall, the two datasets reveal the same pattern at the top end, while having similar but slightly different behavior at the bottom.

Cyclical Inequality. Of course, the focus of this paper is on the cyclical behavior of inequality and risk. To this end, Figure 6 displays the 1-year change in L90-50 and L50-10

[10] 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 6: Change in Top and Bottom End Income Inequality



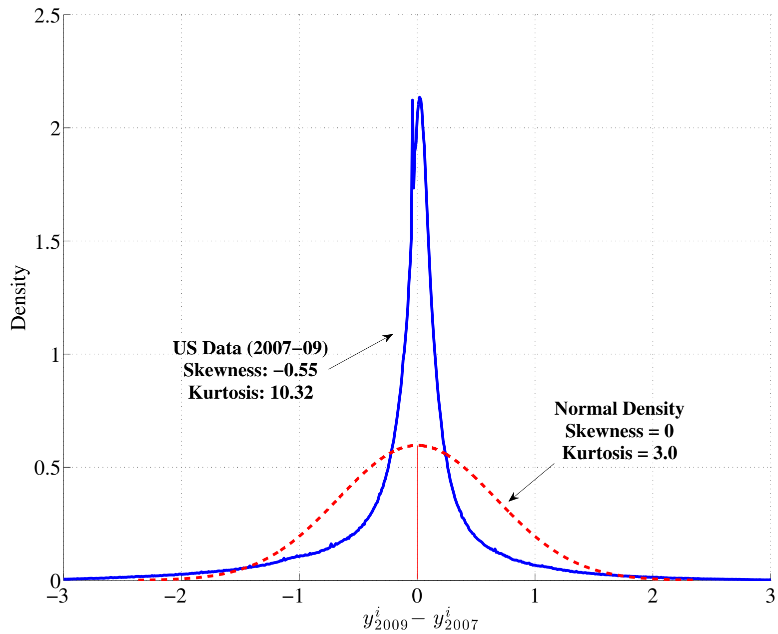
(i.e., the rate of change in Figure 5). To reduce short-term mean reversion in inequality, the solid lines plot the 2-year difference in each inequality measure (divided by two), which is smoother. This time-differencing eliminates the secular trend and allows us to focus on the cyclical change in inequality.

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 inequality itself is clearly 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

Figure 7: Histogram of Δy_t^i: US Data (2007–09) vs. Normal Density with Same Median and Standard Deviation



panel dimension of the MEF dataset.[11]

Figure 7 plots the histogram of the 2007–09 log income change distribution.[12] The histogram is truncated at ± 3, which contains 99.65% of the probability mass. A Gaussian density with the same median and standard deviation is superimposed on this graph for comparison. As seen here, the most striking feature of the data histogram is that it is extremely leptokurtic—i.e., it has a much higher kurtosis (of 10.32) than a Normal distribution (of 3). This can be seen in the figure, with the empirical density having a very pointy center, narrow shoulders, and long tails. To provide a more detailed look, Table II reports, for various time periods, the fractions of individuals whose incomes change less than certain thresholds. Comparing these results with a Normal distribution (last column) with zero mean and the same standard deviation as the 1995–96 period, it is evident that income changes in the US data are far more concentrated near the mean. For example, between 1995 and 1996, about 25.4 percent of individuals had log income changes of less

[11] There are a few studies that have examined the time-series properties of income growth from panel data, although these papers focused on secular trends rather than cyclical behavior (Dynan et al. (2007), Congressional Budget Office (2008), Sabelhaus and Song (2009, 2010), Kopczuk et al. (2010), Solon and Shin (2011), and Moffitt and Gottschalk (2012)).

[12] The general patterns discussed here broadly hold true in other time periods.

Table II: Fraction of Individuals with Selected Ranges of Log Income Change

	Prob($\|\Delta y_t\| < x$)				
$x \downarrow$	1995–96	2007–08	2008–09	**2007–09**	$N(0.00, 0.513^2)$
0.01	0.056	0.060	0.067	0.040	0.015
0.05	0.254	0.262	0.278	0.197	0.078
0.10	0.432	0.439	0.451	0.348	0.154
0.15	0.550	0.550	0.556	0.456	0.229
0.25	0.670	0.675	0.674	0.592	0.374
0.50	0.807	0.810	0.805	0.752	0.670
1.00	0.919	0.919	0.914	0.884	0.949

Note: The standard deviation of the Normal density is chosen to match the data for the 1995–96 period. The empirical distributions are all bimodal. The lower mode for the 2007–09 income growth distribution is at –3.91% (–0.47% nominal growth) and the higher one is at 2.09%.

than 5 log points, and about 44 percent of individuals had changes of less than 10 log points. (These figures are broadly similar for the recession years of 2007–08 and 2008–09.) The corresponding fractions are 7.8 percent and 15.4 percent under a Normal distribution.

The properties of the two-year income growth distribution during the Great Recession are very similar to the one-year change described above, with a slightly smaller kurtosis. Excess kurtosis will reappear throughout our analysis in several important variables and will play a crucial role for understanding some of our main results. Moreover, Figure 7 also shows a pronounced left-skewness in the empirical density, with large drops in income more likely during a recession than large increases.

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 8 plots the evolution of the top and bottom ends of the transitory income shock (i.e., Δy) distribution. The first obvious and important observation is that the top and bottom 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 of the income growth distribution (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 8 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 the rise in L50-10. An alternative measure of shock dispersion—the variance—is plotted in Figure 9 for both persistent and transitory shocks, which shows that the variance 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 a 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 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 variance of income growth (Figure 9) 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).[13] In this paper, we will not dwell much on this trend, except when it is relevant for our analysis of the cyclical

[13]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.

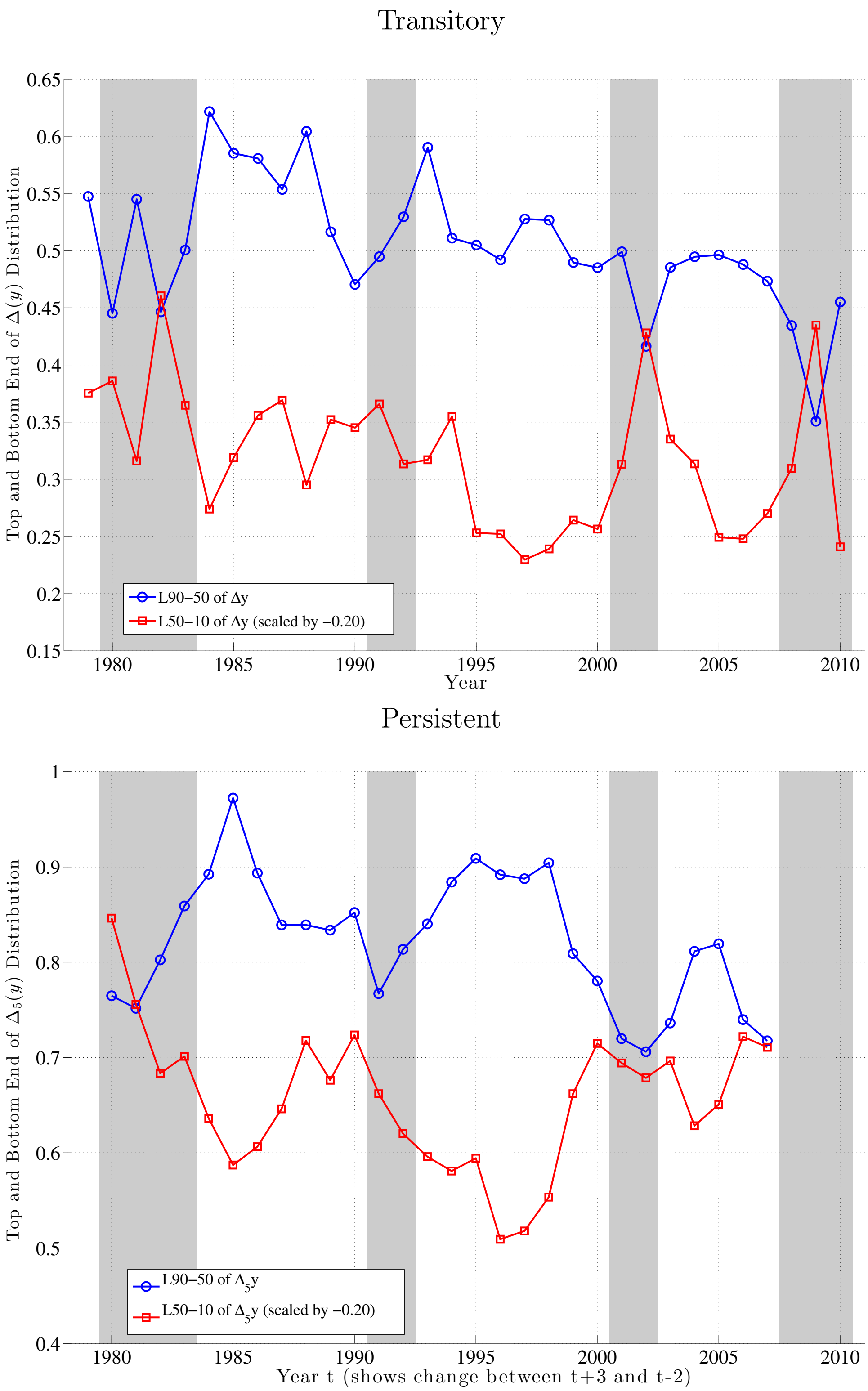
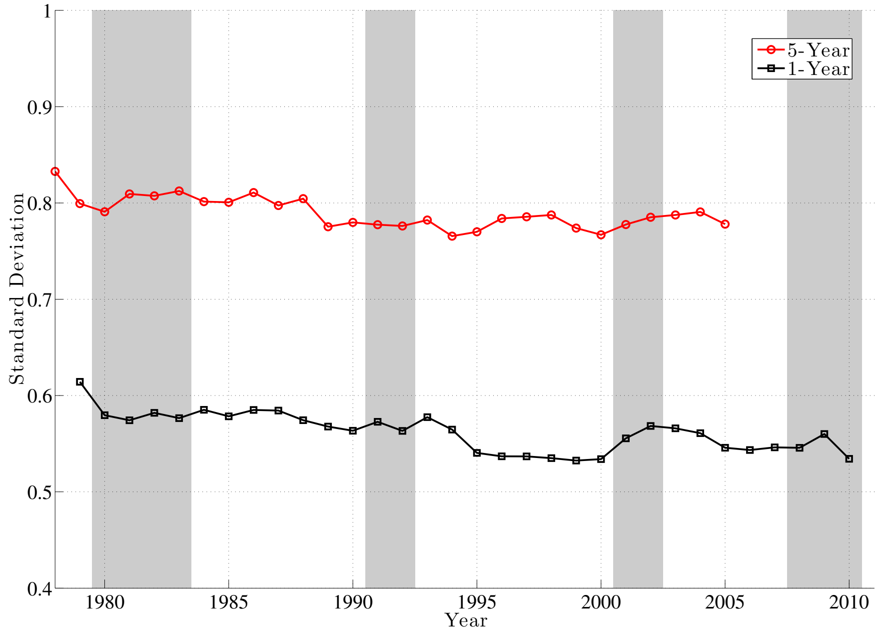


Figure 8: Top and Bottom Ends of Wage Income Change Distribution

Figure 9: Standard Deviation of Transitory and Persistent Income Changes

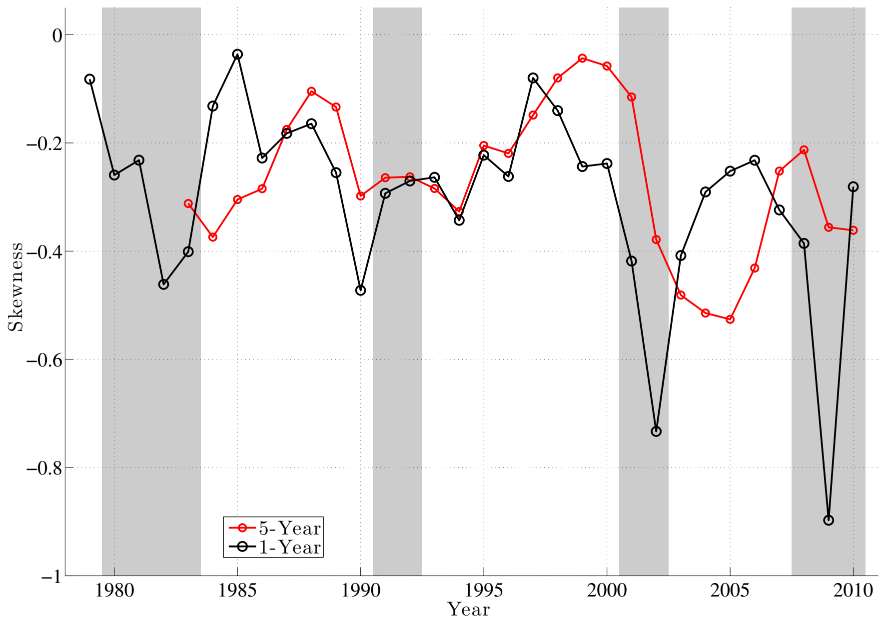


changes in income risk.

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 6. Thus, top-end inequality rises in recessions, whereas the probability of individuals getting very positive income shocks becomes smaller. This is something that we would need to understand and we will dig deeper to that end.

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 10 plots the evolution over time of 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.

Figure 10: Skewness of Transitory and Persistent Income Changes



4 Panel Analysis

The analysis so far intended to provide a general look at how income shocks vary over the business cycle. While this analysis was useful for this purpose, 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.

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 are used to group individuals as of period $t-1$. Consider the following representation:

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

where $g(\theta, h_t)$ is a flexible function of age (h) that captures lifecycle effects in log labor earnings, 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, and λ_t denotes the aggregate shock in the economy. Therefore, this specification allows the distributions of both persistent and transitory shocks to vary across groups and over the business cycle.

Now define 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, take the kth difference of income in equation (1) and modify to allow for a factor structure:

$$y_{t+k}^i - y_t^i = \underbrace{f(\mathbf{V}_{t-1}^i)(\lambda_{t+k} - \lambda_t)}_{\text{factor structure}} \tag{2}$$

$$+ \underbrace{[(\eta_{t+k} + \eta_{t+k-1} + ... + \eta_t)] + (\varepsilon_{t+k}^i - \varepsilon_t^i).}_{\text{stochastic component}} \tag{3}$$

The specification in (3) allows for three different types of business cycle effects. First, the factor structure—captured by the introduction of the function f—allows the conditional mean of income growth to vary systematically over 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 time-varying observable individual characteristics (age, past average income, and past income growth rate as of period $t-1$) as well as a fixed unobservable factor loading that is explained later below. 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 (3) again, this time taking the mean conditional on each group:

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

We have

$$\mathbb{E}(y^i_{t+k} - y^i_t | \mathbf{V}^i_{t-1}) = f(\mathbf{V}^i_{t-1})(\lambda_{t+k} - \lambda_t). \tag{4}$$

Equation (4) provides a simple expression for between-group variation in income growth. Taking the means within each group eliminates both permanent and transitory shocks (since they average zero by assumption). Between any two periods t and $t+k$, each group has a different loading factor $f(\mathbf{V}^i_{t-1})$ on the aggregate shock $(\lambda_{t+k} - \lambda_t)$. The key object of interest is $f(\mathbf{V}^i_{t-1})$, whose shape will tell us about the factor structure of income changes over the business cycle. Below, we will group individuals by four observable characteristics and examine how much of the variation in income shocks can be explained by these different characteristics.

One drawback of this measure is that the left hand side of equation (4) can only be computed using individuals whose incomes are positive in year t and $t+k$ (otherwise log income will be minus infinite). 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 look at an alternative measure for the left hand side of (4). 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 to compute

$$f_2(\mathbf{V}^i_{t-1}) \equiv \log \mathbb{E}(Y^i_{t+k} | \mathbf{V}^i_{t-1}) - \log \mathbb{E}(Y^i_t | \mathbf{V}^i_{t-1}). \tag{5}$$

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

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:

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

Two points to note 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 5 years 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 components, 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 computed 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 refer to them as transitory and persistent variances, respectively.

Finally, each of these variances can also differ across groups of individuals. Below, we will compute various statistics, including variances, separately for each one of these groups to investigate the nature of such variation.

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 four characteristics that can be used to form groups. Not every characteristic will be used in the formation of groups in every experiment.

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 is 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 (i) were in the sample between $t-5$ and $t-1$ and (ii) during those five years, had at least three years of income above $Y_{\min,t}$, *including* in year $t-1$. In every experiment that uses average earnings, our sample contains only individuals who satisfy these additional criteria. 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 effects, 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 earnings 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 of 25 year old individuals used in the regression. Then for a worker of age h in year t, we normalize his earnings for each year from $t-5$ to $t-1$ with the (exponential of the) corresponding age dummy in that year: $Y^i_{t-s}/e^{d_{h-s}}$. We then average these earnings from $t-5$ to $t-1$ (and set earnings below $Y_{\min,t}$ to equal the threshold). This 5-year average earnings is denoted with $\overline{Y}^i_{t-1}$.[14]

[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 bunch together individuals whose incomes are similar at more recent dates. To this end, for a given t, define the weight $w_{t,s} = (6-s)1\{Y_{t-s} \geq Y_{\min,t}\}$, which is zero for ineligible observations and declines with s otherwise. Average earnings are: $\overline{Y}_{t-5,t-1} \equiv \frac{1}{N^*_t}\Sigma^5_{s=1} w_{t,s} Y_{t-s}$, where N^* is an appropriate scaling factor: $N^*_t \equiv \Sigma^5_{s=1} w_{t,s}$. This made almost no change to the results reported here.

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.

Putting Them Together. The most detailed grouping we explore conditions jointly on the three variables described above. To form these groups, we first sort individuals within each age group according to $\overline{Y}^i_{t-1}$ and then separately according to $\Delta_5(y^i_{t-1})$. Along each dimension, we compute the thresholds for the $N_{\overline{Y}}-$ and $N_{\Delta y}-$quantiles, respectively. We use these thresholds to assign each individual into groups formed by the intersection of age, recent average income, and income growth categories. This process yields a total of $7 \times N_{\overline{Y}} \times N_{\Delta y}$ groups. This conditioning allows us to treat individuals within each group as fairly similar at time t. To give an idea about the bounds of a typical group, for the analysis below on 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.[15]

In the next two sections, we examine income changes between year t and $t+k$ (where k varies by experiment) both between and within each of these groups. Below, we shall also consider another classification of individuals—based on their inherent sensitivity to business cycles. That is somewhat more involved and is thus postponed until later.[16]

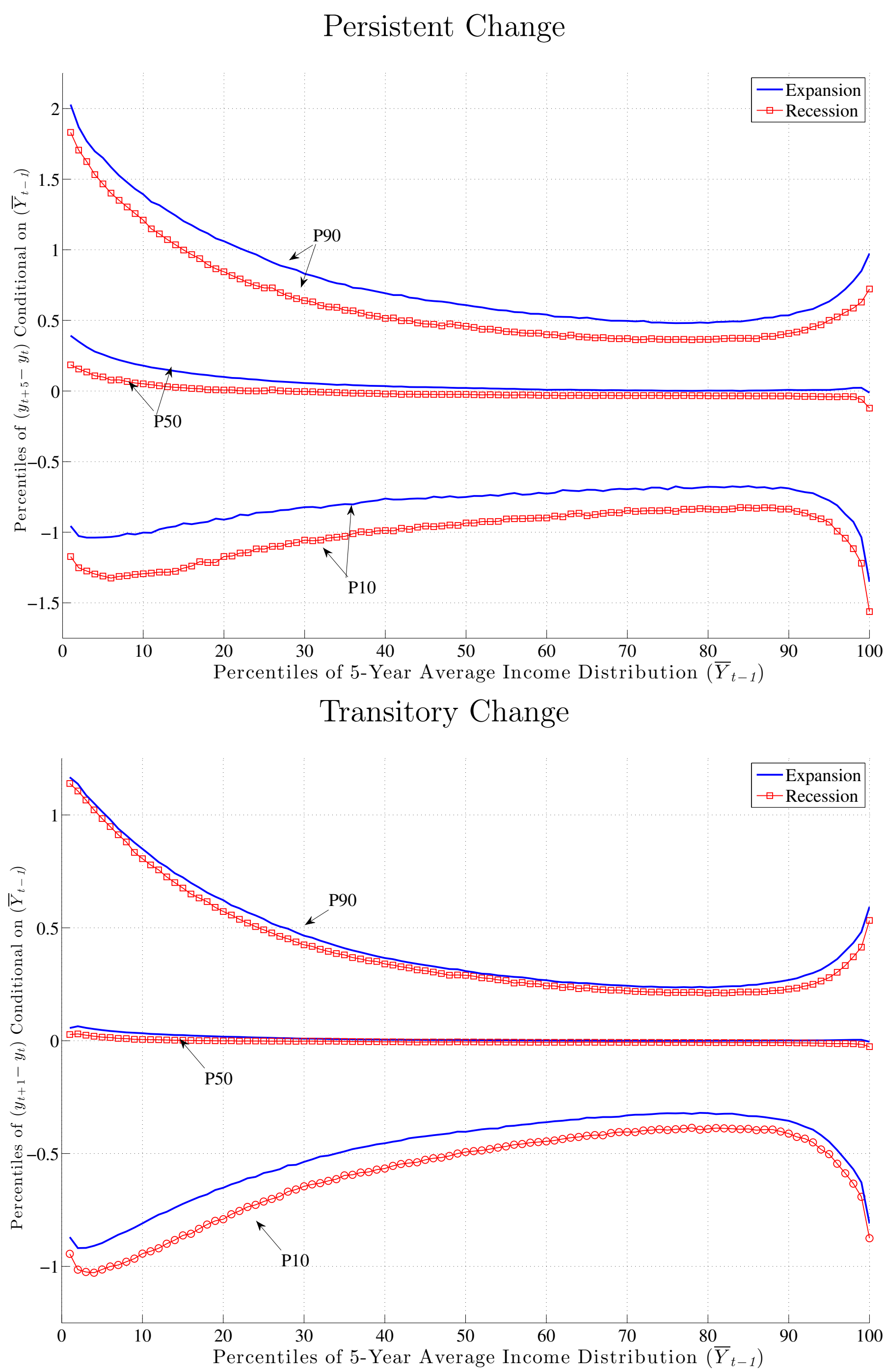


Figure 11: Percentiles of Income Growth Distribution: Recession vs Expansion

5 Within-Group Shocks

We begin with the cyclicality of idiosyncratic shocks—i.e., within-group variation in income growth rates. An important question is whether or not idiosyncratic shocks have countercyclical variances, an idea popular in the existing literature. 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 $\overline{Y}^{i}_{t-1}$. A graphical construct that will be used to study within-group variation begins by plotting the quantiles of $\overline{Y}^{i}_{t-1}$ for a given age group on the x-axis and plots 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 11 is the first use of this graphical construct and contains a lot of information that will be referred to repeatedly in the rest of this section. The top panel displays P90, P50 (median), 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.[17]

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

[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.

[16] 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.

[17] 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.

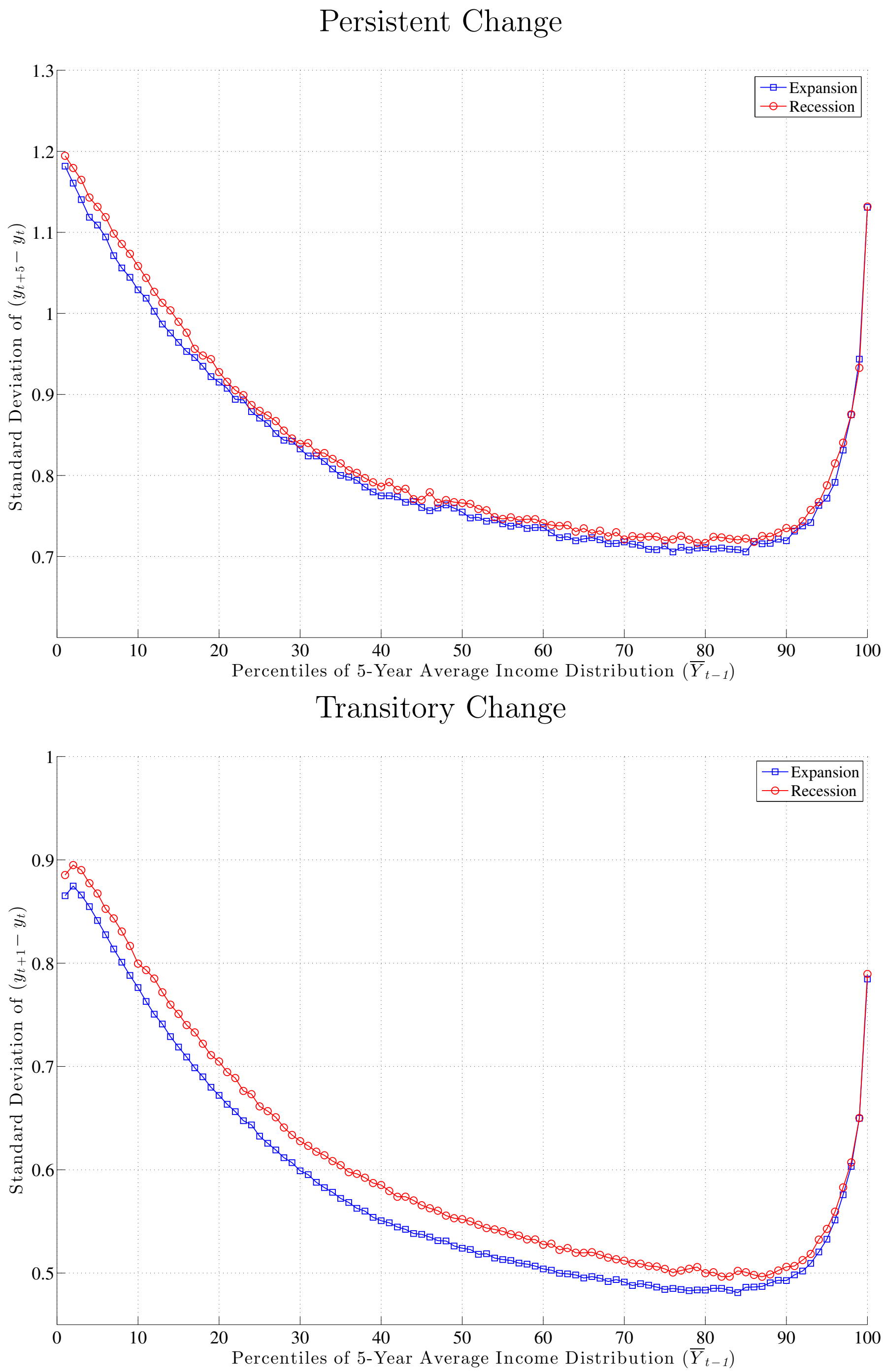


Figure 12: Dispersion of Transitory and Persistent Income Changes

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 80th 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 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).[18] 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 80th and 90th percentiles. Of course, the scales of both graphs are different: the persistent shock distribution (measured both in terms of L90-50 and L50-10) is much wider than the transitory shock distribution, 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.[19]

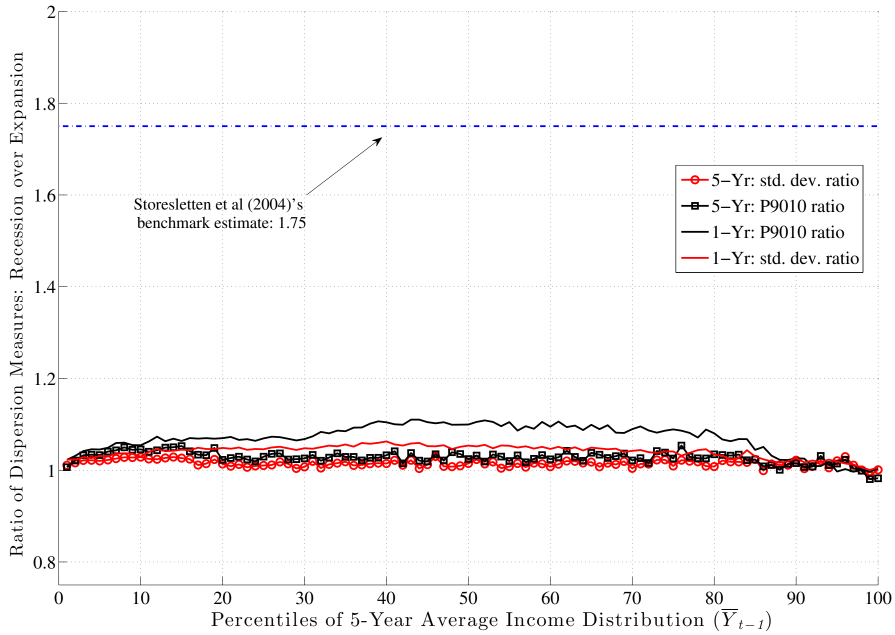
Now we turn to two key questions of interest. First, what happens to idiosyncratic shocks in recessions? Are idiosyncratic shock variances countercyclical? And second, how does any potential *change* in idiosyncratic shocks *vary* across income levels (i.e., the cross-partial derivative)? In other words,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

Are Shock Variances Countercyclical? The countercyclicality of income shock variances has received much attention in the literature. These earlier studies have focused on the cyclicality of persistent shocks, so this is where we also start (top panel of Figure 11). First, both P90 and P10 shift downward by similar amounts from expansion to recession (with P10 shifting down a bit more, but the difference is not very large). As can be anticipated from this, the gap (L90-10) changes by little over the business cycle. Indeed, this

[18]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.

[19]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.

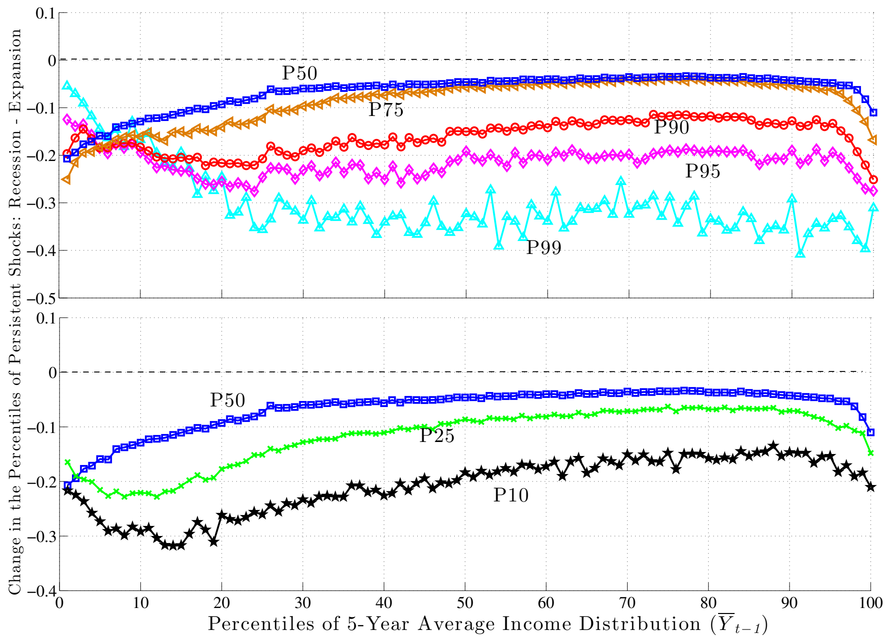
Figure 13: Ratio of Shock Dispersion: Recession over Expansion



is the case, as we shall see momentarily. Furthermore, following the same steps as the one used to construct these graphs, one can also compute the variance of $y_{t+5} - y_t$ conditional on $\overline{Y}_{t-1}^i$ during recessions and expansions, which is plotted in the top panel of Figure 12. 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 13 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 during expansions and 0.21 for recessions. The ratio is 1.75 (marked on the figure for comparison) compared with the 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 action relative to persistent shocks: the standard deviation is higher by about 4% (averaged across the x-axis) and L90-10 is higher by about 7%. These findings suggest that to the extent that recessions involve larger dispersion of shocks, these are to be found

Figure 14: Distribution of Persistent Shocks in Recessions: Compression of Top and Expansion of Bottom



in short-term shocks without much long-term effects. Having said that, these numbers are still very small compared with the values used in the literature.

A second question that was raised above was whether the effects of recessions were felt 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 $\overline{Y}_{t-1}^{i}$ (the cross-partial derivative is zero). This is seen in the three figures just discussed, but is most apparent in Figure 13, where the ratios are quite flat, especially for persistent shocks. Therefore, we conclude that when it comes to the variance of shocks, different income groups are affected similarly by business cycle fluctuations.

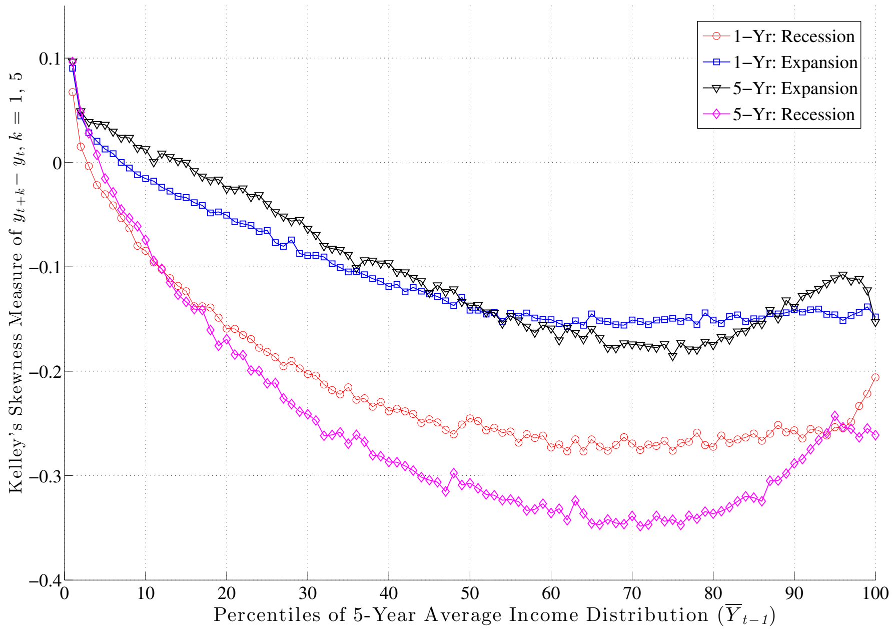
5.1 Countercyclical (Left-)Skewness

The obvious question now is: Do recessions have *any* effect on income shocks? The answer is yes, which could already be anticipated from Figure 11, 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 pervasive across the entire distribution of future income growth rates. This can be seen in Figure 14, 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, again 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 can compute measures of skewness to document this. Figure 15 plots the skewness of shocks within each percentile. 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. 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 15, first, notice that individuals in higher income percentiles face a

Figure 15: Skewness of Transitory and Persistent Income Changes



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, though, 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 $(43\% - 57\%)/100\% = -0.14$) 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 (except, perhaps, for the very low-income individuals in the left corner of the x-axis).

To understand how different this conclusion is from a simple countercyclical variance formulation, let us take a look back at Figure 1, which plots the densities of two Normal random variables: one with zero mean and a standard deviation of 0.13 (corresponding

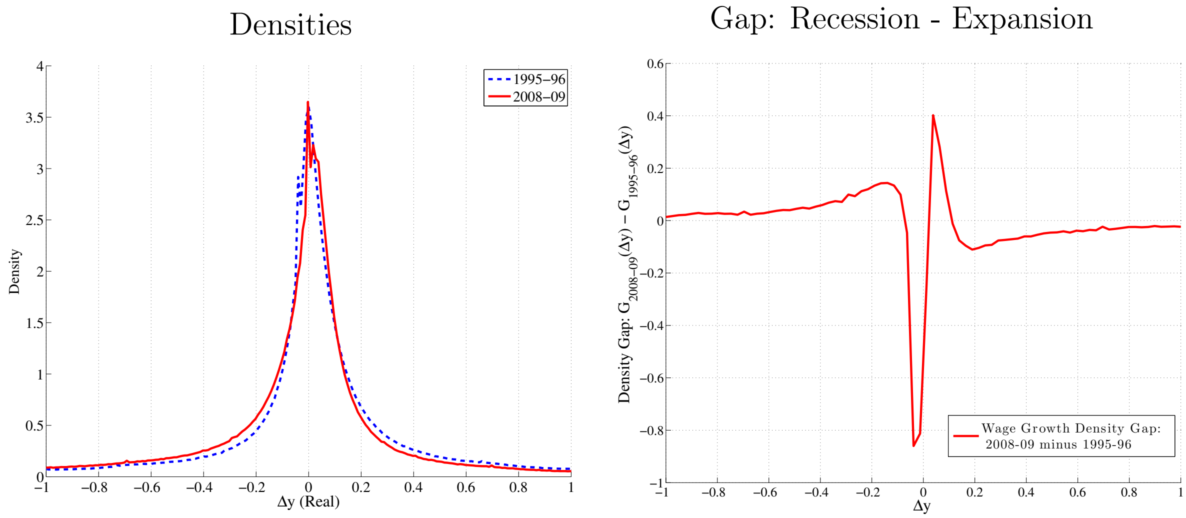
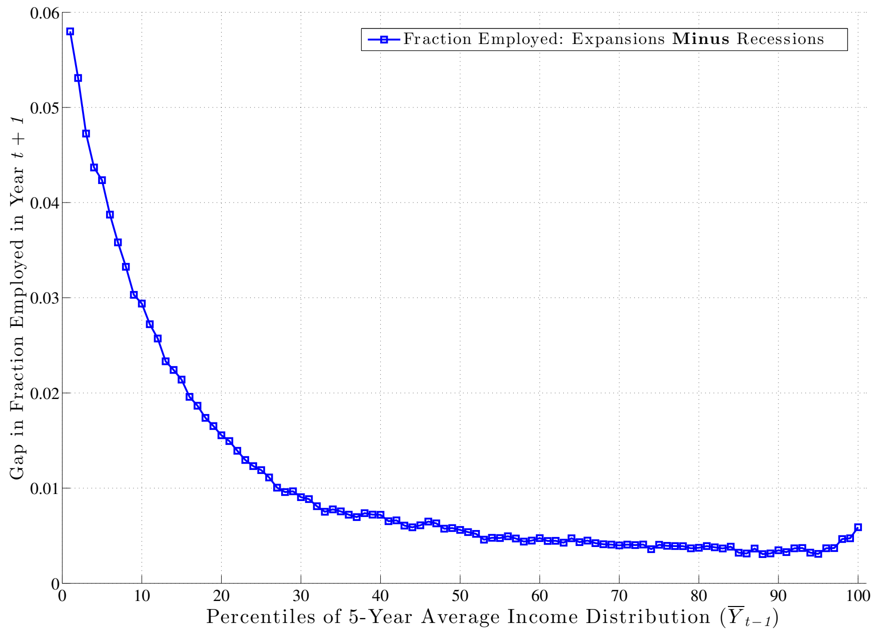


Figure 16: Histogram of $\Delta(\log Y_t)$: US Data 1995–96 vs. 2008–09

to innovation variance in expansions according to Storesletten et al. (2004)) and a second one with a mean of –0.03 and a standard deviation of 0.21 (recession). 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 16 plots the empirical densities of income change 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 right side of the shock distribution is compressed while at the same time the left tail expands. Thus, recessions are times when it is quite unlikely for anybody to make a large upward move in incomes, whereas the risk of falling off the income ladder becomes significantly higher.

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.

Figure 17: Probability of Full Year Non-Employment: Expansions Minus Recessions



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., excluding the long-term unemployed). On average, an excluded 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 17 plots the *gap* (expansions minus recessions) in the fraction of individuals who were employed (at least part year) in t and also had labor income (higher than $Y_{\min}$) in $t+1$. Put differently, the figure plots the *rise* in the risk of full year *non*-employment during recessions relative to expansions as a function of past 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 affect the results presented so far, which mostly 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 17 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. Thus, to the extent that full-year non-employed individuals were incorporated into the analysis above, they would strengthen the findings in the paper. 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, given the relatively modest change in composition, except at the very low end of the past income distribution.

6 Between-Group (Systematic) Business Cycle Risk

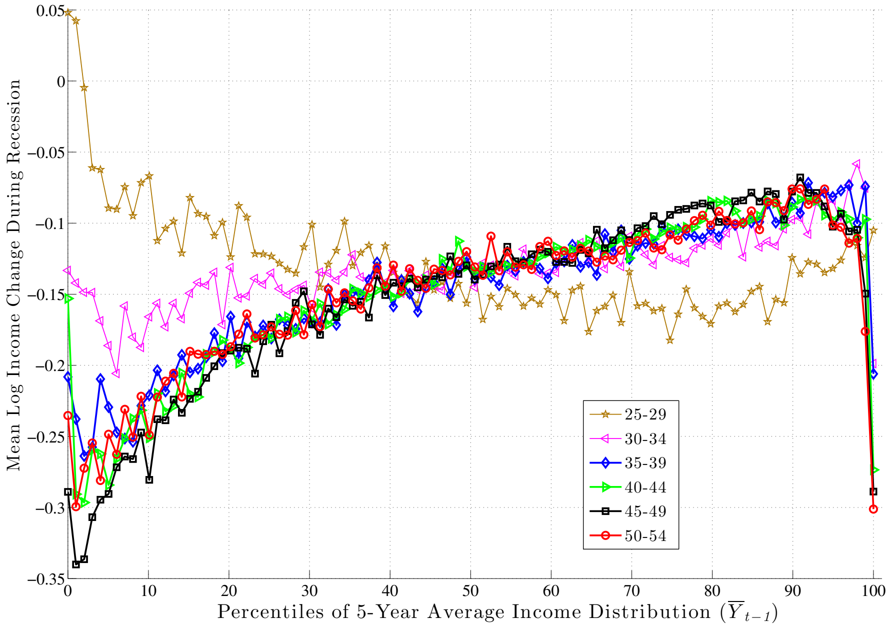
We now turn to the systematic, or between-group, component of income risk. The goal here is to understand the extent to which income growth during a business cycle episode can be predicted by available observable characteristics prior to the episode.

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

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 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 (5).[20] 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 (between-group) variation in income growth across groups of individuals with different $\overline{Y}^i_{t-1}$. If f_2 is upward sloping, that would

[20] We will also construct the f function and compare with f_2 later.

Figure 18: Growth in Log Average Income during the Great Recession (2007–10)



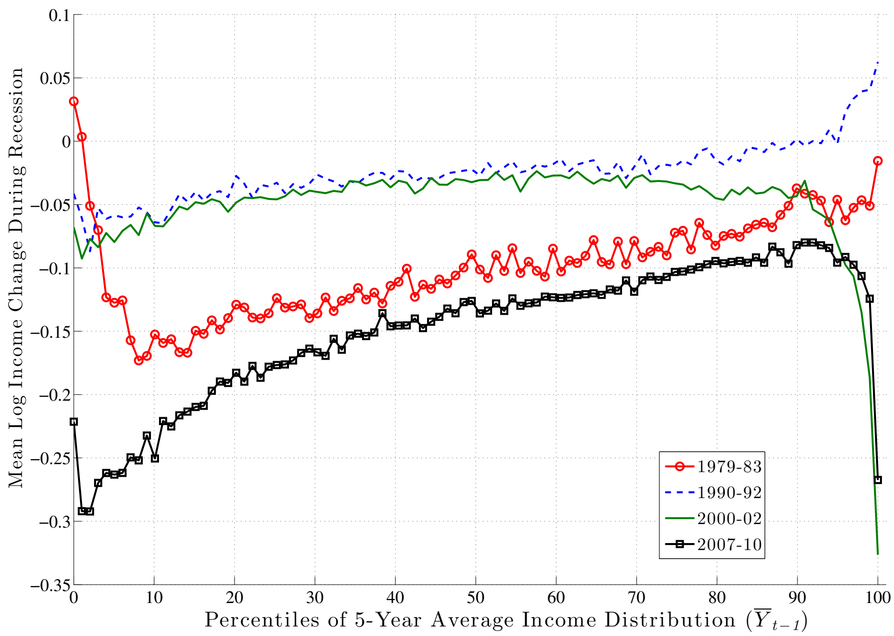
indicate a factor structure in favor of individuals with high $\overline{Y}^i_{t-1}$.[21]

Figure 18 plots the estimated function f_2 for the 2007–10 period, for each of the six age groups (the last one excluded) defined above. 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).

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 19 plots the aggregated f_2 function for prime-age males for all four recessions during our sample period. For the 2007–10 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

[21] This statement relies on shocks *not* exhibiting mean reversion. If z_t was a mean-reverting process, f_2 would be downward sloping in $\overline{Y}_{2006}$. So any sign of upward slope despite this would be an indication of a factor structure.

Figure 19: Growth in Log Average Income during Recessions, Prime Aged (35–54) Males



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 the 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 99th percentile. One conclusion we draw from this analysis is 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

[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.

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.

Finally, we turn to the other two recessions during this period. Consistent with the general view that these were mild recessions, both of them feature modest falls in average income—about 3 log points for the median ($\overline{Y}^i_{t-1}$) 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!

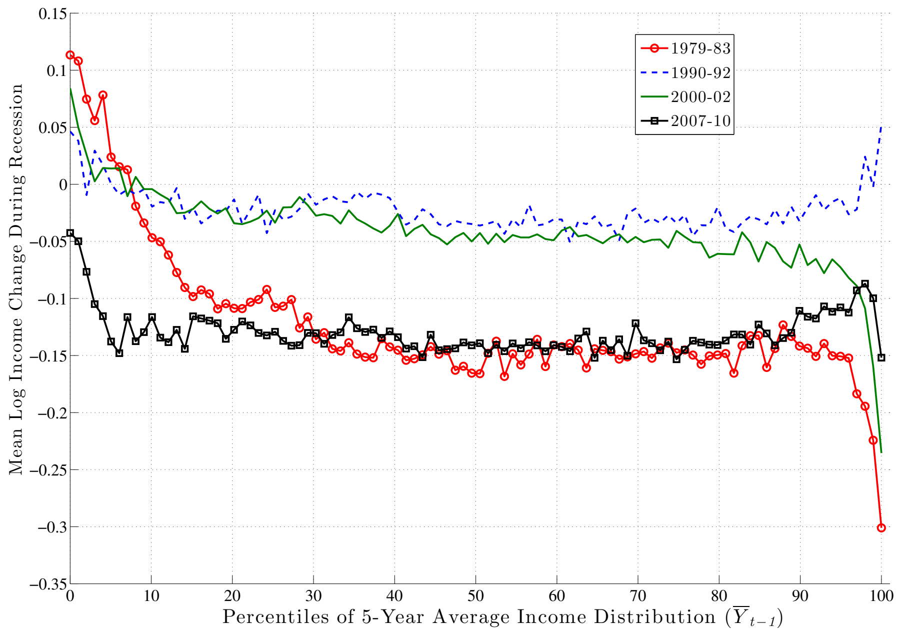
Inspecting the behavior of f_2 above the 90th percentile reveals an interesting pattern. For the first two recessions in our sample period, very high-income individuals fared better than anybody else in the population, whereas for the latest two recessions, there has been a remarkable reversal of these fortunes and the highest- income individuals suffered the most. We conjecture that this new pattern may have to do with the rise of bonuses in the compensation of highly paid workers, which are more likely to be procyclical than base salary. In any case, this is an interesting fact that deserves further study.

Taking Stock. To sum up our findings so far, there is a very clear systematic pattern to average income growth during recessions for prime-age males. For the substantial majority

[23] We conjecture that this has more to do with the fact that for the 1979–83 recession, we were limited to using only year 1978 income 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] It should be stressed that 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 solution, because even during longer recessions, the largest income falls have been concentrated within one- or two-year periods (2008–09, for example).

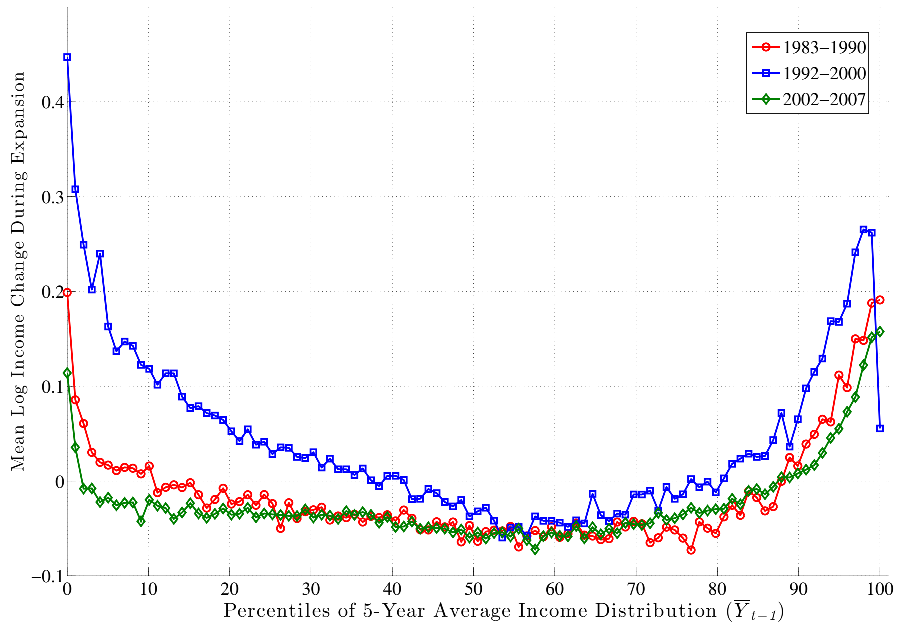
Figure 20: Growth in Log Average Income during Recessions, Young (25–34) Males



of individuals below the 90th percentile, income loss during a recession varies almost linearly with pre-recession average income. 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 4–5 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.

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 seemed to 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

Figure 21: Growth in Log Average Income during Expansions, Prime-aged (35–54) Males

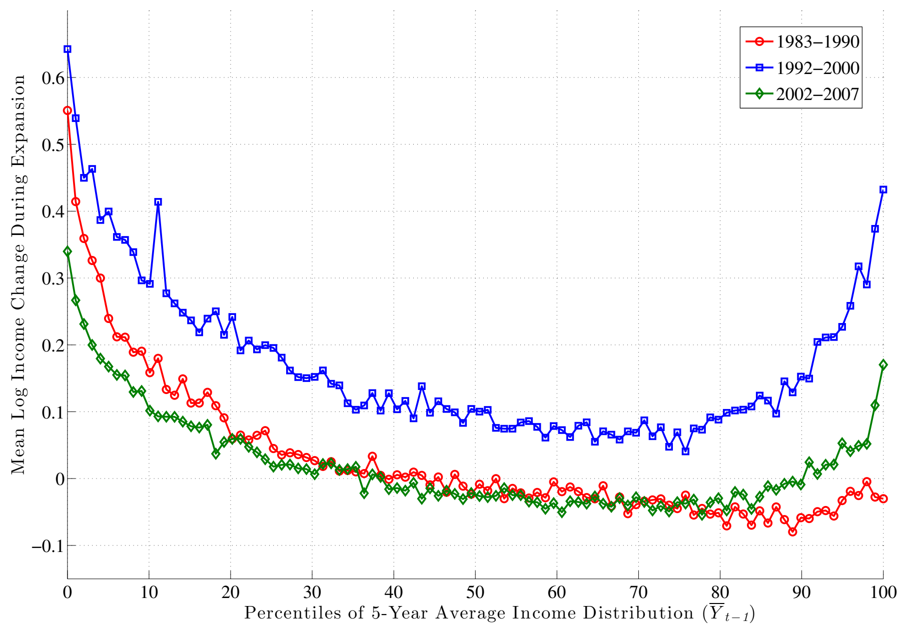


younger individuals, pre-recession income is a far less useful predictor of fortunes during a recession.

Expansions. The next two figures (21 and 22) plot the counterparts of the f_2 function during expansions for prime-age and young males, respectively. Broadly speaking, for all workers in all expansions, f_2 has a U-shape, which is in stark contrast to the upward sloping shape that emerges during recessions.

For prime-age males, there is a clear pattern for workers that entered the expansion with an average income above the median: the f_2 function is upward sloping, indicating further spreading out of the income distribution at the top during the expansion. 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 for all expansions, showing a catching up for all workers below the median.

Figure 22: Growth in Log Average Income during Expansions, Young (25–34) Males



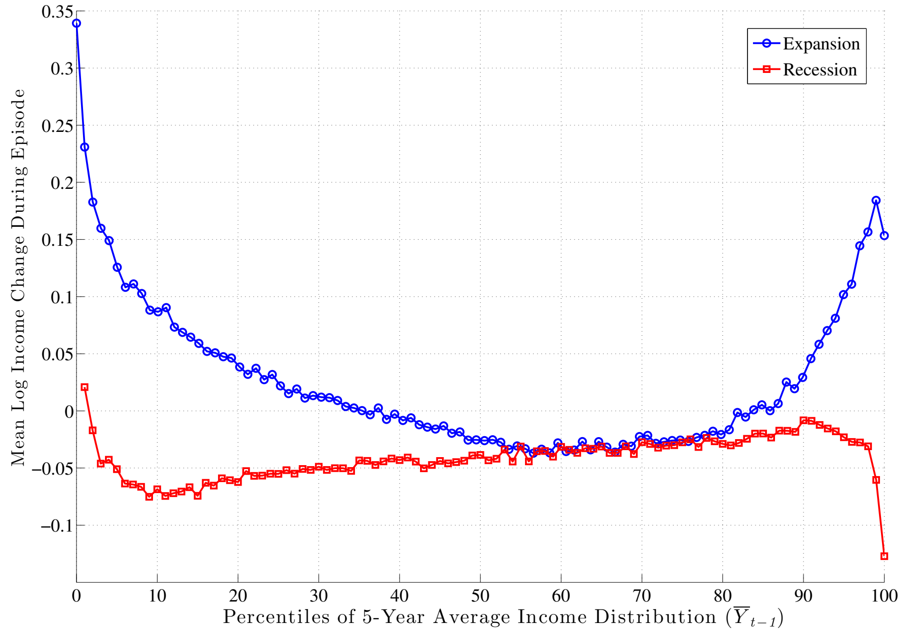
To summarize these patterns, Figure 23 aggregates f_2 across all age groups and combines separate recessions and expansions. As seen here, there is a clear U-shape emerging during expansions—indicating a compression of the income distribution at the bottom and expansion at the top. In contrast, the recessions reveal an upward sloping figure, indicating a widening of the entire distribution except at the very top (above the 95th percentile). Therefore, we conclude that the key cyclical impact of business cycles is felt below the median, which expands during recessions and compresses during expansions.

6.2 Rising Stars versus Stagnant Careers

We now control for two characteristics—$\overline{Y}^i_{t-1}$ and $\Delta_5(y^i_{t-1})$—simultaneously. To this end, we proceed as follows. For each episode under study, we first sort individuals according to their $\overline{Y}^i_{t-1}$ and $\Delta_5(y^i_{t-1})$ (independently in each dimension) and compute 40- and 50-quantile thresholds, respectively. We cross the two variables (indexed by j and p) to obtain 2,000 cells, for each of which we compute the average (residual) labor earnings: $y^{j,p}_t$ and $y^{j,p}_{t+k}$.[25] We then regress

[25] 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

Figure 23: 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.

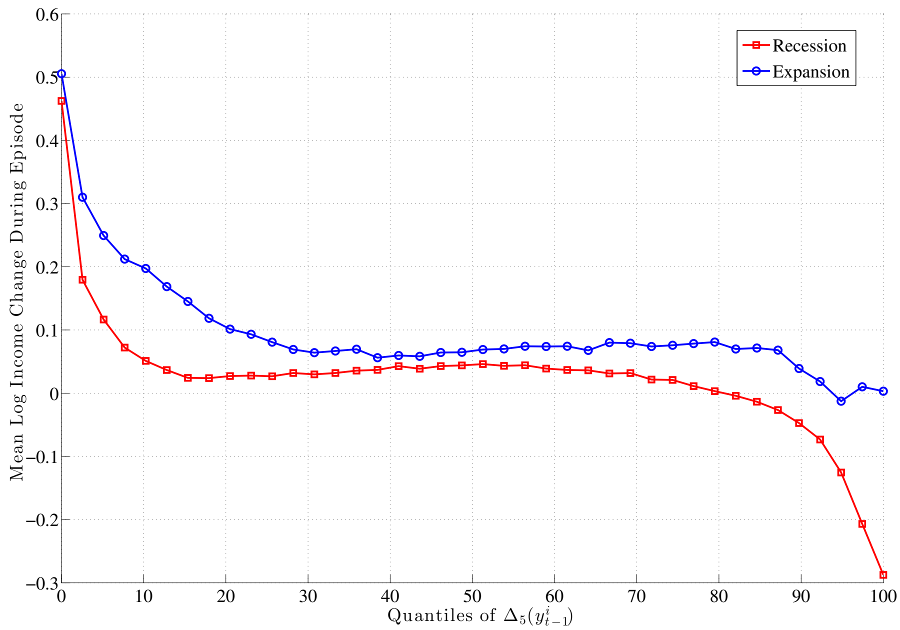
$$y_{t+k}^{j,p} - y_t^{j,p} = \sum_{j=1}^{50} \alpha_j d_{\overline{Y}}^j + \sum_{p=1}^{40} \gamma_p d_{\Delta y^i}^p + u_t^{j,p}, \tag{6}$$

where $d_{\overline{Y}}^j$ is a dummy variable that equals one if the left hand side group belongs in the jth 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 as before via ordinary least squares.

Two main results emerge. First, 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. Second, past income growth itself has a significant effect on future income growth. This is shown in Figure 24, 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

drop cells that have less than 30% of the maximum number of observations.

Figure 24: Growth in Log Average Income by Quantiles of Recent Growth Rate



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.

6.3 Cyclical Incomes: Worker β^is

A different perspective on the factor structure and the systematic component of business cycle risk is the following. Can we predict the fortunes of a worker during a given recession if all we know is how his income changed in previous recessions and expansions? In other words, how useful is a notion of systematic risk in individual incomes?[26]

First, we begin with a version of (1), which we write here as:

$$y^i_{t,h} = \beta^i \times (\overline{y}^A_t - \overline{y}^A) + \left[a^i + b^i h + c^i h^2\right] + \xi^i_t, \tag{7}$$

where β^i is the key individual-specific parameter that measures how sensitive an individ-

[26]This notion is directly analogous to the CAPM literature that seeks to identify the systematic component of each stock return.

ual's income is to business cycle fluctuations. The latter is captured by the movements in log average labor earnings, $\overline{y}_t^A$, which is normalized by its time series average $\overline{y}^A$ so as to have zero mean over time.[27] Here $f(\mathbf{V}_{t-1}^i) \equiv \beta^i$ and $\lambda_t \equiv \overline{y}_t^A$ and the lifecycle component allows for an individual-specific quadratic polynomial. Finally, ξ_t^i is a potentially autocorrelated residual. The advantage of this specification is that by making the factor λ_t observable and the loading term fixed over time, it allows us to estimate this regression separately for each individual using a time series of his earnings.

One goal of this analysis is to assess the out-of-sample predictive power of β^i (that has not been used in estimating equation (7)) for an individual's fortunes during a recession. Thus, for a given recession (e.g., 1979–83, 2007–10), we estimate the regression above by excluding that episode from our sample. (To save on notation, we do not explicitly indicate the episode that has been excluded using a subscript.) We first quantify the differences in income growth between groups of individuals with different levels of β^i. Furthermore, the analysis in the previous section has shown that past average income level and growth rate can be important predictors of fortunes during subsequent recessions.[28] Thus, we also want to quantify the effects of these three risk factors jointly (β^i,$\overline{Y}_{t-1}^i$, $\Delta_5(y_{t-1}^i)$). Because each factor may have very nonlinear effects—and these non-linear effects may also vary by the quantile of the income growth distribution—we will use, as before, an approach that is as non-parametric as possible.

The need to have a long time series of earnings requires us to focus on a subsample of cohorts that are between the ages of 22 to 31 in 1978. These individuals are guaranteed to have at least 30 years of data between the ages of 25 and 60 during our sample period (which may involve years with zero annual earnings). Moreover, because each regression is estimated by taking out a certain episode that ranges from 2 to 4 years, this leaves us with individuals who have at least 26 years of data that can be used for estimation. In addition, to ensure at least a moderately strong labor market attachment for individuals in our sample, we require an individual to have earnings above the minimum threshold during at least 2/3 of the sample period—22 years. There are 1,553,002 individuals that

[27] To compute $\overline{y}_t^A$ for a given year, we include all males between the ages of 25 and 60, including those with zero income. The first difference of this series was reported in column 4 of Table I.

[28] At least two other papers have run regressions of the form (7). Solon et al. (1994) use PSID data and emphasize composition bias for understanding why this regression implies a small cyclicality when using aggregate data but much larger values with micro data. Schulhofer-Wohl (2011) uses a similar regression to measure the cyclicality of earnings for workers with low and high risk aversion using Social Security data that is capped at the taxable limit for most of the period.

satisfy the described criteria and, hence, as many regressions.[29]

So, how is equation (7) estimated? We include zero earnings into this regression by setting them equal to $Y_{\min,t}$. Consequently, the distribution of y_t^i is very non-symmetric due to occasionally large drops to $\log(Y_{\min,t})$. We conducted a Monte Carlo study, which suggested that a quantile regression for the median was more robust and yielded much less bias for β^i (as well as less bias for all other parameters.) Thus, we estimate equation (7) via a quantile regression for the median.

Table III reports summary statistics for the distribution of the estimated parameters of equation (7) using data from 1978 to 2006. The estimated β^is will be used subsequently for analyzing their predictive power for the 2007–10 income changes. Starting from β^i, it has a mean of 0.74 and a large standard deviation of 4.13. The finding that the mean β^i is close to 1 is plausible, although one should not expect it to be exactly equal to 1 for at least two reasons. First, because we track a fixed set of cohorts, they represent a subsample of the population (whose age is also changing over time). Therefore, aggregating the left hand side does not yield the average labor earnings in the economy, which is the $\overline{y}_t^A$ term on the right hand side. Second, because we include an individual-specific polynomial in age, this component can soak up some of the time variation in the left hand side. Considering these factors, the value of 0.74 is quite close to 1.

Second, notice that β^i has a high standard deviation of 4.13. The 10th percentile of the β^i distribution is –2.5 and the 90th percentile is 5.0, indicating that many individuals have significant cyclicality in their labor earnings. Furthermore, the distribution of β^i is very symmetric (skewness of zero) but extremely leptokurtic (kurtosis of 9.93). Finally, Table III also reports the distributions of other parameters. The common finding is one of substantial dispersion across individuals and very high kurtosis for all parameters estimated.

Putting The Pieces Together: β^i and $\overline{Y}_{t-1}^i$. To estimate between-group differences, we proceed as before, now replacing $\Delta_5(y_{t-1}^i)$ with β^i in regression (6). Figure 26 plots the estimated coefficients γ_p, for $p = 1, .., 40$, measuring the log average income growth by β^i quantile during five episodes—four recessions and a short expansion period: 1994–96. As mentioned above, this analysis is conducted with a fixed sample, which is aging over time. So, age effects should also be taken into account in interpreting patterns over different

[29]The three oldest cohorts (aged 29 to 31 in 1978) are older than 60 years old in 2010. Therefore, these three cohorts are excluded from the analysis when we examine the 2007–2010 recession.

Table III: Individual-level Regressions: Characterizing Business-Cycle Comovement

	$y^i_{t,h} = \beta^i \times (\overline{y}^A_t - \overline{y}^A) + [a^i + b^i h + c^i h^2] + \xi^i_t$					
	β^i	a^i	b^i	$c^i \times 100$	σ_ξ	pseudo-R^2
Mean	0.744	9.819	0.077	–0.197	0.491	0.323
Std dev	4.13	1.277	0.162	0.505	0.466	0.206
Skewness	0.05	-1.115	0.313	–0.772	1.680	0.059
Kurtosis	9.93	7.22	6.93	7.48	8.312	4.73
P10	–2.558	8.181	–0.075	–0.795	0.034	0.072
P25	–0.695	9.406	0.00	–0.366	0.110	0.163
P50	0.380	10.082	0.058	–0.116	0.386	0.304
P75	1.938	10.543	0.142	0.004	0.721	0.470
P90	5.007	10.935	0.276	0.239	1.106	0.611
P95	7.963	11.258	0.370	0.461	1.376	0.683
P99	13.830	12.451	0.558	1.131	2.028	0.797

Note: A separate regression is estimated for each individual that is between ages 25 and 60 and has at least 30 years of potential income observations from 1978 to 2010 (cohorts aged 22 to 31 in 1978). There were a total of 1,553,002 individuals who satisfied these restrictions in our sample and, hence, as many regressions. The equation was estimated using quantile regression for the median using data from 1978 to 2006.

episodes over time. The graphs are all normalized to zero at the 10th percentile of β^i to separate the significant nonlinearities to the left of this point.

First, looking at the expansion episode (solid blue line), it slopes up throughout the entire range, consistent with the idea that individuals with high β^i have more procyclical earnings. Of course, the particular β^i used in this exercise is estimated excluding the 1994–96 period, so it is interesting to note the strong predictive power it has for those two years: individuals in the 90th percentile of the β^i distribution experience an income growth that is 13 log points higher than those in the 10th percentile during these two expansion years (1994 to 1996).

Turning to the recessionary episodes, the graphs flip and do so especially strongly for the first two recessions in this period. For example, during the 1979–83 recession, individuals in the 90th percentile of the β^i distribution had their income fall by 15 log points more than those in the 10th percentile. Similarly, during the 1990–92 recession, the gap in income growth between the 90th and 10th percentiles of β^i was +11 log points. The contrast with the subsequent expansion is striking: individuals with β^i in the 90th percentile had their

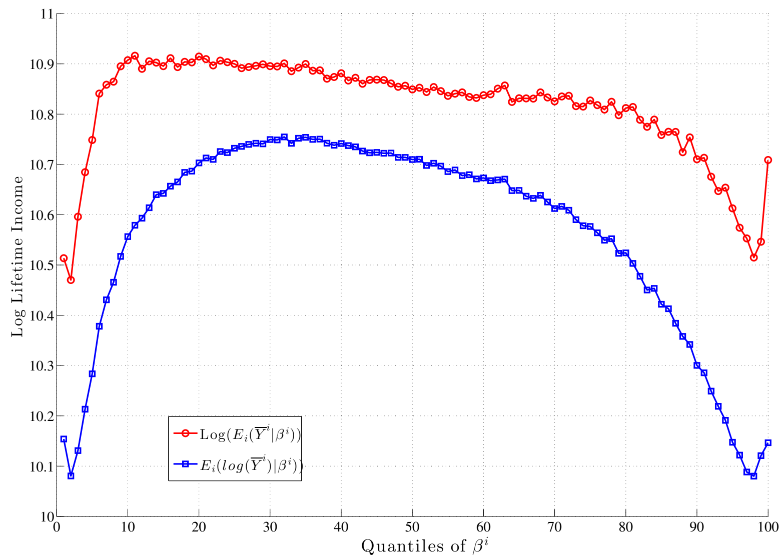


Figure 25: Log Lifetime Income by Quantiles of β^i. The red line with circle markers plots the log of average lifetime income for individuals within a β quantile. The blue line marked with squares plots the average of log lifetime income for individuals within a β quantile. Because of the dispersion of lifetime income across individuals with given β^i and Jensen's inequality, the former is always higher than the latter.

income fall by 15 log points more during the recession than did those with low β^i (those in the 10th percentile), only to have their fates reversed during the expansion: a 13 log point rise relative to the same group.

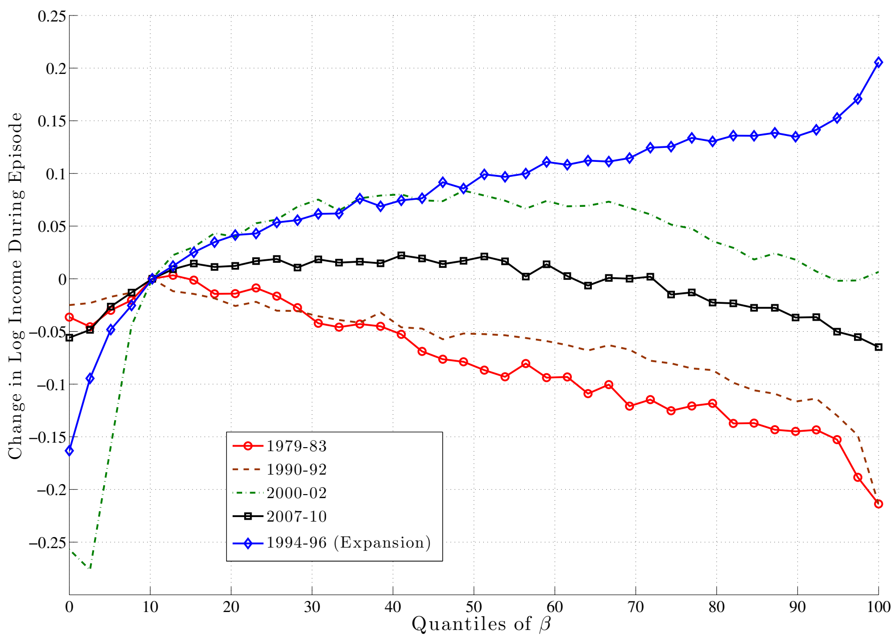
The last two recessions display a similar picture, although a bit more muted, which is likely due to the aging population. For example, during the Great recession, workers with a β^i in the 90th percentile saw their earnings fall by 4 log points more than those in the 10th percentile, whereas during the 2000–02 recession, the gap was reversed but remained very small (about +2 log points).

To conclude, these results reveal a very strong factor structure along two dimensions: pre-episode average income and β^i. These patterns *themselves* also vary strongly between young and prime-age workers.

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

Figure 26: Mean Log Income Growth during Business Cycle Episode by Quantiles of β^i



longitudinal 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 analysis shows that income shock variances are not countercyclical. Instead it is the left-skewness of shocks that is countercyclical. That is, 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 is very stable and moves very little compared to either tail. 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. 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.

Second, we examine the systematic component of business cycle risk. We study four separate dimensions: whether the income loss during recession depends on (i) age, (ii) pre-episode average earnings, (iii) pre-episode earnings growth rate, or (iv) one's inherent sensitivity to business cycles (β^i) measured as the comovement of earnings with the cycle (excluding the episode under study). First, we find important age variation in the patterns that we document between young (25–34) and prime-age (35–54) male workers, but not much variation within each group. Second, two variables (ii and iii) turn out to be excellent predictors 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 remarkably 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.

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.

Campbell, John Y., Martin Lettau, Burton Malkiel, and Yexiao Xu, "Have Individual Stocks Become More Volatile? An Empirical Exploration of Idiosyncratic Risk," *Journal of Finance*, 02 2001, *56* (1), 1–43.

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.

Fama, Eugene F. and Kenneth R. French, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 1992, *47* (2), 427–465.

_ **and** _ , "Common Risk Factors in The Returns on Stocks and Bonds," *Journal of Financial Economics*, 1993, *33* (1), 3–56.

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.

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.

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.

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.

_ , Robert Barsky, and Jonathan Parker, "Measuring the Cyclicality of Real Wages: How Important is Composition Bias," *Quarterly Journal of Economics*, 1994, *109* (1), 1–25.

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.