

The Distributions of Income and Consumption Risk: Evidence from Norwegian Registry Data

Elin Halvorsen Hans A. Holter* Serdar Ozkan† Kjetil Storesletten

February 15, 2017

Preliminary Extended Abstract Version Prepared for the SED 2017 Meeting

Abstract

Using the *Norwegian Registry Data*, containing income and wealth information for the entire Norwegian population, we study the distributions of idiosyncratic income and consumption risk over the life-cycle and over the business-cycle. For this purpose, we first document moments (including higher order moments) from the distributions of growth rates of labor income, business income and capital income, after tax and after transfer income both at the individual level, for males and females, and at the household level. We then decompose the growth in labor earnings into changes in wages and changes in labor hours, in particular, changes in extensive and intensive margins. At the household level, we also study the distribution of consumption risk and the degree of consumption insurance towards labor market risk.

We find that for individual labor income the Norwegian data is qualitatively remarkably similar to the recent studies on population wide U.S. registry data by Guvenen *et al.* (2015, 2014) (quantitatively there is more inequality and larger risk in the U.S.). The much richer Norwegian data, however, allows us to go beyond individual labor income. So far we find (i) The strong negative skewness of individual labor income, which have previously been documented, is due to negative skewness of work hours. (ii) Both capital income and the progressive Norwegian tax- and transfer system contribute significantly towards reducing the effect of the negatively skewed labor income on total individual income.

JEL Codes: E24, J24, J31.

Keywords: Earnings dynamics, consumption dynamics, kurtosis, skewness.

*University of Oslo; hans.holter@econ.uio.no; <https://sites.google.com/site/hansaholter/>

†University of Toronto; serdar.ozkan@utoronto.ca; www.serdarozkan.me

1 Introduction

Quantifying and modeling income and consumption risk is central in many fields of economics, including labor economics, public finance and asset pricing. Until recently, empirical studies have relied on relatively small samples of survey data and there is still high uncertainty about the distribution of income risk and how much of it that translates into consumption risk. An overwhelmingly large part of the literature does for instance, without much empirical support, assume that income risk is normally distributed.

Güvenen *et al.* (2015, 2014) use administrative data from the US Social Security agency to shed light on the distribution of individual annual earnings changes. One important feature of the data is that earnings changes are extremely leptokurtic (kurtosis is as high as 35 compared with 3 for a Gaussian distribution). Namely, in a given year most individuals experience very small changes in earnings relative to the overall standard deviation, but a very few experience extremely large shocks. In addition, earnings data display strong negative skewness, which implies that workers are more likely to experience very large negative changes in their earnings (disaster shocks) compared to very large positive shocks.

Being able to utilize a large panel of registry data, Güvenen *et al.* (2015, 2014), is a first step in the right direction for the empirical analysis of earnings dynamics. However, several important questions cannot be answered due to limitations of the SSA data. One shortcoming of the SSA dataset is the lack of detailed information on components of annual individual earnings, i.e., weekly or hourly wages, unemployment spells and weekly hours. It is not possible to decompose the annual earnings change distribution into its components using the SSA data. Thus we do not know what explains the concentration at the center of the distribution, the very large tails and the negative skewness. Are the non-Gaussian properties of annual earnings changes due to leptokurtic and negatively skewed distribution of wage changes or of hours changes. The answers to these questions are central to understanding the economic mechanisms behind these features of the data and in turn, the consumption insurance available for households against these changes.

In this paper we study the distributions of income and consumption risk for individuals and households, using the Norwegian Registry data. We begin by documenting moments (including higher order moments) from the distributions of growth rates of labor income, business income and capital income, after tax and after transfer income both at the individual level, for males and females, and at the household level. We find that the distribution of individual labor income growth rates in Norway is qualitatively very similar to the U.S. distribution, although quantitatively there is larger variance in the U.S. Similar to what has previously been documented for the U.S. labor income risk in Norway is negatively skewed. However, looking at income, including capital income, the negative skewness is smaller and income after taxes and transfers is even less negatively skewed. This illustrates the roles of saving and of the progressive Norwegian tax and transfer system in insuring against labor market risk.

Next, we decompose the growth in labor earnings into changes in wages and changes in labor hours, in particular, changes in extensive and intensive margins. We find that the negative skewness and high kurtosis of individual labor income is driven by the negative skewness of work hours. The distribution of hourly wages is closer to log-normal.

Furthermore, Norwegian Registry data allows us construct household consumption by providing data on household wealth (along with the data on income). Thus we exploit household budget constraint to construct consumption at the household level (as in Fagereng *et al.* (2015)). Using this measure of consumption we study the distribution of household consumption growth rates.

Last, we document how idiosyncratic risk evolves when one moves from individual labor income to household income before and after taxes and transfers.

2 Empirical Methodology

2.1 Data

A more detailed data description to follow. The variables that we have in our data set include, but is not limited to, the following: wage and salary income, business income, capital income, unemployment benefits at the individual level for both males and females, household labor income, household business income, household total income, household after-tax after-transfer income, household consumption (constructed from detailed wealth and income data).

2.2 Sample Selection

We follow a similar empirical methodology with Guvenen *et al.* (2015, 2014). Our *base sample* is a revolving panel that maximizes sample size (important for precise computation of higher-order moments for finely defined groups) and allows a stable age structure over time. It is constructed as follows. First, an individual-year earnings observation must satisfy two criteria to be *admissible* for that year: the individual (i) must be between 25 and 60 years old (working age) and (ii) must have participated in the labor market—that is, have earnings (sum of wage/salary income and two thirds of business income/self employment income) above $Y_{\min,t}$. The threshold $Y_{\min,t}$ is chosen as the annual earnings level corresponding to one quarter of full-time work (13 weeks at 40 hours per week) at half of the legal minimum wage. This condition is fairly standard in the income dynamics literature and ensures that we select individuals with a reasonably strong labor market attachment (see, e.g., Abowd and Card (1989) and Meghir and Pistaferri (2004)).

For each year t , we will construct moments by conditioning on individuals' average earnings from $t - 5$ to $t - 1$ and compute statistics based on income changes from year t to $t + k$ ($k = 1, 5$); see Figure 1. The revolving panel for year t selects individuals for which this conditioning can be done in a sensible way. Thus, an individual is in the *base sample* if he

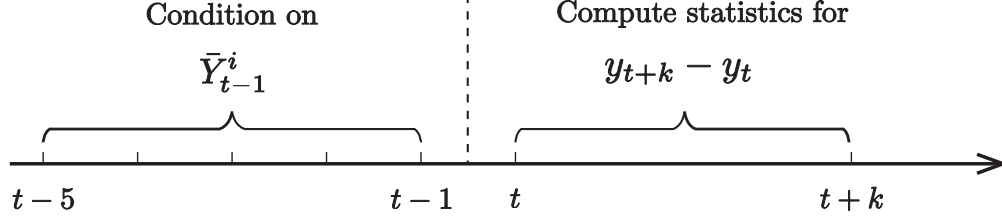


Figure 1: Timeline for Rolling Panel Construction

is admissible in $t - 1$ and in at least two more years between $t - 5$ and $t - 2$. This ensures that the individual was actively in the labor market and has earnings records from which we can compute a measure of average recent earnings, as we describe in a moment. For some statistics, such as change from y_t to y_{t+k} , we also require the individual to be admissible in t and $t + k$. This is specified later below.

Recent Earnings. We now define a measure of what we call “recent earnings,” a term used throughout the paper. Let $\tilde{y}_{t,h}^i \equiv \log(\tilde{Y}_{t,h}^i)$ denote the log earnings (sum of wage/salary earnings and two thirds of business income) of individual i who is h years old in year t . (We will suppress dependence on age whenever it does not create confusion.) For a given worker, we compute his average earnings between years $t - 1$ and $t - 5$. We set earnings observations below $Y_{\min,t}$ to the threshold for this computation. We also control for age effects as follows. We first estimate age dummies, denoted d_h , by regressing log individual earnings on a full set of age and (year-of-birth) cohort dummies. We then construct five-year average earnings in the population from ages $h - 5$ to $h - 1$: $\sum_{s=1}^5 \exp(d_{h-s})$. Finally, we normalize the worker’s average past earnings with this measure to clean age effects. Thus, our measure of *recent earnings* (hereafter, RE) is

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

We will we first group individuals in the base sample into five-year age bins based on their age in year $t - 1$: 25–29, 30–34,..., 50–54, and 55–60. For each year t , we group individuals based on their age and recent earnings as of time $t - 1$, $\bar{Y}_{t-1}^i$. If these groupings are done at a sufficiently fine level, we can think of all individuals within a given age/RE group to be ex

ante identical (or at least very similar).

Growth Rate Measures. Once we construct finely defined groups of workers we can compute the cross-sectional moments of different measures of income changes, between t and $t + 1$ and t and $t + 5$ such as labor income, total individual income (sum of labor income and capital income), after transfer total income (sum of total income and unemployment benefits). It will be useful to distinguish between income growth over short and long horizons, i.e., t and $t + 1$ and t and $t + 5$. We think of these as roughly corresponding to “transitory” and “persistent” earnings shocks.

We consider two measures of earnings growth rates, each with its own distinct advantages. Let $z_t^i = \tilde{z}_{t,h}^i - d_h^z$ denote log type- z income (which could be labor income, total income, etc.) net of age effects of type- z income, where $\tilde{z}_{t,h}^i$ is simply the log type- z income.¹ The first growth measure is the familiar

$$\text{log change: } \Delta_k z_t^i \equiv (z_{t+k,h+k}^i - z_{t,h}^i) = (\tilde{z}_{t+k,h+k}^i - d_{h+k}^z) - (\tilde{z}_{t,h}^i - d_h^z).$$

This measure is well known, and its higher-order moments for a lognormal distribution are familiar to readers (zero skewness and a kurtosis coefficient of 3). While this familiarity makes it a good choice for the nonparametric and descriptive analysis, it also has a well-known drawback: because log of zero is $-\infty$, we need to drop observations close to zero so as to obtain sensible statistics. Thus, when we use $\Delta_k z_t^i$ to compute a statistic, we drop individuals from the base sample whose data are not admissible (type- z income below $Y_{\min,t}$) in t or $t + k$. This forces us to ignore valuable information in the extensive margin. Thus, we borrow a second and closely related measure, which is commonly used in the firm dynamics literature (see, e.g., ?) where firm entry and exit are key margins and lead to the same

¹We estimate age dummies, denoted d_h^z , by regressing log type- z income on a full set of age and (year-of-birth) cohort dummies.

difficulties with logs. The measure is

$$\text{arc-percent change: } \Delta_{\text{arc}} Z_{t,k}^i = \frac{Z_{t+k}^i - Z_t^i}{(Z_{t+k}^i + Z_t^i)/2},$$

where Z_t^i is the level of type- z income. This measure allows computation of time differences even when the individual has zero income in one of the two years, thereby also capturing the extensive margin of earnings changes.

Please note that we keep our conditioning income variable, which is recent earnings $\bar{Y}_{t-1}^i$, same for all measures of income (i.e., for any type- z income) whose future growth rate we compute.

Household-Level Income We can apply the above methodology to the household-level income and household consumption.

3 Results

3.1 The Distributions of Individual Labor Income Growth in Norway and the U.S. are Qualitatively Similar

In this section we study moment of male individual labor earnings in Norway and compare to those documented for the U.S. by Guvenen *et al.* (2015, 2014). We also report the moments for females.

3.1.1 Second Moment: Variance of Earnings Growth

To illustrate how the dispersion of earnings shocks vary over the life-cycle Figure 2 plots the standard deviation of one-year and five-year earnings growth of Norwegian males by age and recent earnings (hereafter, RE) groups (as defined in Section 2.2). The following patterns hold true for both short- and long-run growth rates. First, for every age group,

there is a pronounced U-shaped pattern by RE levels, implying that earnings changes are less dispersed for individuals with higher RE up to about the 90th percentile (18th quantile along the x-axis). This pattern reverts itself inside the top 10% as dispersion increases rapidly with recent earnings. Second, over the life cycle, the dispersion of shocks declines monotonically up to about age 50 (with the exception of very top earners) and then rises slightly for middle- to high-earning individuals from ages 50 to 55.

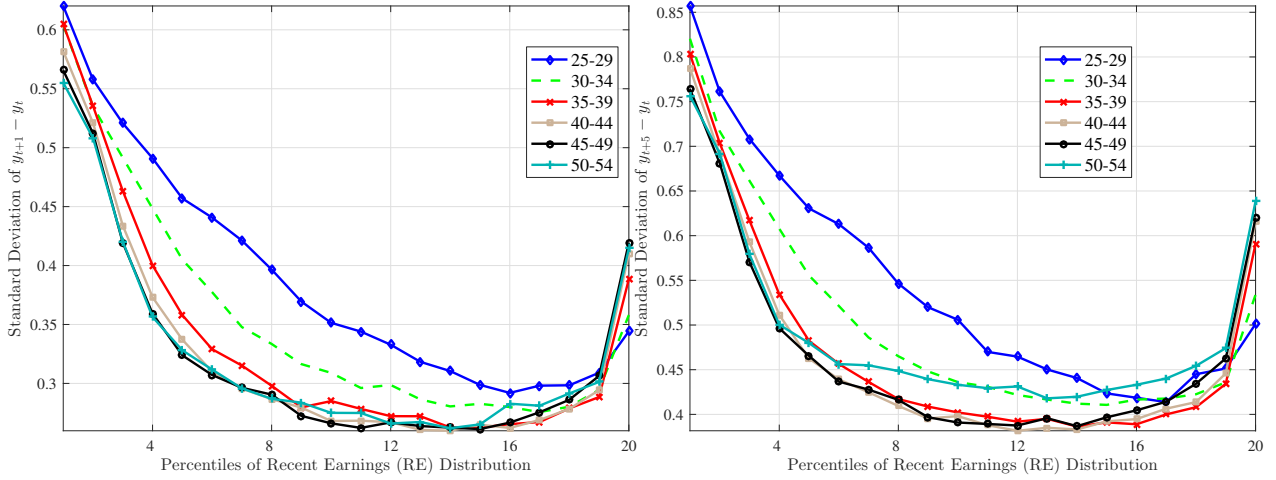


Figure 2: Standard Deviation of Male Earnings Growth in Norway

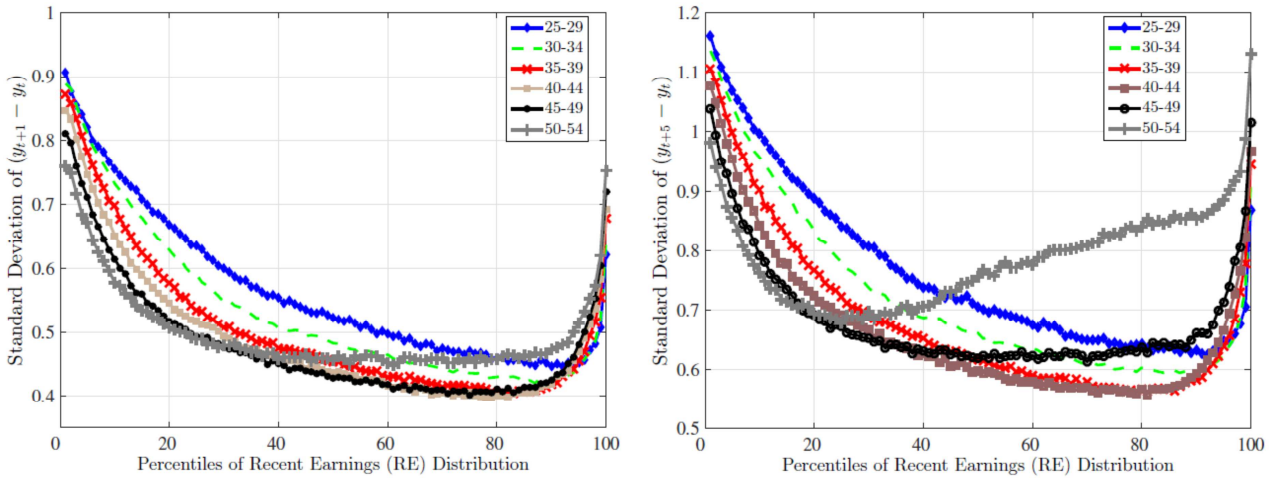


Figure 3: Standard Deviation of Male Earnings Growth in the U.S. (Figure from Guvenen *et al.* (2015))

Figure 3, taken from Guvenen *et al.* (2015) displays a qualitatively identical picture of

the standard deviation of earnings growth of U.S. males by age and recent earnings. The difference between Norway and the U.S. is that the standard deviations are generally larger in the U.S. Figure 18 in the Appendix displays the standard deviation of earnings growth for Norwegian women. The results are both qualitatively and quantitatively quite similar to those for Norwegian men. One difference is that the U-shape with higher standard deviations for the highest earners is less pronounced for women.

3.1.2 Third Moment: Skewness

With the assumption of log-normality of earnings shocks, which is commonly used in the literature, the skewness of is zero. Figure 4 plots the skewness, measured here as the third standardized moment, of one-year (left) and five-year (right) earnings growth for Norwegian males. The first point to observe is that almost every part of every graph in both panels lie below the zero line, indicating that earnings changes are negatively skewed at every stage of the life cycle and for all earnings groups. The second point is that skewness is increasingly more negative for individuals with higher earnings and as individuals get older. Thus, it seems that the higher an individual's current earnings, the more room he has to fall and the less room he has left to move up. And this is true for both short-run and long-run earnings changes. Curiously, and as was the case with the standard deviation, the life-cycle pattern in skewness becomes weaker at the top of the earnings distribution.

Figure 5 displays the skewness of one-year (left) and five-year (right) earnings growth for U.S. males. The pictures are both qualitatively and quantitatively very similar to those for Norwegian males. Figure 19 in the Appendix displays the skewness of earnings growth for Norwegian females. The negative skewness is even more pronounced for women but varies similarly to that of men over age- and earnings group

Another measure of asymmetry is provided by Kelly's measure of skewness, which is defined as:

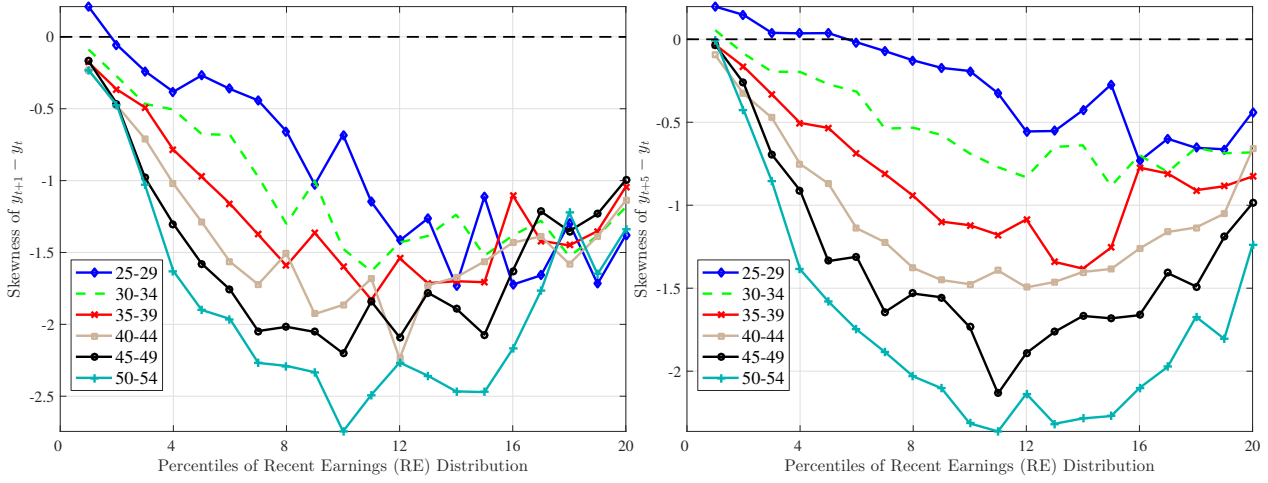


Figure 4: Skewness of Male Earnings Growth in Norway

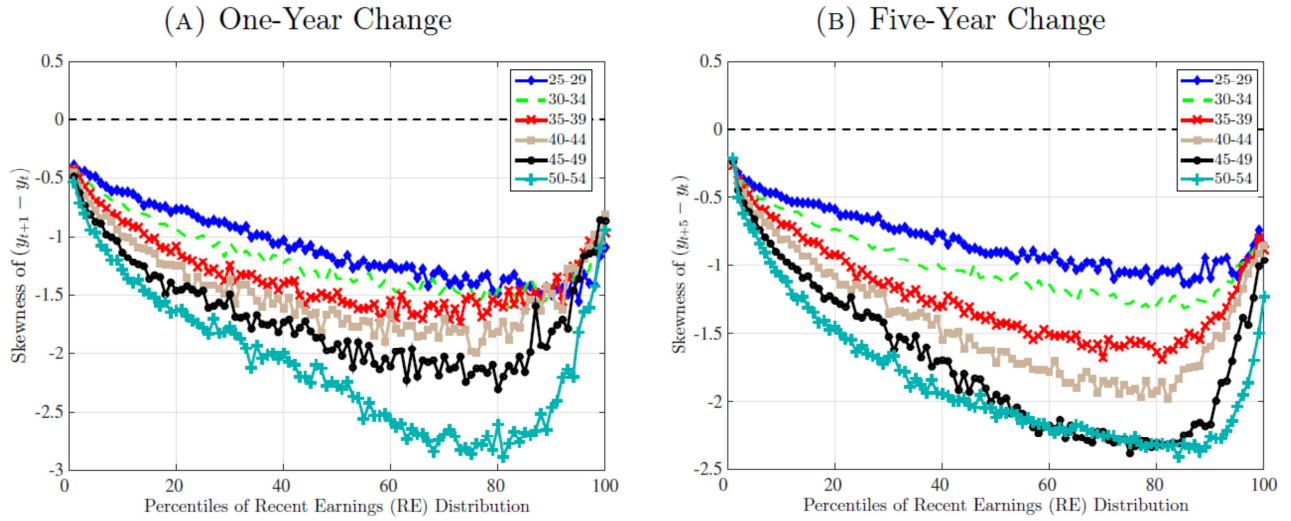


Figure 5: Skewness of Male Earnings Growth in the U.S. (Figure from Guvenen *et al.* (2015))

$$S_k = \frac{(P_{90} - P_{50}) - (P_{50} - P_{10})}{P_{90} - P_{50}} \quad (1)$$

where P_{xy} refers to percentile xy of the distribution under study. Basically, S_k measures the relative fractions of the overall dispersion ($P_{90} - P_{10}$) accounted for by the upper and lower tails. An appealing feature of Kelly's skewness relative to the third standardized moment is that a particular value is easy to interpret. A negative value of SK implies that the lower tail

(P50-P10) is longer than the upper tail (P90-P50), indicating negative skewness. Another property of Kelly's measure is that it is less sensitive to extremes (above the 90th or below the 10th percentile of the shock distribution). Instead, it captures the shift in the weight distribution in the middling section of the shock distribution, whereas the third moment also puts a large weight on the relative lengths of each tail.

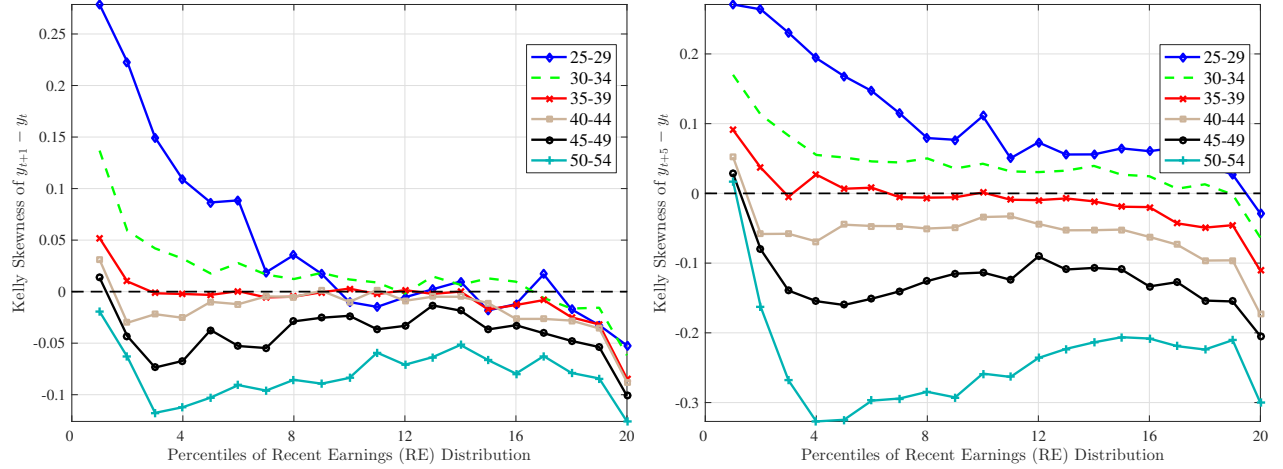


Figure 6: Kelly Skewness of Earnings Growth in Norway
(A) One-year Change (B) Five-Year Change

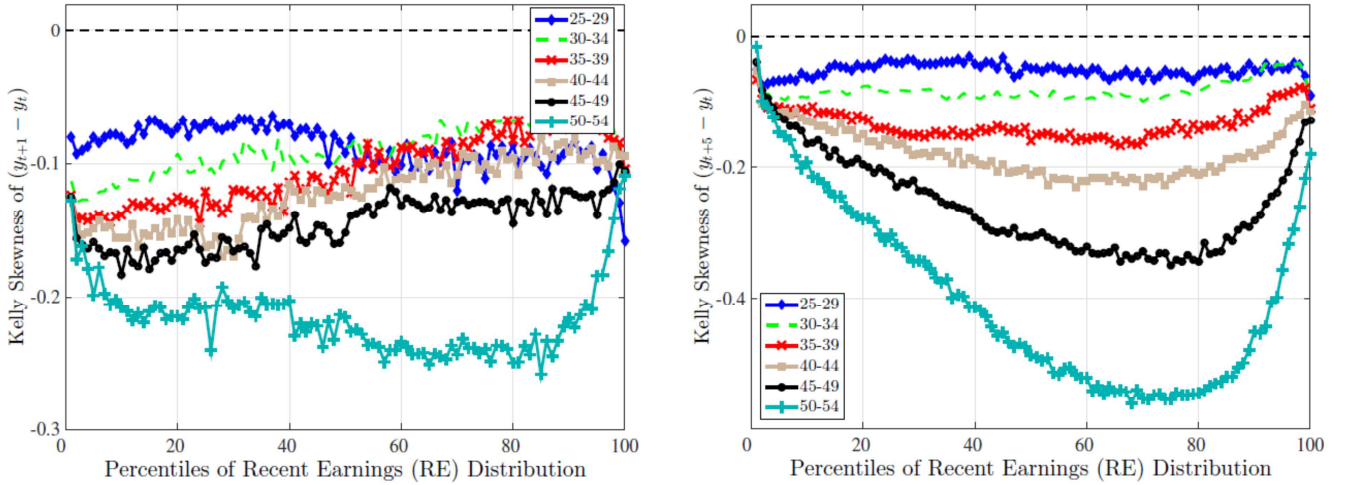


Figure 7: Kelly Skewness of Earnings Growth in the U.S. (Figure from Guvenen *et al.* (2015))

In Figures 6 and 7 we plot the Kelly skewness of labor income growth for Norwegian and U.S. males by age and income percentile. The Kelly skewness is negative for most age and income groups in Norway and all age and income groups in the U.S. The Kelly skewness is

generally slightly larger for U.S. men than for Norwegian men. Figure 20 shows that the Kelly skewness is somewhat more negative for women than for men in Norway.

3.1.3 Fourth Moment: Kurtosis

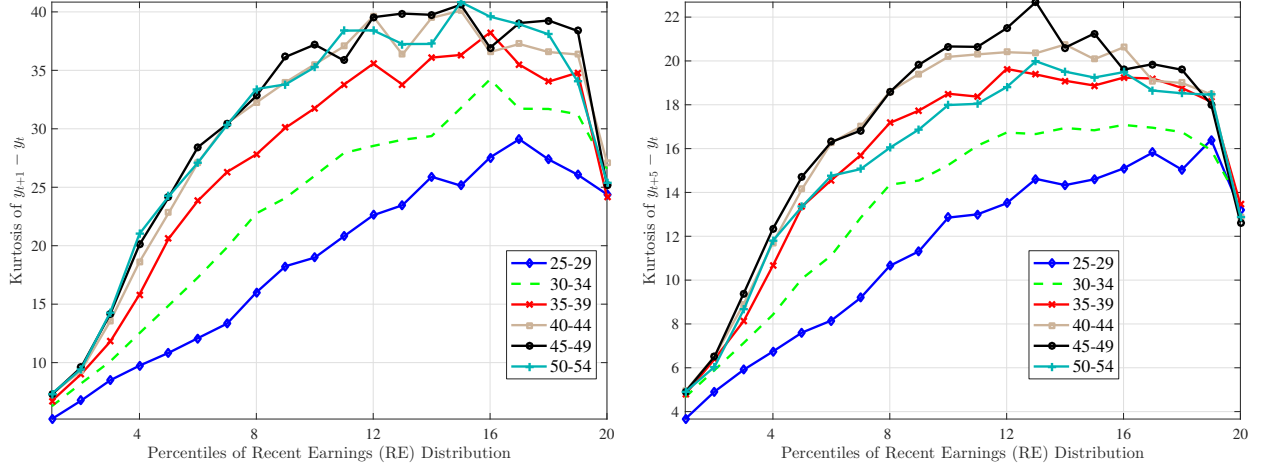


Figure 8: Kurtosis of Earnings Growth in Norway

(A) Annual Change

(B) Five-Year Change

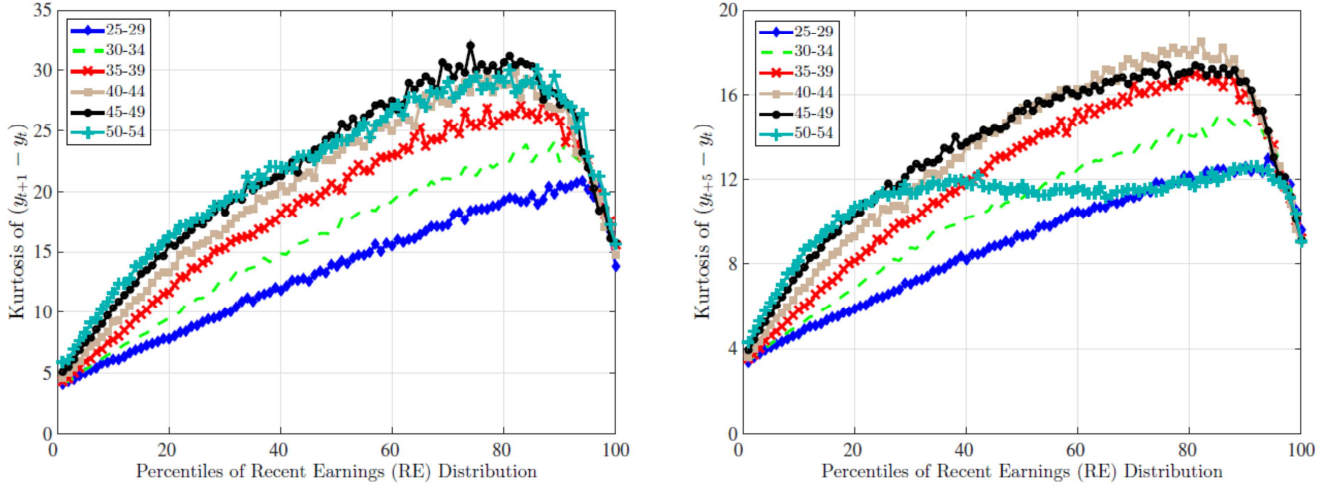


Figure 9: Kurtosis of Earnings Growth in the U.S. (Figure from Guvenen *et al.* (2015))

3.2 The Negative Skewness of Labor Income Shocks is Driven by Work Hours

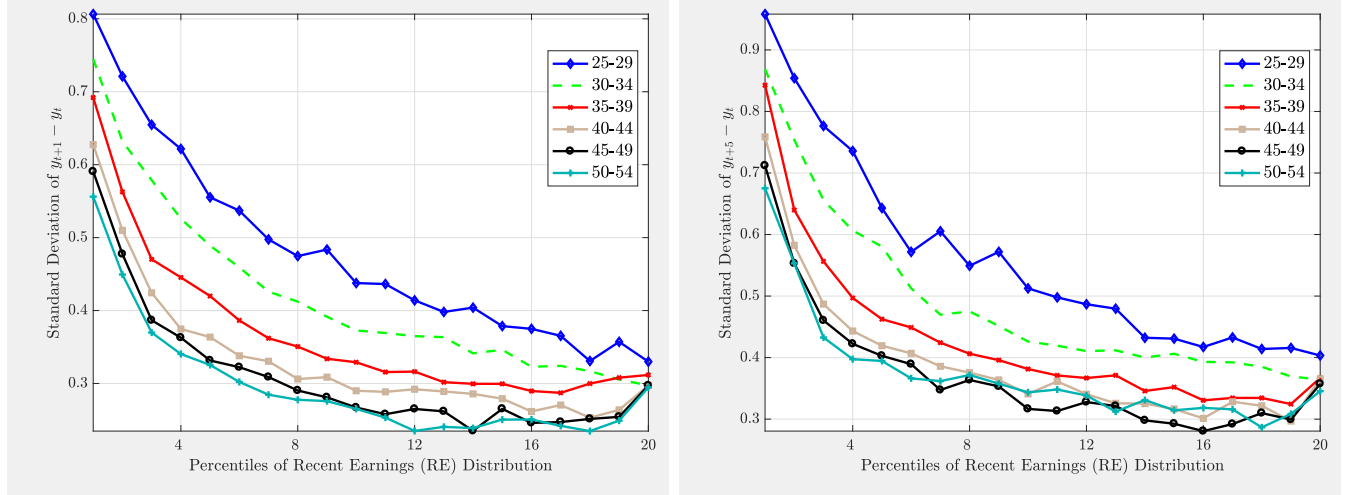


Figure 10: Standard Deviation of Male Hours Growth in Norway

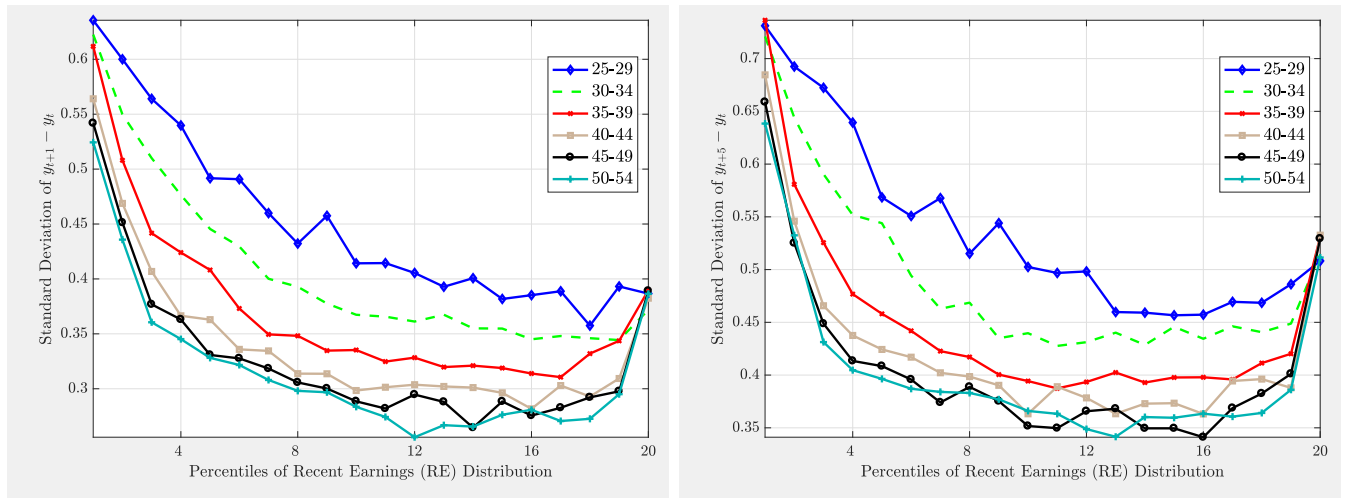


Figure 11: Standard Deviation of Male Hourly Wages Growth in Norway

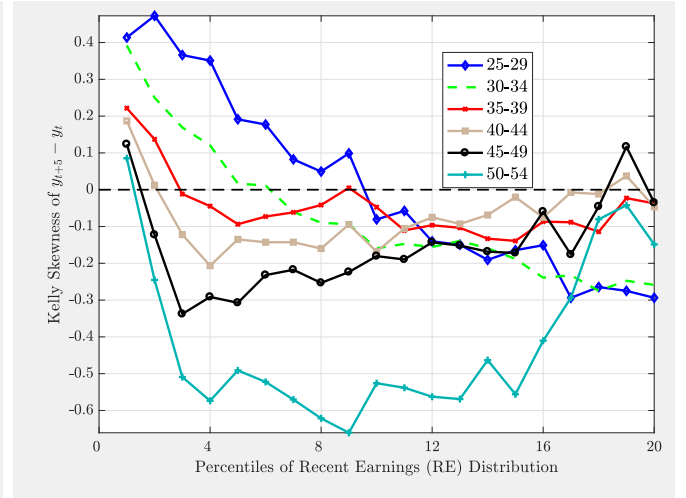
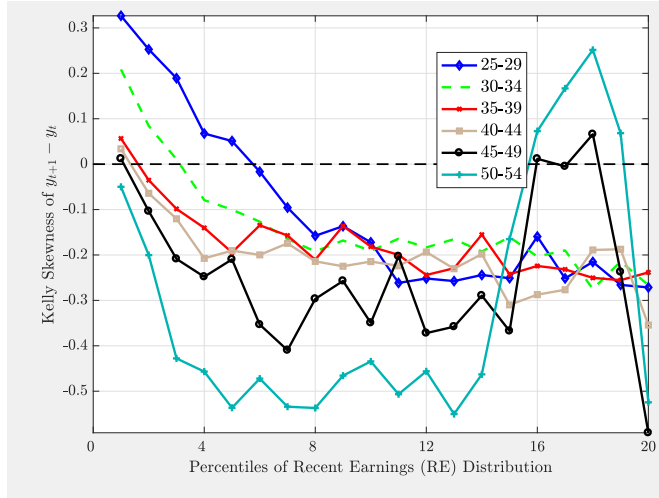


Figure 12: Kelly Skewness of Male Hours Growth in Norway

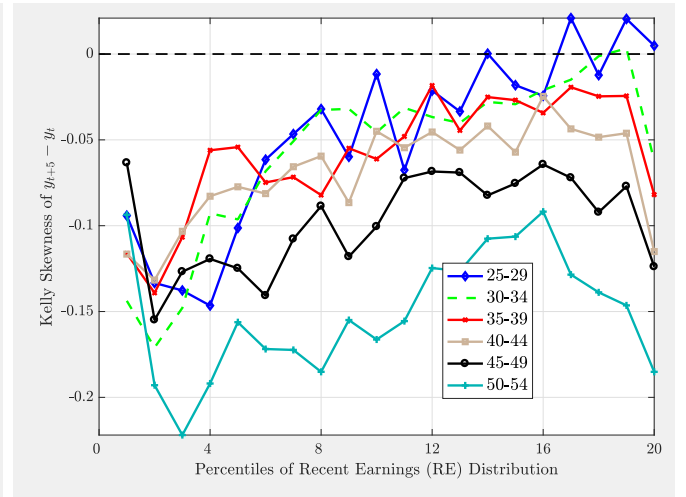
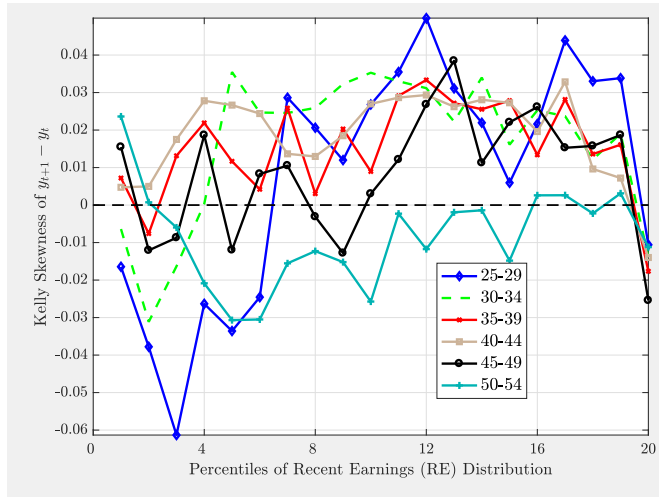


Figure 13: Kelly Skewness of Male Hourly Wage Growth in Norway

3.3 The Welfare State Works: How Transfers Help Remove the Negative Skewness of Labor Income Shocks

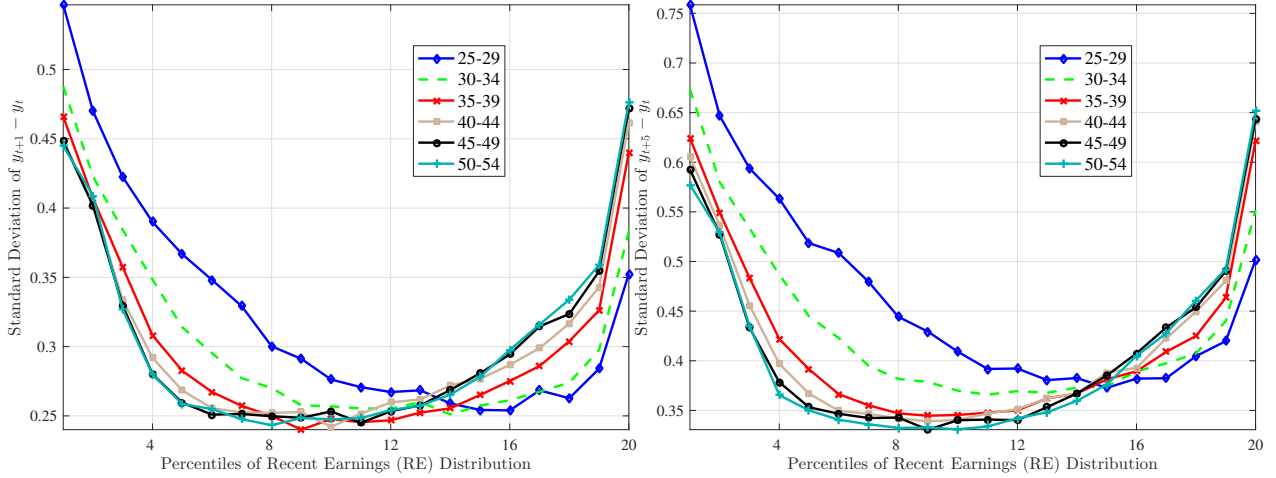


Figure 14: Standard Deviation of Male Gross Income Growth in Norway

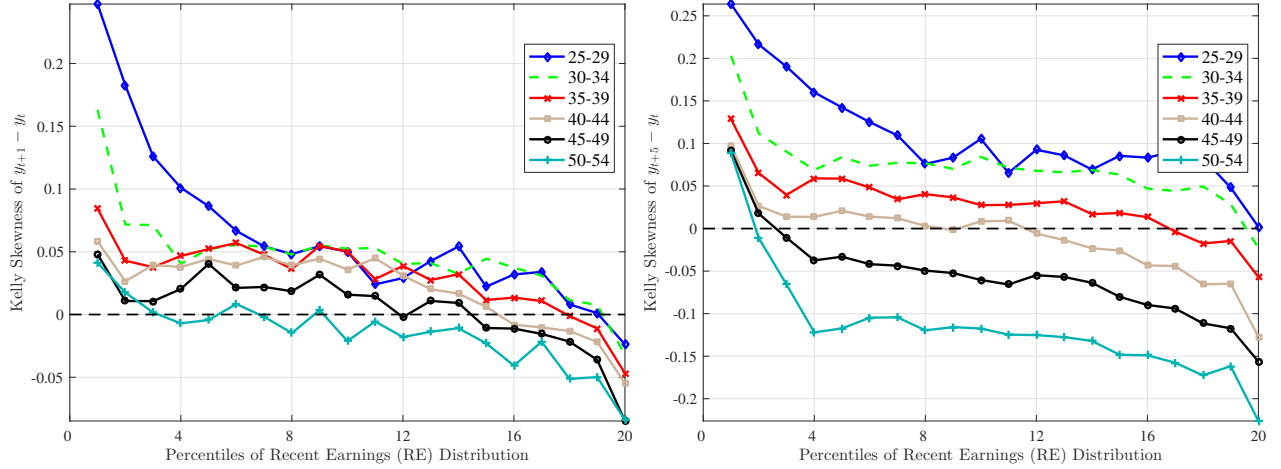


Figure 15: Kelly Skewness of Male Gross Income Growth in Norway

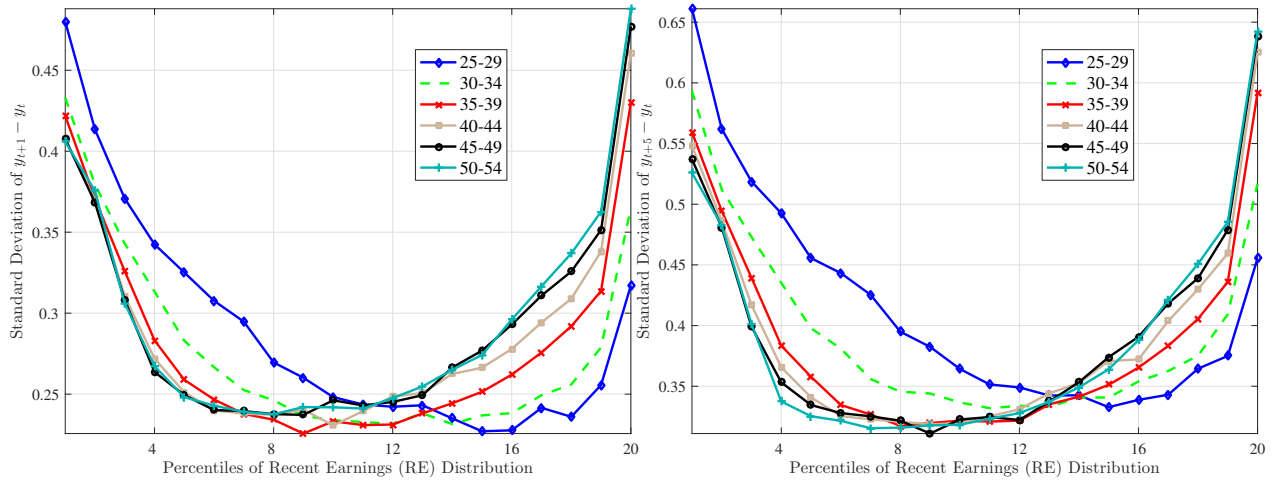


Figure 16: Standard Deviation of Male After Tax Income Growth in Norway

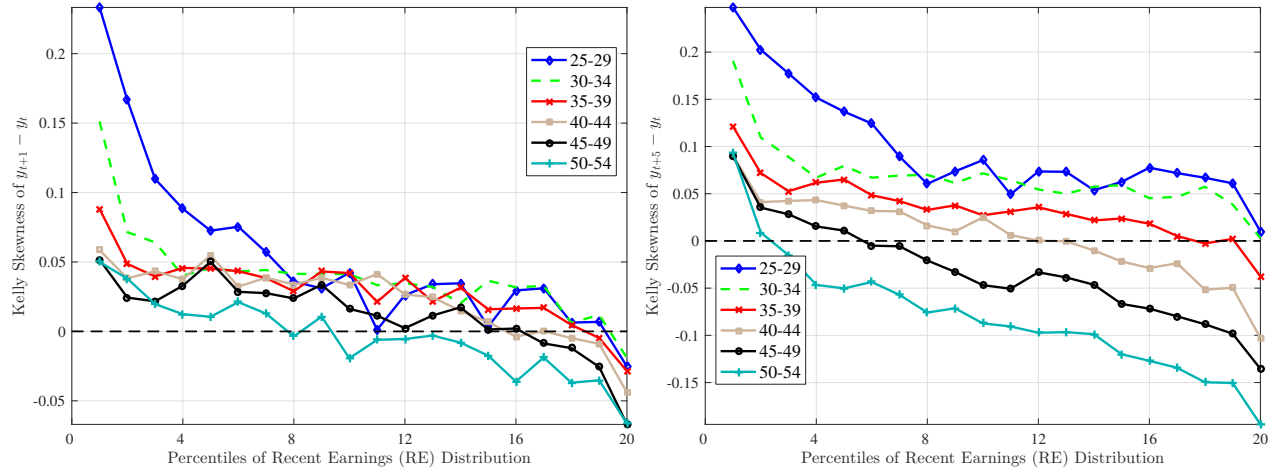


Figure 17: Kelly Skewness of Male After Tax Income Growth in Norway

3.4 The Distributions of Household Labor Earnings and Disposable Incomes

3.5 The Distribution of Consumption Shocks

3.6 Consumption Insurance Towards Income Shocks

4 Appendix

4.1 Additional Figures and Tables

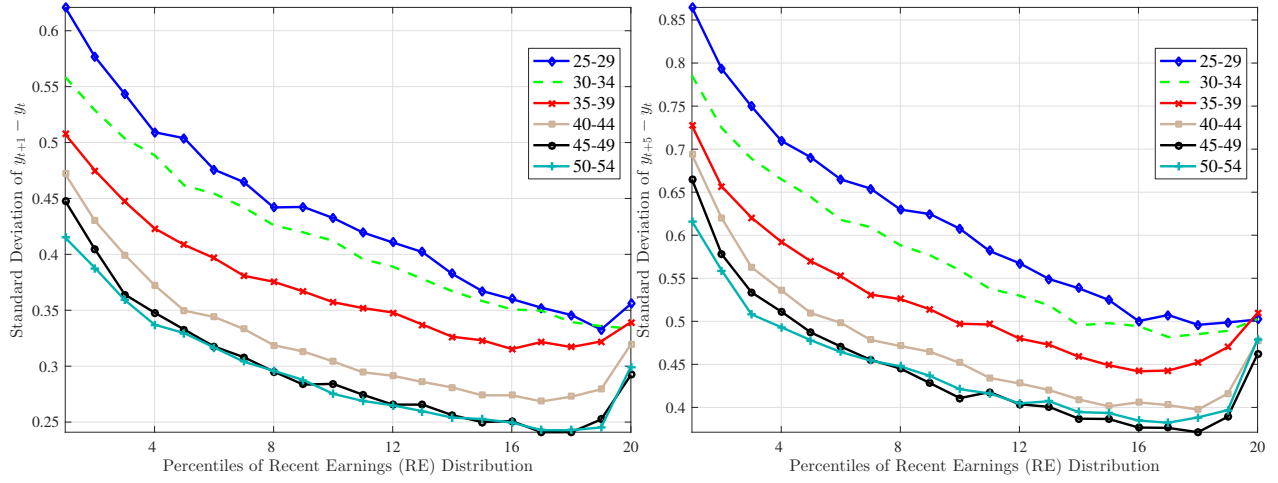


Figure 18: Standard Deviation of Female Earnings Growth in Norway

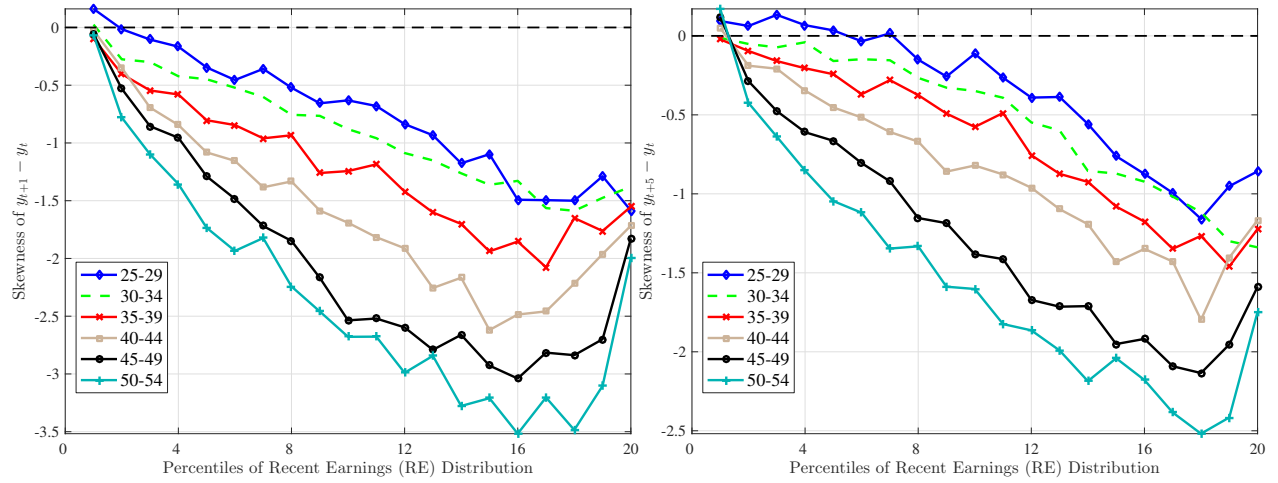


Figure 19: Skewness of Female Earnings Growth in Norway

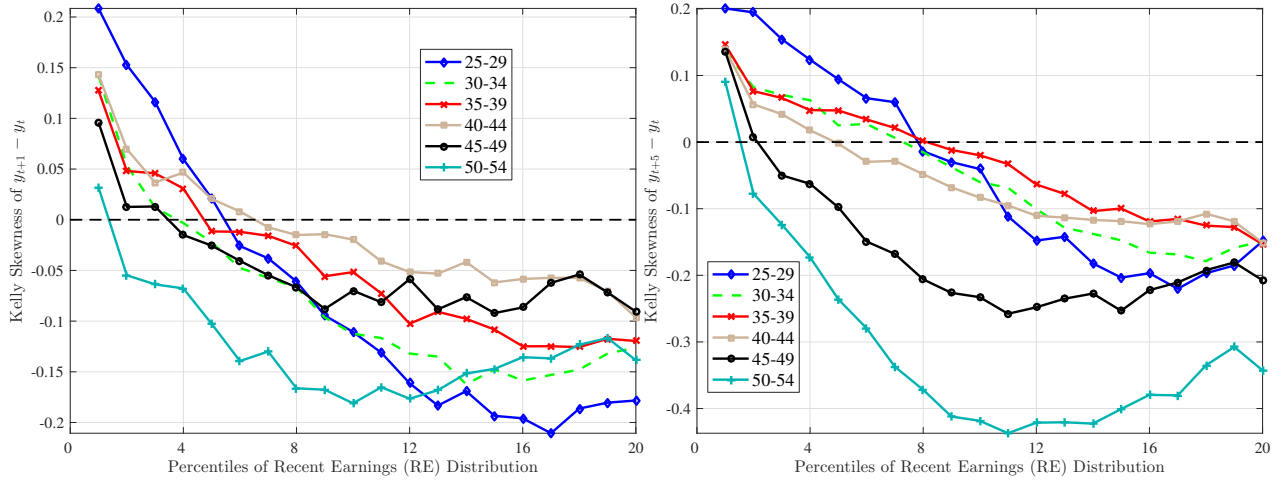


Figure 20: Kelly Skewness of Female Earnings Growth in Norway

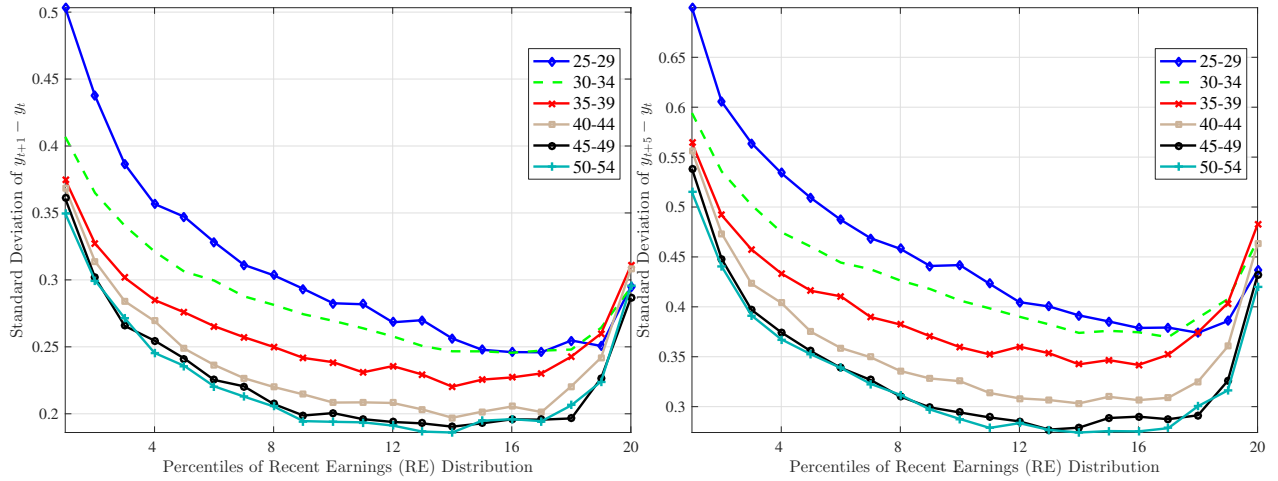


Figure 21: Standard Deviation of Female Gross Income Growth in Norway

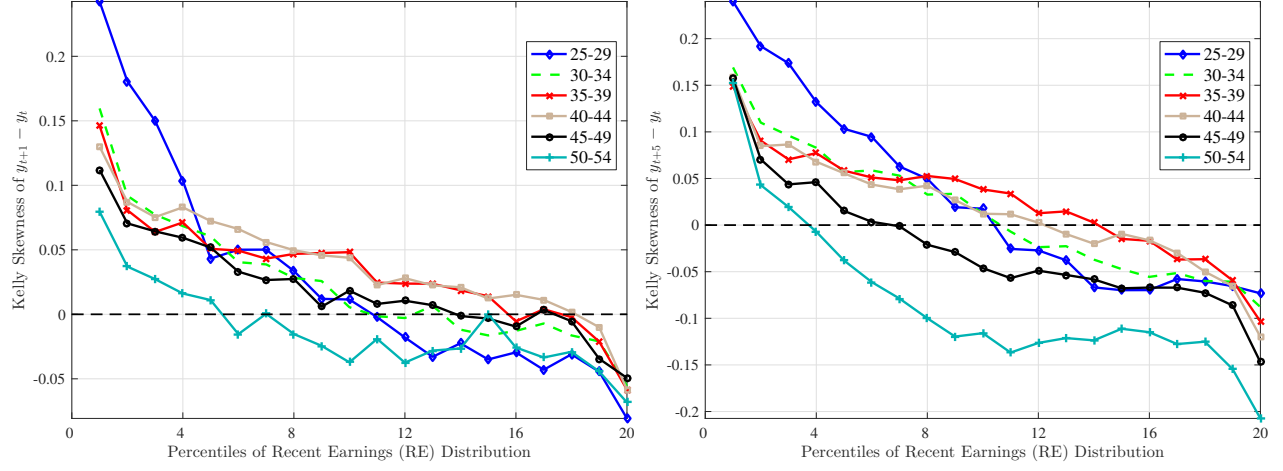


Figure 22: Kelly Skewness of Female Gross Income Growth in Norway

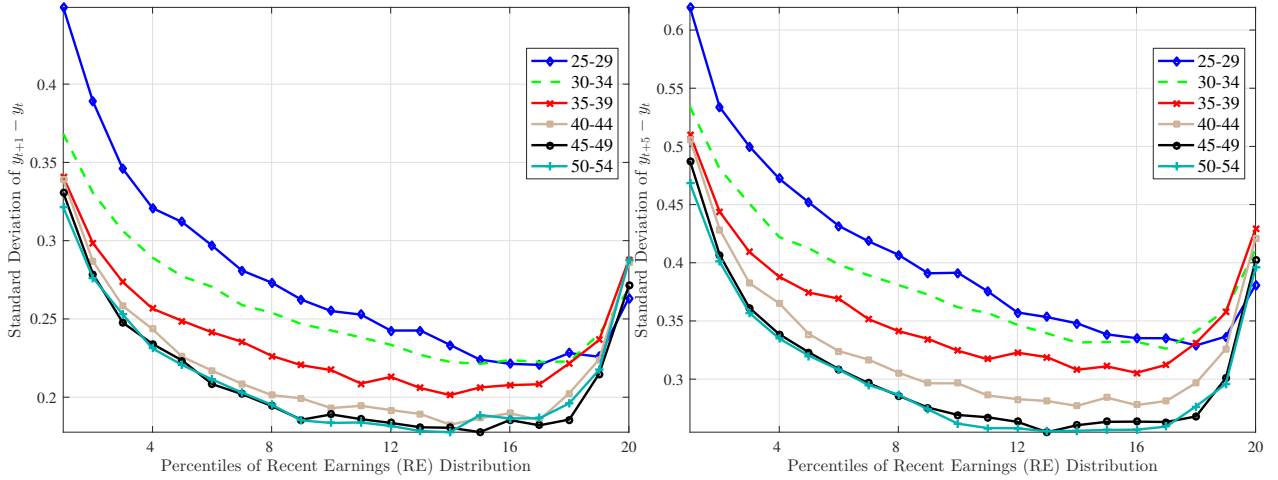


Figure 23: Standard Deviation of Female After Tax Income Growth in Norway

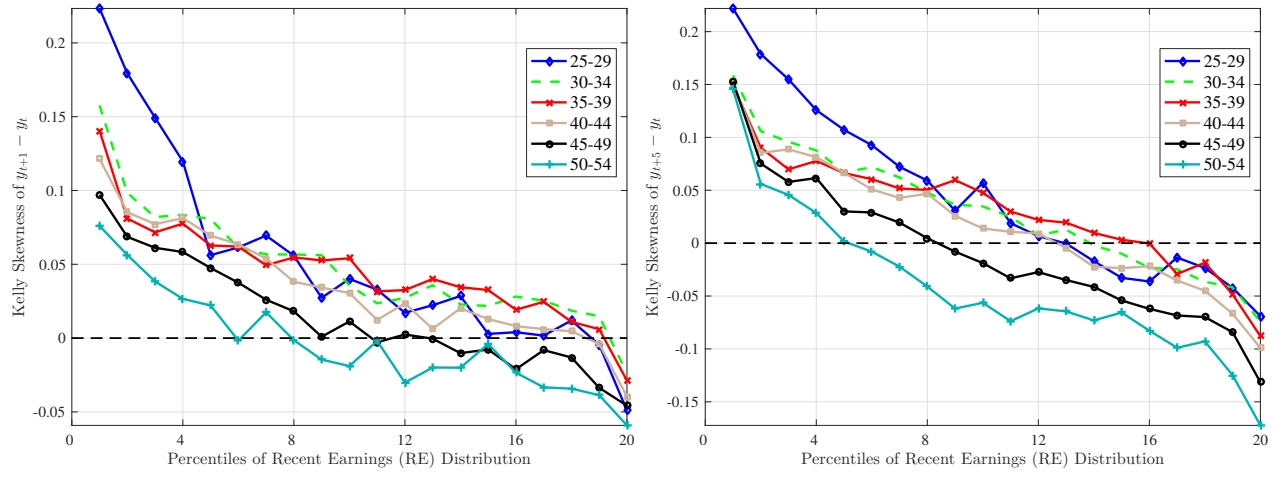


Figure 24: Kelly Skewness of Female After Tax Income Growth in Norway

References

- ABOWD, J. M. and CARD, D. (1989). On the covariance structure of earnings and hours changes. *Econometrica*, **57** (2), 411–45.
- FAGERENG, A., HALVORSEN, E. *et al.* (2015). *Imputing consumption from Norwegian income and wealth registry data*. Tech. rep.
- GUVENEN, F., KARAHAN, F., OZKAN, S. and SONG, J. (2015). *What Do Data on Millions of U.S. Workers Say About Labor Income Risk?* Working Paper 20913, National Bureau of Economic Research.
- , OZKAN, S. and SONG, J. (2014). The nature of countercyclical income risk. *Journal of Political Economy*, **122** (3), 621–660.
- MEGHIR, C. and PISTAFERRI, L. (2004). Income variance dynamics and heterogeneity. *Econometrica*, **72** (1), 1–32.