CENTER FOR
ECONOMIC
BEHAVIOR &
INEQUALITY

CEBI WORKING PAPER SERIES

Working Paper 01/23

SUBJECTIVE EARNINGS RISK

Andrew Caplin

Victoria Gregory

Eungik Lee

Søren Leth-Petersen

Johan Sæverud

ISSN 2596-447X

CEBI
Department of Economics
University of Copenhagen
www.cebi.ku.dk

Subjective Earnings Risk

Andrew Caplin
New York University

Victoria Gregory
FRB St. Louis

Eungik Lee
New York University

Søren Leth-Petersen
University of Copenhagen

Johan Sæverud
University of Copenhagen

January 3, 2024

Abstract

We introduce a survey instrument to measure earnings risk allowing for the possibility of quitting or being fired from the current job. We find these transitions to be the key drivers of subjective risk. A link with administrative data provides multiple credibility checks for correspondingly aggregated data. Yet it reveals subjective earning risk to be many times smaller than traditional estimates imply even when conditioning richly on demographics and job history. A life-cycle search model calibrated to match data on job transitions and earnings can replicate the distribution of subjective beliefs reported in the survey. Job-match quality, which directly impacts subjective risk but is impossible to identify in administrative data, contributes significantly to earnings risk. This highlights the importance of administratively-linked subjective risk measures.

Keywords: earnings risk, job transitions, subjective expectations

JEL classification: D31, D84, E24, J31

1 Introduction

Subjective earnings risk impacts search and labor market decisions, savings/insurance choices, and inequality. Yet it has proven challenging to measure. We introduce a survey instrument to measure this risk, which we define as the spread of an individual's probability distribution over future earnings.[1] We make explicit allowance for job transitions: both the possibility of quitting and the risk of being fired from the current job. We find these transitions to be the key drivers of subjective earnings risk. A link with administrative data provides multiple credibility checks, but also reveals subjective risk to be many times smaller than traditional estimates based on the cross sectional distribution of realized earnings growth, imply. A life-cycle search model calibrated to match data on job transitions by tenure and the level of earnings can replicate the distribution of subjective beliefs reported in the survey. The model suggests that differences in job-match quality, which are very hard if not impossible to identify in administrative data, contribute significantly to earnings risk. This indicates the importance of administratively-linked subjective measures such as ours.

One key finding from the survey is the high level of heterogeneity in peoples' beliefs about their future earnings. Respondents vary in terms of their probabilities of job separation, how long they expect to be out of work if they separate from their jobs, and in terms of their probability distribution over future wage growth conditional on job separation. The link to the administrative data reveals that much of this heterogeneity survives even after accounting for demographics and earnings histories. The match with the administrative data does however provide many broad credibility checks when correspondingly aggregated up. Average job-separation rates calculated from the administrative data are very similar to the average job-separation probabilities in the survey, even when stratified by age groups. Similarly, average expected time out of work following job-separation compares nicely with actual time spent out of work as it is recorded in administrative data. Furthermore, based on the answers to the survey instrument, we construct a measure of the overall subjective probability distribution over one-year ahead earnings growth and show that the mean of this lines up tightly with realized earnings growth for the period that the expectations concern. Finally, we pool the subjective distributions over one-year ahead earnings growth for the entire survey sample and compare with the cross sectional

[1] In the literature estimating income processes, the year-to-year earnings volatility, including earnings fluctuations that may have been the result of a choice and that cannot be explained by observable characteristics, is often interpreted as risk. This amounts to assuming that workers receive a risky but exogenous flow of earnings in each period. The well-known permanent-transitory income process fitted to administrative data falls within this category.

distribution of realized earnings growth that we observe in the administrative data. Also here we find that the distribution constructed from the survey data aligns closely with the distribution of realized earnings growth in the administrative data. In sum, the survey answers agree closely with third-party reported administrative data in a sequence of validation exercises, thus giving credence to the quality of the survey.

Despite these matches when the survey data are averaged, we find that the subjective earnings risk at the individual level, measured as the interdecile range of the subjective probability distribution over future earnings growth, is two to six times smaller than would be estimated in the administrative data. This is the case even when taking into account observable characteristics in a very detailed nonparametric fashion. The reason for this is that the cross sectional distribution of earnings growth confounds heterogeneity in expected subjective earnings growth and actual subjective earnings risk. This suggests that individuals have more information than an econometrician who only observes administrative data on realized earnings growth across the population and therefore cannot separate risk from heterogeneity. Our unique combination of survey and administrative data allows us to confirm this and to quantify how much administratively inferred earnings risk overstates subjective risk. By documenting that respondents' expected earnings growth align with their actual subsequent earnings growth, as shown above, we confirm that individuals have superior information.

The survey design allows us to examine the role of job transitions for subjective probability distributions over one-year ahead earnings growth. We show that higher order moments are important for characterizing these distributions. In particular, the typical subjective distribution is negatively skewed and displays excess kurtosis that is increasing in age. When we consider only the probability distributions over the contingency where people stay in the current job then the typical subjective spread is reduced and there is no skewness nor excess kurtosis. In other words, we find that job transitions are responsible for most of the spread that we observe in subjective probability distributions over future earnings including higher order moments.

We organize our key findings on subjective earnings risk in a model of search over the life cycle due to Menzio et al. (2016)[2] The model is an appealing benchmark for us because it emphasizes job mobility, which our survey points out is critical for earnings risk.[3] We

[2] This model is a life-cycle extension of the directed search model of Menzio and Shi (2011).

[3] There is an extensive literature of search models with detailed characterizations of heterogeneity on both the firm and worker side, including Low et al. (2010), Altonji et al. (2013), Moscarini and Postel-Vinay (2013), Bagger et al. (2014), and Bagger and Lentz (2019). Especially relevant to our study are Hubmer (2018), Jung and Kuhn (2019), and Karahan et al. (2022) which have explicit focuses

calibrate the model to match administrative data on job transitions by levels of tenure and the level of earnings, and show that it is able recover, to a first order approximation, the distribution of beliefs that we observe in the survey data. The reason is that the search model is explicitly designed to describe job transitions over the life cycle. Our earlier finding showed that the survey expectations concerning job transitions and time out of work match the actual frequency and time out of work in the administrative data, meaning that the model matches them as well. Because we also showed that these factors summarize the most important aspects of subjective earnings risk, it implies that the model calibrated to realized job transitions by tenure and the level of earnings can also capture the basic features about subjective beliefs, including subjective earnings risk. We conclude that in order to capture the beliefs about earnings risk documented from our data, one should explicitly model the search process and account for job transitions, rather than relying on a reduced-form income process identified off of the variance of earnings growth in administrative data

The model provides a structural interpretation of the heterogeneity in subjective beliefs that people report in the survey. Like in the survey data, agents in the model differ in their beliefs over future earnings growth even conditional on age and earnings. The model attributes this to match quality, a latent variable that is not observable in administrative data. Match quality is a determinant of future transitions and consequently also a determinant of subjective beliefs over future earnings growth. Workers in high quality job-matches correctly expect that they will stay there longer, whereas workers in new or lower quality matches expect to make a transition sooner to a better job match. Workers' knowledge about match quality is thus a potential source of heterogeneity in expected earnings growth that goes beyond what can be explained by basic observable characteristics like age and the level of earnings. In this way the model rationalizes that individuals have superior information about their expected earnings growth, and this heterogeneity can potentially explain how subjective earnings risk is many times smaller than risk inferred from administrative data on earnings growth alone because the latter confounds heterogeneity in expected earnings growth and subjective risk. The model also rationalizes agents having different risk perceptions. Knowledge about (future) match quality varies across agents and, consequently, so does the variability of future earnings growth, i.e., subjective risk.

Our paper builds on recent advances in our understanding of earnings dynamics and

on earnings risk over the life cycle. These models all incorporate the same mechanisms as our model where people face the risk of job separation because the job-match quality is unknown when entering a new job-match.

earnings risk measured in large administrative data (e.g., Guvenen et al., 2021).[4] This literature shows the importance of higher-order moments in the distribution of realized earnings growth, heterogeneity in earnings dynamics across ages and levels of earnings, and the critical role of job transitions and periods out of the labor force for earnings risk, all of which are mirrored in our subjective data.[5] The other key literature on which we build is the pioneering research of Dominitz and Manski (1997), measuring probabilistic beliefs about one-year ahead earnings, with important subsequent work by De Bruin et al. (2011), Dominitz (1998, 2001), Guiso et al. (2002), and Pistaferri (2001, 2003). In recent work complementary to ours, Koşar and van der Klaauw (2023) and Wang (2023) study wage expectations related to staying in the current job as elicited in the Survey of Consumer Expectations conducted by the New York Fed. Wang finds that subjective wage risk associated with staying in the current job is lower than wage risk inferred from wage realization for job stayers. Our study follow recent research that combine at the individual level subjective information collected from surveys with objective information from administrative data facilitating direct comparison of subjective and objective information (e.g., Andersen and Leth-Petersen, 2021; Hvidberg et al., 2023; Epper et al., 2020). Another related literature studies subjective labor market expectations (Manski and Straub, 2000; Stephens, 2004; Campbell et al., 2007; Hendren, 2017), and in terms of methodology, our work builds on a branch of the literature that measures conditional expectations (e.g. Arcidiacono et al., 2020, Wiswall and Zafar, 2021). Finally, our work is related to a recent literature using subjective expectations data to inform and discipline structural models of labor market dynamics and savings (Conlon et al., 2018; Balleer et al., 2021; Bick et al., 2021; Mueller et al., 2021; Faberman et al., 2022; Jäger et al., 2022; Stoltenberg and Uhlendorff, 2022; Balleer et al., 2023; Wang, 2023).

The paper is organized as follows. Section 2 introduces the conditional earnings survey instrument. Section 3 compares survey responses with linked administrative data. Section 4 compares subjective earnings risk with its administratively-estimated counterpart. Section 5 links this risk with job transitions. Section 6 presents and calibrates a life-cycle search model to administrative data on job transitions, tenure and the level of earnings, compare the implied cross sectional distribution of subjective risk to its counterpart

[4] See also Busch et al. (2022), Druedahl and Munk-Nielsen (2020), Halvorsen et al. (2023), Guvenen et al. (2022) and other papers published as part of the Global Repository of Income Dynamics (GRID) project, `https://www.grid-database.org`.

[5] The prominent role of job transitions for earnings risk, and in particular for higher order moments in the distribution of realized earnings growth, is documented earlier by Low et al. (2010). The importance of job transitions for the level of earnings is also revealed in separate studies of layoffs and quits (Topel and Ward, 1992; Jacobson et al., 1993; Von Wachter et al., 2009).

from the survey, and provide a structural interpretation of the nature of subjective risk. Section 7 concludes.

2 The Conditional Earnings Instrument

In this section we introduce the conditional earnings survey instrument through which respondents are asked in January 2021 about their expectations concerning job transitions and earnings throughout 2021. We first present the branching structure and the survey questions.[6] We then introduce the *Copenhagen Life Panel* in which it was implemented and give a branch-by-branch bird's-eye-view of survey responses. We end by explaining how key variables are constructed and providing a high level overview of quantitative findings.

Figure 1 illustrates the branching structure of our survey and our naming convention for each of the components. Starting from the left, we first ask about the probability of job transitions, i.e., the probability of staying in the current job (p_i^S), the probability of being laid off (p_i^L), and the probability of quitting (p_i^Q). For the layoff and quit branches we then ask about the expected time out of work following the separation (n_i^L, n_i^Q). Finally, we elicit the conditional probability distributions over one-year ahead earnings in each of the three branches. We subtract last year's earnings, which we also ask about in the survey, from this to arrive at branch-specific distributions of growth rates of earnings, and we denote these distributions (f_i^S, f_i^L, f_i^Q). For each respondent we collect all these eight objects.

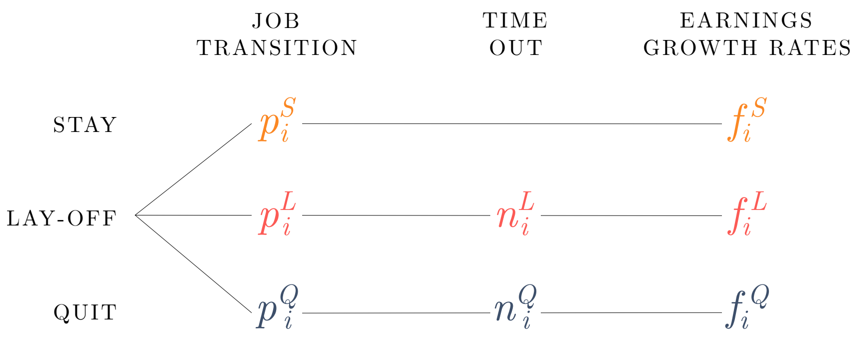
2.1 Job transitions

The expectations instrument opens by asking all respondents who report being employed in January 2021 about the likelihood of job transitions during 2021:

- *Please think about your possible relationship with your current employer in 2021. Assign the probability in each possible case. The sum of the probabilities should be 100.*

    1. *Staying with your current employer during 2021*
    2. *Being laid off from your current employer at some point during 2021*
    3. *Quitting from your current employer at some point during 2021*

[6] The questionnaire is reported in Online Appendix A.



Note: The survey instrument consists of three branches, each representing a job transition (Stay, Layoff, Quit), and three domains for each branch: for each individual i we elicit job transition probabilities, p_i^B, time out of work, n_i^B, and distributions of conditional earnings growth rates, f_i^B, where $B \in \{S, L, Q\}$.

Figure 1: Survey instrument overview

4. *Separating from your current employer for some other reason during 2021*

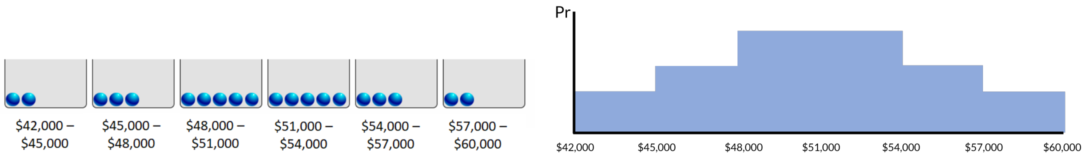
For each individual (i) we denote the branch-specific probability p_i^B, where $B \in \{S, L, Q\}$.

For those who report a positive layoff probability, we follow up by asking about how long they expect to be out of work, and we do this by asking the likelihood of being re-employed within four different horizons: 1, 3, 12, and 24 months:

- *Suppose you were to be laid off from your current employer during 2021. What is the probability that you would start working for pay again within 1/3/12/24 months of termination?*

For those who report a positive probability of quitting during 2021, we ask a similar question, where the probabilities now refer to finding a job within each time horizon after quitting. We use this information to calculate the expected time out of the labor force following a separation (n_i^L and n_i^Q) for each individual. The process is described in detail in Section 2.3.

Finally, we ask each respondent their probabilistic beliefs about future earnings for each of the subjectively possible job transitions. This is straightforward for the stay branch as this is just the uninterrupted continuation of the current job. For the layoff and quit branches, we ask about the earnings in the 12 months following the start of the new job, i.e., the annual earnings taking into account that the new job may begin following a period out of work. Here is the basic design for the case of being laid off from the job during 2021.



(a) Elicitation (b) Distributions

Note: Panel (a) shows the sample screen for the elicitation and Panel (b) shows how we interpret the distribution of the answer in Panel (a) as a mixture of uniform distributions.

Figure 2: Balls in bins

- *Suppose you were to be laid off from the current employer during 2021 and to start to work for pay at some point in the following 2 years. Think about your possible earnings during the first 12 months in this new job*

In order to elicit the full distribution of future annual earnings in each branch we apply the "balls in bins" method developed by Delavande and Rohwedder (2008), which is intuitive and visually oriented.[7] Respondents are first asked to state the minimum and maximum values for possible future earnings, as in Dominitz and Manski (1997). Then the range between the stated minimum and maximum is divided into six equally sized bins. Respondents are then instructed to move 20 balls into the six bins to reflect how likely their future earnings are to fall in each of the ranges defined by the bins. Figure 2(a) illustrates the "balls in bins" task as it appears in the online survey.

We construct branch-specific subjective distributions of earnings based on the answers to these questions. Since there are 20 balls available, we interpret one ball as representing a probability of 5%. We also assume that probabilities are uniformly distributed within each bin. For example, in Figure 2(a) two balls are placed in the first bin and we interpret this to mean that there is a 10% likelihood of realizing earnings in the interval 42,000 to 45,000 DKK (the first bin). Combining all the bins enable us to characterize the entire subjective probability distribution and to calculate various moments for each respondent's conditional distribution. For example, Figure 2(b), shows the distribution that the "balls in bins" answers in panel (a) are converted to. The mean of this distribution is 51,000 and the standard deviation is 4,896.

[7] Goldstein and Rothschild (2014) show that bins and balls elicitation increases the accuracy of reported distribution compared to other non-graphical elicitation methods.

In the survey we ask respondents about last year's earnings.[8] By subtracting this from the conditional distributions collected using the survey instruments outlined above we arrive at branch-specific distributions of growth rates of earnings, which we denote f_i^B, where $B \in \{S, L, Q\}$.

2.2 Copenhagen Life Panel

Our survey instrument is implemented in the newly developed *Copenhagen Life Panel* [9] (CLP) which is an online panel survey implemented in Denmark. We invite a random selection of individuals, who are recorded in the Danish population registry and aged between 20 and 70, to participate in the survey. The population registry is a complete registry of all persons who are born or have ever had an address in Denmark. It contains a personal identifier (CPR-number) applied universally to record any contact an individual has with the public sector. Invitations to participate were sent out using an official email account, called *e-boks*, which all Danes are equipped with. For the purpose of this paper we consider questions about earnings expectations and job transitions that were included in CLP issued in January 2021.[10]

Upon survey completion, answers are linked to the administrative records for all individuals who are invited to the survey as well as the rest of the entire Danish population. These data include standard demographic information, such as age, gender, education, household composition, and household wealth, all collected at the annual frequency. All the administrative data are longitudinal by nature and are currently available up to and including 2021.

For this study we include respondents between age 20 and 65 which is the typical working age span. The gross sample includes 14,875 respondents. We restrict the sample to include 10,945 people who are employed at the time of the survey. This is to make sure that we are not dealing with individuals who are permanently or temporarily out of the labor market. In Online Appendix B, we compare the average earnings, age, gender, and

---

[8] Last years earnings is the earnings realized in 2020. This has a direct counterpart in the administrative data. In Online Appendix C.4 we directly compare the survey responses concerning earnings in 2020 with earnings for 2020 recorded i the income-tax register. The comparison show that survey answers line up very closely with the administrative data.

[9] The *Copenhagen Life Panel* is an ongoing survey that was initiated in 2020 and is issued every year in January.

[10] Those who finished the survey participated in a prize lottery with 50 respondents receiving prizes worth 1,000 DKK (approximately, 140 USD) and one a grand prize of 10,000 DKK (approximately, 1,400 USD).

educational attainment between the survey sample and the full population belonging to the same age groups. There is wide variation across age and earnings in our sample. In comparison with the larger population, the average survey participant is slightly older, more educated, and has a somewhat higher level of earnings. For the subsequent analysis, we apply population weights that we construct from the administrative data.[11]

2.3 Job Transitions

In Figure 3 we present an overview of the answers collected. Starting from the left, the probabilities of job transitions $\bar{p}^B$ represent the average job transition probabilities stated by the respondents. With an average likelihood of $\bar{p}^S = 82\%$, the most likely event is remaining with the current employer, followed by quitting, $\bar{p}^Q = 12\%$ and being laid off, $\bar{p}^L = 6\%$. In Online Appendix C.1, we plot the distributions of each job transition probability.

Moving to the right in Figure 3, we report the average expected time out of work upon quitting or being laid off, $\bar{n}^B$. To arrive at a summary measure of the duration out of work we aggregate over the likelihood of being out of work for the four horizons that we ask about for the quit/layoff branches. Focusing on time out of work following a layoff, respondents report $(n^L_{i,1}, n^L_{i,3}, n^L_{i,12}, n^L_{i,24})$ as their reemployment probabilities within 1, 3, 12, and 24 months, respectively. Assuming that reemployment takes place in the middle of the four time intervals, the expected reemployment period is calculated as:

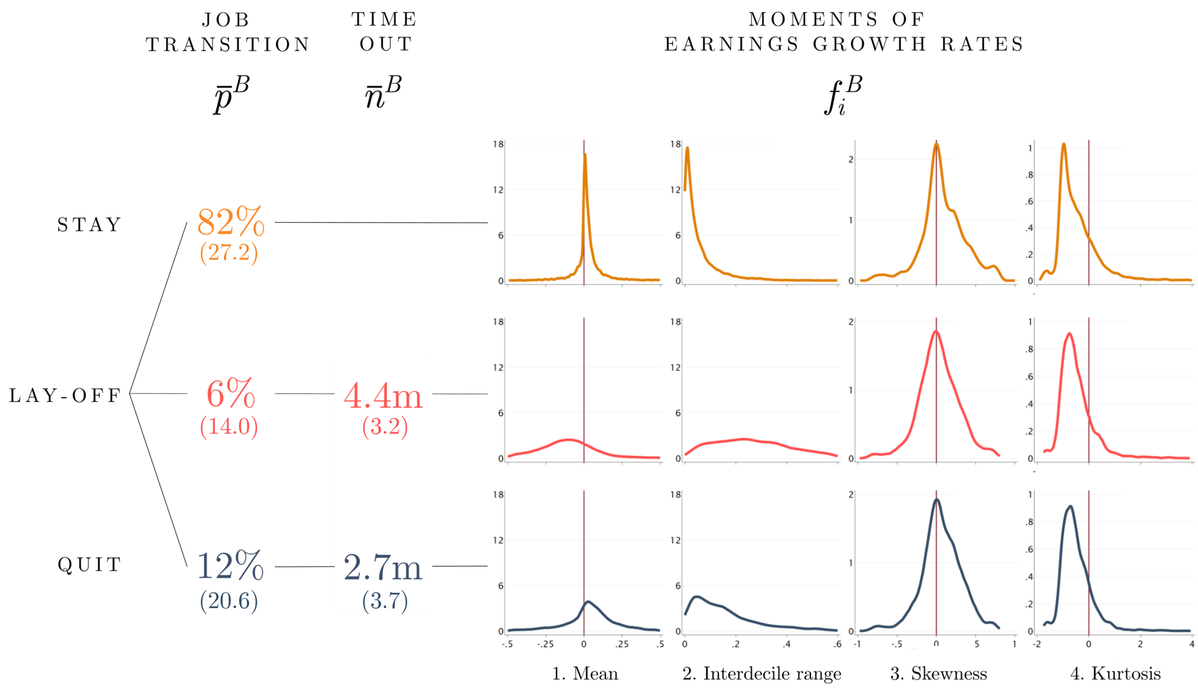
$$n^L_i = 2(n^L_{i,3} - n^L_{i,1}) + 7.5(n^L_{i,12} - n^L_{i,3}) + 12(100 - n^L_{i,12})$$

We use the same procedure for the re-employment period after a quit, n^Q_i.

The numbers under "Time out" in Figure 3 are the average periods out of work following a job separation, $\bar{n}^B$. We find that respondents expect to spend 4.6 (2.7) months on average to find a new job after being laid off (quitting). These results imply that the respondents anticipate spending more time out of work following a layoff than following a quit, as might have been expected. In Online Appendix C.1, we also plot the distributions of time out labor force.

The fact that many survey respondents expect to spend a short time out of work following

[11] To construct these weights, we estimate a probability model of survey participation using the 2020 administrative data with information about the demographics of the Danish population who are active in the labor market and use the inverse of the predicted propensity scores as weights. For a detailed description of the construction of the population weights see Online Appendix B.



Note: The figure shows answers to the questions in the conditional survey instrument, where the rows correspond to the branches "Stay", "Layoff", and "Quit". The first column shows the average probabilities of each branch, $\bar{p}^B$. The second column shows the average of the expected reemployment period in each branch, $\bar{n}^B$, in months. The distributions show the cross-sectional distribution of the 1st to 4th moment of the subjective conditional earnings distributions, f_i^B. The parenthesis below each value shows the corresponding standard deviation. We measure the second moment by the interdecile range, $p_{90} - p_{10}$. We measure skewness using Kelley's measure of skewness: $S_K = \frac{(p_{90}-p_{50})-(p_{50}-p_{10})}{(p_{90}-p_{10})}$. We use the Crow-Siddiqui measure of excess kurtosis $K_{CS} = \frac{(p_{97.5}-p_{2.5})}{(p_{75}-p_{25})} - 2.91$. Survey results are weighted using population weights. In Online Appendix C.1, we plot the distributions of job transition and time out respectively.

Figure 3: Overview of branch-by-branch survey responses

a quit contrasts with the standard registry-based assumption that quits correspond to direct job-to-job transfers. The anticipated time out of the labor force after quitting may reflect either an anticipated break or an anticipated job search. In unreported analyses we regress expected time out of work following a quit on liquid assets relative to disposable income, an often used indicator of being liquidity constrained (Leth-Petersen, 2010). We find that workers with less liquid wealth expect to spend less time out of work after quitting, as if pressured back to work more quickly.

2.4 Conditional Earnings Risk

A key innovation in our survey is that we obtain subjective distributions of expected earnings growth **conditional** on job transitions. We are therefore able to calculate the moments of their subjective earnings distributions for each respondent i in each branch B. We simulate the empirical distributions of conditional earnings growth rates for each survey respondent in each branch, $\hat{f}_i^B$, by taking 20,000 random draws from the mixture of uniform distributions of expected earnings, which is illustrated in Figure 2, panel b. We convert expected earnings levels to logs and subtract the log of earnings in 2020 (self-reported) to obtain a distribution of one-year-ahead log earnings growth. This procedure imposes minimal assumptions on the shape of the empirical subjective conditional distributions, $\hat{f}_i^B$.[12]

The last four columns of Figure 3 show the cross sectional distribution of the first four moments of subjective earnings growth distributions.[13] Each row corresponds to a different branch, B. Turning first to the means, the average respondent expects a 3% increase in earnings if staying with their current employer. Following a layoff, individuals on average expect an 11% decrease in annual earnings when they find a new job, and a 7% increase when they find a new job after a quit. These two branches also exhibit considerable het-

[12] We also fitted beta distributions to the subjectively reported data from the balls-in-bins answers. This did not change the results in any important way (not reported).

[13] Following the practice in the literature we measure the second moment, skewness, and kurtosis for each of the subjective distributions using robust, quantile based measures. We measure the second moment by the interdecile range, $p_{90} - p_{10}$. We measure skewness using Kelley's measure of skewness: $S_K = \frac{(p_{90}-p_{50})-(p_{50}-p_{10})}{(p_{90}-p_{10})}$. S_K measures the relative length of the right and left tails of the distribution. If $S_K > 0 (< 0)$, then it means the right (left) tail is longer and large positive (negative) draws are more likely than large negative (positive) draws. Therefore, this captures the extent to which individuals perceive larger upside or downside risk. Finally, we use the Crow-Siddiqui measure of excess kurtosis $K_{CS} = \frac{(p_{97.5}-p_{2.5})}{(p_{75}-p_{25})} - 2.91$. This measure compares the range of the middle 95% of the distribution to that of the middle 50%. The statistic is normalized by 2.91, such that the Crow-Siddiqui measure of excess kurtosis for a normal distribution is zero. Excess kurtosis is informative about the extent to which expected earnings growth is concentrated in the center of the distribution or in the tails. Large excess kurtosis means larger risk of extreme changes.

erogeneity in the means relative to the stay branch. Among those who report a positive probability of being laid off, 73% of the respondents expect a decrease in earnings if this state is realized. In contrast, among those who report a positive probability of quitting, 81% of the respondents expect to increase earnings if that state materializes.

The next column shows the interdecile range, $p_{90} - p_{10}$. This measures how much risk people associate with their earnings prospects in each branch. As might be expected, the results show that people tend to associate least risk to their earnings growth in the stay branch and most risk in the layoff branch. There is also less heterogeneity in responses in the stay branch where a considerable amount of the mass is bunched toward zero. In contrast, in the layoff branch different respondents report very different perceptions of earnings risk.

The distribution of skewness is similar across all branches. In all cases, it is clustered around zero and symmetric. This means that the modal respondent is creating symmetric distributions with their bins and balls on all the branches. However, it is noticeable, that there is a lot of heterogeneity and many individuals report distributions that are skewed.

Lastly, we turn to the distribution of excess kurtosis. The final column of Figure 3 shows that also these distributions appear similar across branches, but with a lot of heterogeneity across respondents. On average, excess kurtosis is negative, which means that the average subjective distribution is not as peaked as a normal distribution. This means that most respondents have entered distributions with relatively more mass between the center and the tails than a normal distribution.

Being laid off leads to the worst outcomes, on average, and respondents associate most risk to what may happen here. Staying with the current employer is expected to lead to small increases in earnings and respondents are most certain about the outcome in this state. Quitting leads to the best outcomes, and the level of risk is between that of staying and being laid off. Overall, the data uncover massive heterogeneity in expectations of future earnings, and this is reflected in all four moments and across all labor market transitions.[14]

3 Comparing Survey and Administrative Data

The Danish research data infrastructure allows us to directly compare measures elicited in the survey with their realized counterpart in the administrative data and thereby to assess

[14] In Online Appendix C.2, we reproduce Figure 3 using standard measures of the moments.

the credibility of the survey instrument. To that end, the survey data is combined at the individual level with administrative data made available by Statistics Denmark from different sources with third-party reported records from various sources. The Danish administrative data are known to be of high quality (Kleven et al., 2011) and have been used extensively in previous studies, see for example, Browning et al. (2013), Leth-Petersen (2010) and Chetty et al. (2014). The data are made available with a time lag, with data through 2021 currently available for research thus allowing us to directly compare expectations with their realized counterparts. Data gathered in this manner includes earnings from work and job transitions as well a host of other administrative data providing background information about each respondent. For our comparison between survey and administrative earnings, we use monthly data about employer matches and earnings to identify job transitions, time spent out of work, and annual earnings. We also use a standard battery of administrative data compiled by Statistics Denmark.

We start out by comparing job transitions and time out of work following a job separation. We then introduce a method of aggregating our conditional survey responses to arrive at a holistic measure of subjective earnings risk that we then pool to compare with the corresponding numbers in the register data. Our baseline comparison will be based on administrative data for 2021. To allay one possible concern, note that the COVID-19 pandemic hit the Danish economy lightly and respondents seemed to recognize that this would be the case. Massive furloughing schemes were set in place very quickly by the Danish government. As a result, the lowest employment level during 2020 was only 40,000 below the baseline pre-pandemic level of 2,768,766 (February 2020), and by the end of the year this small loss had been recovered. During 2021 employment accelerated and by end of 2021 total employment was 2,916,139, about 5 percent above the pre-covid level. To verify whether this is consistent with the data that that we collect, we will also verify that subjective expected earnings growth aligns with the realized earnings growth during the period that expectations concern

3.1 Job Transitions

In the survey, we ask about the probability of staying with the current employer, the probability of being laid off and the probability of quitting. In the administrative data we only observe job-separations but not the reason for the job separation. Consequently, as a first-order check on the survey answers to these items, we compare the probability of staying with the same employer, since this object can be computed directly from the administrative data.

From the survey we consider the average reported probability of staying with the same employer, $\bar{p}^S$. In the administrative data we observe employer-employee matches at the monthly frequency and obtain a direct counterpart to $\bar{p}^S$ by calculating the share of employees who stay with the same employer throughout the calendar year. Figure 4, panel (a) shows the average stated probability (solid line) of staying with the same employer throughout 2021 and the fraction of stable job-matches (dashed line) throughout 2021 in the administrative data, both summarized by age. Generally, the likelihood of remaining in the same job throughout the year is lower among the young and there is only a low likelihood that workers aged 40+ separate from their job. The alignment between survey and registry is striking.

3.2 Time Out of the Labor Force

In the survey we ask about time spent out of work following a potential quit and lay-off. As before, we cannot confront these objects directly with administrative data as we do not know the reason for an observed separation. We therefore compare time spent out of work following any type of separation between the survey and the administrative data. In the administrative data we observe employer-employee matches at the monthly frequency and are able to track the number of months spent out of work following a separation. As we currently only have administrative data until 2021, we consider time spent out of work following job separations that took place in 2020 such that we can follow periods out of work that extend into 2021. From the survey, the expected time out of work in the survey is calculated as a combination of quits and layoffs:

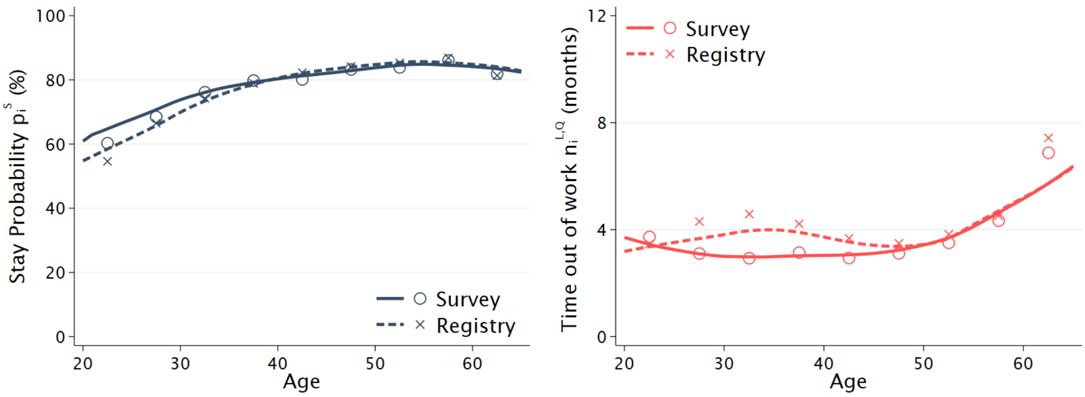
$$n_i^{L,Q} = \frac{p_i^Q n_i^Q + p_i^L n_i^L}{p_i^Q + p_i^L} \tag{1}$$

The result of the comparison is shown in Figure 4, panel (b). According to the survey (solid line), the average expected time spent out of work following a job separation is about 3.5 months for people aged less than 50 and the expected time out of work increases dramatically for workers aged 50+. The pattern is similar when the corresponding measure from the administrative data is plotted (dashed line).

3.3 Overall Subjective Earnings Expectations

We now turn to compare the overall subjective earnings expectations to the realized earnings growth for 2021 that we observe in the administrative data.

To do this, we first aggregate the conditional answers for each respondent into one dis-



(a) Stay probability (b) Time out of work

Note: The figure shows how two components from the survey (solid line) and the administrative data (dashed line) compare over the life-cycle. Panel (a): In the survey, we directly ask about the probability of staying with the same employer for the next 12 months. In the registry, we compute the proportion of workers who stayed from Dec 2020 to Dec 2021 with the same employer. Panel (b): We calculate the expected time out of work after a separation in the survey using Equation (1). In the administrative data we consider job separations that took place during 2020 and follow time spent until reemployment occurs, possibly extending into 2021. In the registry, 1) we exclude the workers going back to the same employer (possibly seasonal work) and 2) we consider the first separation and do not take into account further separations within a year. "o" and "x" represent the empirical mean across 5-years age bins for the survey and the administrative data, respectively. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights. Online Appendix C.3 shows the corresponding figures for 2019.

Figure 4: Job separations and time out of work in the survey and in the administrative data

tribution that characterizes the respondent's overall subjective expectations concerning one-year ahead earnings growth and we denote this subjective expected *holistic* earnings growth. This object summarizes the overall subjective probability distribution over one-year ahead earnings growth taking into account all the different contingencies that we asked about.

To construct the subjective holistic expected earnings growth distribution we weight together each of the branches, $B = \{S, Q, L\}$, for individual i,

$$g_i = p_i^S f_i^S + p_i^Q f_i^Q (1 - n_i^Q) + p_i^L f_i^L (1 - n_i^L), \tag{2}$$

where p_i^S, p_i^Q, and p_i^L are the probabilities of staying, quitting, and being laid off, n_i^Q and n_i^L are time out of work following a quit and a layoff. f_i^S, f_i^Q, and f_i^L are the subjective probability distributions over one-year ahead earnings growth rates for each of the three branches, staying, quitting and being laid off. The subjective holistic probability distribution over one-year ahead earnings growth, which we denote g_i, captures the total earnings growth risk, as perceived by individual i.

In practice, we simulate the empirical distribution of $\hat{g}_i$ by making a large number of random draws for each respondent based on the stated transition probabilities, p_i^B, and the individual empirical distributions of $\hat{f}_i^B$ and $\hat{n}_i^B$,[15] which are then weighted together according to Equation (2). The goal is to construct a measure of overall one-year ahead earnings expectations. Some of the contingent questions pertain to a period that goes beyond the 12 months following the interview. For example, the probability distributions over earnings growth in the case of quit and lay-off, f_i^Q and f_i^L concern the 12-months period following separation from the previous job and potential time out of work. When weighting together all the components in equation (2) we take this into account such that the overall probability distribution over one-year ahead earnings, g_i, exactly matches the 12 months following the interview. In this way, our measure of g_i pertains to the exact same period that the administrative data measure, i.e., the calendar year.

We simulate the empirical distribution $\hat{g}_i$ by drawing 20,000 job transition events for each individual based on the stated job transition probabilities. From each of these simulated job transitions, we simulate time out of work and the conditional earnings distribution for the relevant branch based on the empirical distributions of $\hat{n}_i^B$ and $\hat{f}_i^B$. In this way, we

[15] In addition to the individual point estimates, n_i^L and n_i^Q, we also construct individual empirical distributions of time out of work following a separation $\hat{n}_i^L$ and $\hat{n}_i^Q$. These distributions are simulated out of stated probabilities of being reemployed within a certain time frame. We refer to Online Appendix D for specifics.

simulate 20,000 synthetic realizations for each respondent based on the reported survey answers. We give a complete account of the simulation protocol in Online Appendix D.[16] In the following we refer to g_i but in practice we use $\hat{g}_i$.

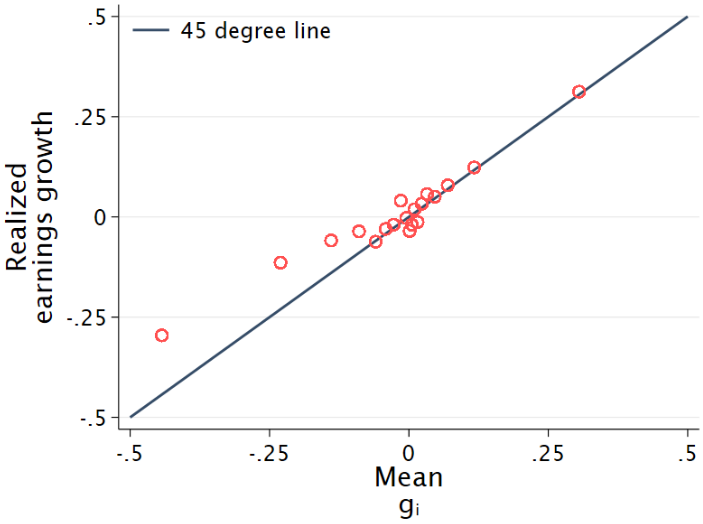
g_i is the subjective probability distribution over earnings growth during 2021. The mean of $\mathbb{E}_i[g_i]$ has a direct counterpart in the administrative data, namely the realized earnings growth, Δy_i, for 2021. In Figure 5 we show a direct comparison of $\mathbb{E}_i[g_i]$ and Δy_i. The figure displays a binned scatter plot of the $\mathbb{E}_i[g_i]$ against realized earnings in 2021 as observed in the administrative data. The figure is constructed by ranking $\mathbb{E}_i[g_i]$ for all respondents and dividing them into 20 equally sized groups. Within each of the 20 groups the average realized earnings growth in 2021 is calculated. The figure shows that there is a remarkably close alignment between subjective earnings growth expectations and the subsequent realizations for the vast majority of the sample. In other words, respondents are well-informed about their own future earnings growth. Finally, note that the survey was fielded in January 2021, before the COVID-19 lockdown. The close alignment of expectations with realizations confirms that the COVID-19 pandemic did not significantly misalign people's expectations with the actual outcomes.

3.4 The Cross-Sectional Distribution of Earnings Growth

Finally, we turn to compare the cross sectional distribution of expected earnings growth in the survey to the cross sectional distribution of realized earnings growth for 2021 in the administrative data. The first characterizes the *expected* earnings growth variability in the population and the second characterizes the corresponding *realized* earnings growth variability in the population.

In order to arrive at a cross sectional distribution of expected earnings growth, we pool the distributions of subjective expected holistic earnings growth, g_i, for all N individuals in our sample to arrive at a cross sectional distribution that describes the expected earnings growth variability in the population that is conceptually comparable to the cross sectional distribution of realized earnings growth that we observe in the administrative data. In doing this, we pool within cells in which individuals share the same observable

[16] Equation (2) implicitly assumes that job separations take place at the beginning of the period. In practice, expected earnings in case of job separation may be a convex combination of earnings in the old job and earnings in the new job following time out of work. We have a more detailed discussion in Online Appendix D.



Note: The figure shows a binned scatter plot of the mean of subjective expected earnings growth, $\mathbb{E}_i[g_i]$ and the corresponding realized earnings growth, Δy_i. We first divide the sample into 20 equal-sized bins and plot the empirical mean of the X and Y-axis. The regression coefficient of mean g_i on realized earnings growth is 1.03. For the plot, we focus on the range of mean g_i in between -0.5 and 0.5.

Figure 5: Comparing mean g_i and realized earnings growth

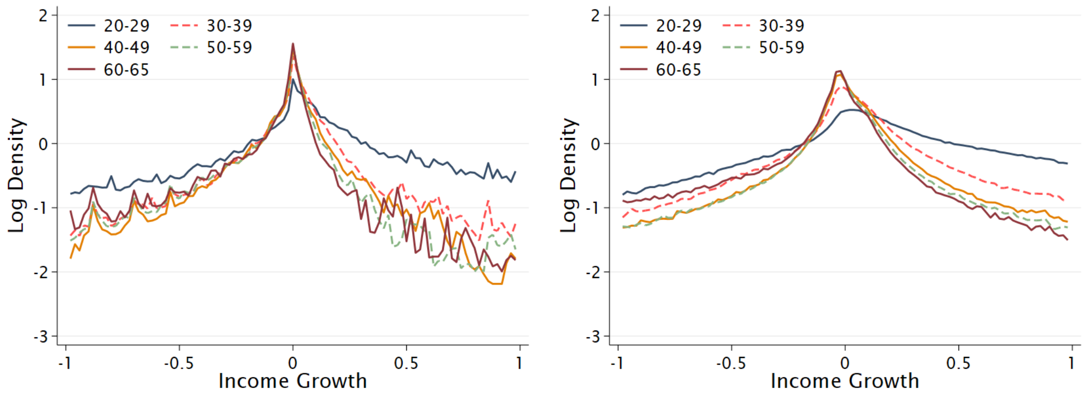
characteristics X:

$$h_S^X = \frac{1}{N^X} \sum_{i=1}^{N^X} g_i^X \tag{3}$$

The pooled distribution, h_S^X, reflects the total variability of expected earnings growth in the population and it is thus directly comparable to the distribution of realizations earnings growth observed in the administrative data, which we denote h_A^X, for individuals who have similar observable characteristics, X.

We start out by plotting the distribution of pooled holistic expected earnings growth from the survey, h_S^X, cf. equation (3), within broad age groups and compare it to the corresponding distributions of realized earnings growth from the administrative data, h_A^X.[17]

[17] As a further check we compare the stated level of earnings in 2020 from the survey to the level of earnings recorded in the administrative data. In the survey, which is conducted in January 2021, we ask about total earnings throughout 2020. Information about total earnings for 2020 is also reported directly from employers to the tax agency and is made available in the administrative data. In Online Appendix C.4 we compare these two different measures of earnings in 2020, and it turns out that survey answers line up accurately with the administrative data. This is an indication that respondents are well-informed about their level of earnings.



(a) Survey, h_S^X (b) Registry, h_A^X

Note: Panel (a) plots log density for the pooled distribution of expected holistic earnings growth rates from the survey, h_S^X, where X indicates partitions by age groups. (b) plots the distribution of annual earnings growth from 2020 to 2021 as observed in the administrative data for the full population, h_A^X. For constructing the distribution of earnings growth in the administrative data we dropped observations where the level of annual earnings is less than 24,000 DKK in 2020. Survey results are weighted using population weights. Online Appendix C.5 shows the corresponding figures using administrative data for 2019.

Figure 6: Pooled earnings risk and registry earnings risk

Figure 6 panels (a) and (b) show the distributions of earnings growth in the survey and the registry. The two distributions are similar and have similar life cycle patterns. Generally, the distributions based on the survey data and the administrative data both have thicker and longer tails than a normal distribution. This is analogous to the patterns documented by Guvenen et al. (2021) for the US.[18] In the survey as in the administrative data, younger workers (age 20-29) tend to have a higher density of positive earnings growth and the log density is right tilted, arguably reflecting career progress for individuals in this age group. At older ages, the density of positive earnings growth decreases and the density of negative earnings growth increases. It is also notable that the peak around 0 earnings growth increases. The log density level for each age group is well aligned between the survey and the registry. Panel (b) is based on the full population. In Online Appendix C.6, we present the distribution of realized earnings growth observed in administrative data but only including the individuals from the survey and note that this too looks similar to the distributions shown in Figure 6.

In Figure 7 we examine how the moments of the distributions h_S^X and h_A^X evolve over

[18] Leth-Petersen and Sæverud (2022) document that the distribution of realized earnings growth in Denmark share many of the features that are also observed in the US data.

the life cycle. The graph shows that life cycle patterns are broadly similar between the survey data and the administrative data for all moments. Mean earnings growth, panel (a), decreases with age. Young workers, on average, expect and realize positive earnings growth while the oldest workers expect and realize negative earnings growth. Next, the interdecile range of the distribution of earnings growth, panel (b), is especially high for people in the 20s but is relatively stable after age 30. This means that young workers tend to associate their earnings growth with relatively more risk. Skewness, panel (c), is decreasing with age. One divergence between the survey and the registry is that for those of ages 30-50, skewness is negative while in the registry it is close to zero in the survey. Lastly, note that excess kurtosis, panel (b) increases in age. This means that the earnings growth distribution becomes more peaked and develops fatter tails as age increases.[19]

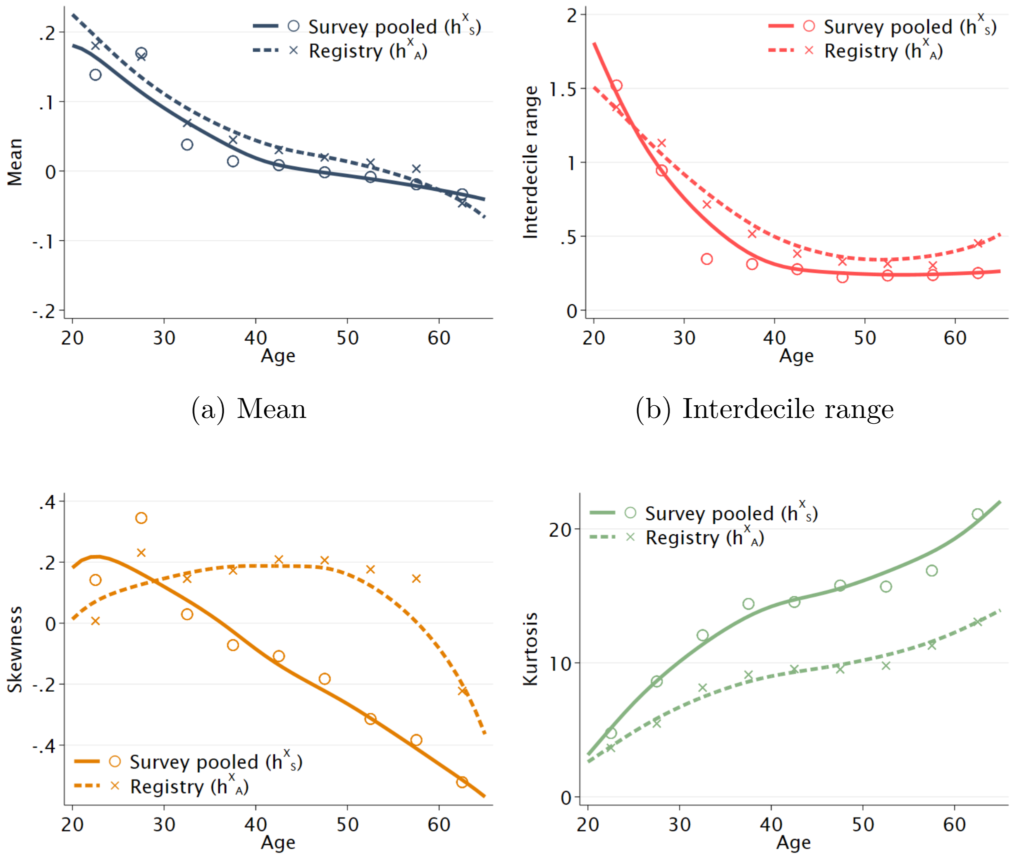
Overall, we find high coherence between the pooled distribution of expected earnings growth based on the survey data and the distribution of actual earnings growth recorded in the administrative data.[20]

4 Administrative versus Survey-Based Earnings Risk

In this section we compare the distribution of realized earnings growth, often used to infer earnings risk, with subjective earnings risk directly measured in survey data. Guvenen et al. (2021), for example, group the population into three broad age groups and percentiles of earnings levels and examines the characteristics of the distribution of earnings growth within these cells. Obviously this method of inferring earnings risk from the moments of the cross sectional distribution of realized earnings growth comes with assumptions about worker homogeneity, i.e. that groups of workers draw earnings realizations from the same underlying distribution which can be characterized by the cross sectional distribution of realized earnings. To explore the validity of such homogeneity conditions, we analyze whether moments calculated from the distribution of earnings growth in the administrative data within these detailed partitions, h_A^X, are able to mimic the moments of the subjective distributions of holistic earnings growth within the same cells, g_i^X.

[19] In Online Appendix C.7, we document that also the standard moment measures (standard deviation, skewness, and kurtosis) show similar patterns in the registry and the survey.

[20] The comparison is based on administrative data for 2021. In Online Appendix C.5 we repeat the exercise using administrative data for 2019, and find that the pattern is practically identical. This confirms that the survey year is not particular.



(a) Mean (b) Interdecile range

(c) Skewness (d) Kurtosis

Note: The figure shows the average value of the 1st to 4th quantile based moments (See notes to Figure 3) over the life cycle of the pooled earnings distribution in the survey, h_S^X, and in the administrative data, h_A^X. "o" and "x" represent the empirical mean across 5-years age bins for the survey and the administrative data for 2021, respectively. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights. Online Appendix C.7 shows the corresponding figures using standard moments.

Figure 7: Higher-order moments of h_S^X and h_A^X over the life cycle

4.1 Coarse Stratification

We start out by illustrating the main insight based on a coarse partition of the administrative data which allows us to summarize the main insight graphically. We then implement a more detailed partition that is close to the most granular researchers could achieve with administrative data.

In the coarse stratification we divide the administrative data for the Danish population into six cells based on three age groups (20-34, 35-49, 50-65) and the earnings level being *High* or *Low* (above/below the median). Within each of these cells we calculate the moments of the cross sectional distribution of realized earnings growth from the administrative data, h_A^X. Each moment within a cell will be a unique number. Next, we calculate the corresponding moments for the pooled distribution of subjective holistic earnings growth expectations, h_S^X, cf. equation (3). Also in this case, each moment within a cell will be a number. If these two objects are similar then the survey data and the registry data are consistent with each other. We then compare with the cross sectional *distribution* of moments of the subjective distributions of holistic earnings growth, g_i^X, cf. equation (2), for individuals from the survey belonging to the cell.

Figure 8 presents estimates of mean earnings growth and interdecile range (IDR) for two of the groups in the coarse stratification described above. The first row is for the cell 20-34:*Low* and the second row is for the cell 50-65:*High*. The first column shows the estimates for the mean and the second column shows estimates of the interdecile range. Panel (a) shows that for the 20-34:*Low* group, the mean calculated from the administrative data, $\mathbb{E}[h_A^X]$, and the mean calculated from the pooled subjective distribution, $\mathbb{E}[h_S^X]$, are practically identical while there is considerable heterogeneity in the means of the subjective distributions, $\mathbb{E}_i[g_i^X]$. Panel (b) shows that the interdecile range estimated from the administrative data, $\text{IDR}[h_A^X]$, and the pooled survey data, $\text{IDR}[h_S^X]$, are also very close to each other, but that the subjective interdecile ranges, $\text{IDR}_i[g_i^X]$, are very heterogeneous and centered at much lower values than the interdecile range calculated from the administrative data and the pooled subjective data.

The second row of Figure 8 shows the corresponding figures for the 50-65:*High* group. The estimate of the mean, panel (c), and the interdecile range, panel (d), based on the administrative data and the pooled subjective data are also very similar. The modal point of the distribution of subjective interdecile ranges is also positioned lower than the estimate of the interdecile range based on the administrative data. However, this group displays less heterogeneity in subjective means and the distance between the estimate of the interdecile range based on the administrative data and the modal point of the distribution of subjective interdecile ranges is smaller than for the 20-34:*Low* group.

The main insight from Figure 8 is that risk inferred from the administrative data tends to be larger than risk inferred from the subjective data and that the degree of overshooting tends to be linked to how much dispersion there is in the distribution of subjective means. This is consistent with the view that the pooled distribution of expected earnings growth, h, is a mixture of underlying subjective distributions, g_i, cf. Equation (3). The theoretical variance of a mixture distribution of N equally weighted subjective distributions with individual means and variances μ_i, σ_i^2 is:

$$\mathrm{Var}(h) = \frac{1}{N}\sum_i^N \sigma_i^2 + \frac{1}{N}\sum_i^N \mu_i^2 - \left(\frac{1}{N}\sum_i^N \mu_i\right)^2 \tag{4}$$

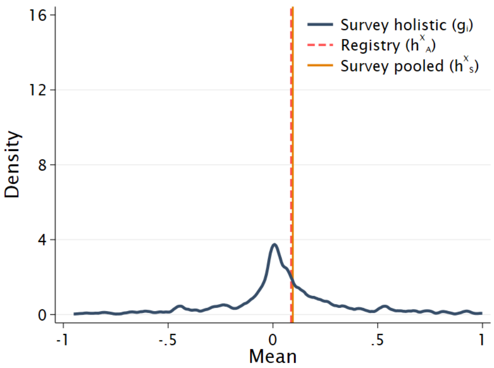
The variance of the mixture distribution is the mixture of the variances of the subjective distributions plus a non-negative term reflecting the differences in means between the subjective distributions. By Jensen's Inequality the average squared mean is weakly greater than the squared average mean, implying that the sum of the last two terms is non-negative and hence that the variance of the mixture distribution is weakly larger than the average variance of the subjective distributions, $\mathrm{Var}(h) \geq \frac{1}{N}\sum_i^N \sigma_i^2$. Put differently, over-dispersion in the pooled holistic distribution, h_S^X, and by extension the distribution from which the registry based variance is calculated from, h_A^X, occurs when the underlying subjective holistic distributions, g_i, have heterogeneous means, and, as a result of this, risk and heterogeneity are confounded.[21]

4.2 Refining the Stratification

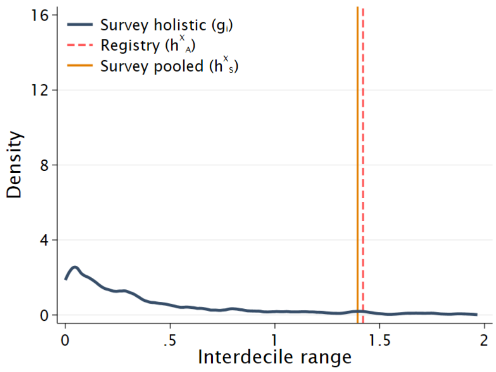
The logic above suggests that the gap between subjective and administratively estimated risk will be lower the more we refine the stratification of the population. To pursue this we now consider a finer stratification. Specifically we make use of the administrative data and partition the distribution of realized earnings growth in the population data into 300 cells by three age groups and earnings percentiles following Guvenen et al. (2021). For each of these cells we perform the same calculations as in the illustration above: We calculate the interdecile range of the distribution of realized earnings growth within each cell, $\mathrm{IDR}[h_A^X]$, and the average of the subjective interdecile ranges within each of these cells, $\frac{1}{N^X}\sum_i^{N^X} \mathrm{IDR}_i[g_i^X]$. The result is shown in Figure 9.[22]

[21] How skewness and kurtosis of the pooled distribution are related to skewness and kurtosis of the underlying subjective distributions is ambiguous. We refer to Online Appendix E.2 for derivations.

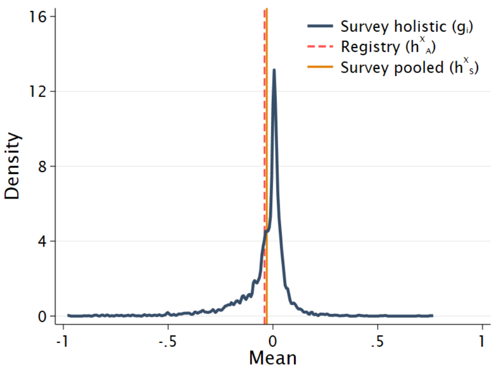
[22] in Online Appendix E.3, we report results for skewness and kurtosis.



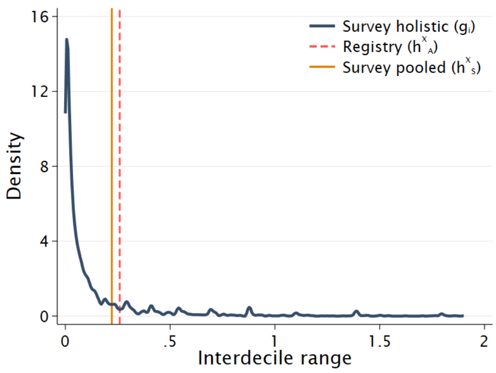
(a) 20-34:*Low*, Mean



(b) 20-34:*Low*, Interdecile range



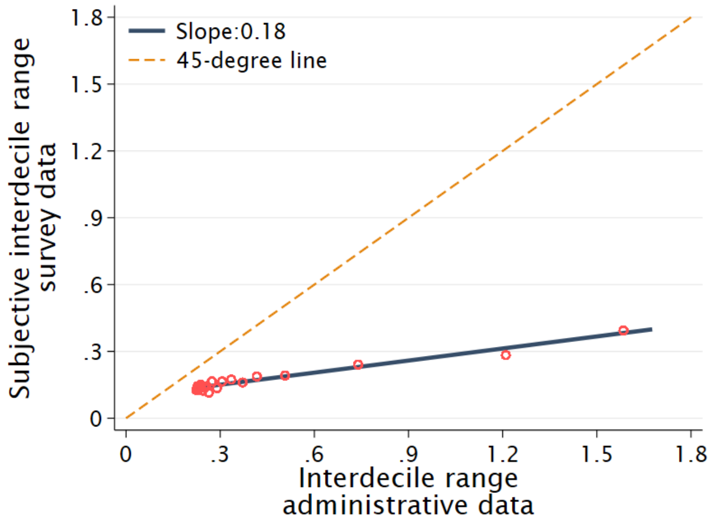
(c) 50-65:*High*, Mean



(d) 50-65:*High*, Interdecile range

Note: The figure shows estimates of the mean and interdecile range for h_S^X, h_A^X, and the distribution of g_i for two subgroups in the data. The top row shows these statistics for individuals aged 20-34 and with below median earnings (20-34:*Low*), and the bottom row shows the corresponding statistics for individuals aged 50-65 and with above median earnings (50-65:*High*). Online Appendix E.1 show the corresponding figures for the remaining subgroups.

Figure 8: Mean and interdecile range of h_S^X and h_A^X, and the distributions of individual means and interdecile range of g_i for two selected subgroups



Note: The figure compares average interdecile ranges of subjective holistic earnings expectations, $\frac{1}{N^X}\sum_i^{N^X} \text{IDR}_i[g_i^X]$, to interdecile ranges calculated from administrative data, $\text{IDR}[h_A^X]$, within 300 cells divided by age groups (20-34, 35-49, 50-65) and earnings percentiles. The panel shows a binned scatterplot (red circles) of $\frac{1}{N^X}\sum_i^{N^X} \text{IDR}_i[g_i^X]$ by vigintiles of $\text{IDR}[h_A^X]$. A regression line based on the 300 data points is overlaid.

Figure 9: Comparing interdecile ranges calculated from subjective expectations and from administrative data

We find that the average of the subjective interdecile ranges, $\frac{1}{N^X}\sum_i^{N^X}\text{IDR}_i[g_i^X]$, within each cell is much smaller than the interdecile range calculated from the administrative data within the same cell, $\text{IDR}[h_A^X]$. We find that within each cell there is a lot of heterogeneity in the subjective mean growth rates (not reported). Consistent with the idea that the pooled distribution of earnings growth rates is a mixture of individual distributions of expected earnings growth rates, this finding suggests that heterogeneity is assigned to chance when earnings risk is inferred from the distribution of realized earnings growth and, as a consequence, that risk is systematically overstated compared to how the majority of individuals experience it.

Assigning heterogeneity to risk could potentially be the result of not applying a sufficiently fine partition by observable characteristics. In Online Appendix E.4 we present results for an even finer grid with 1,800 cells for age, earnings deciles, gender, and university education and find results that are practically identical. Furthermore, we also try a version where we include the individual growth rate of earnings in the covariate set. A branch of the literature assumes that individual earnings grow deterministically at an unobserved rate. This is known as the heterogeneous income profiles model (HIP, e.g., Guvenen, 2009; Browning et al., 2013). In order to account for this possible type of heterogeneity we construct an alternative version of Figure 9 where we expand the covariate set to include also the average growth rate of earnings within the past five years. This essentially allows for an individual fixed effect in the growth rate of earnings. The resulting figure is practically identical to Figure 9. These results are reported in Online Appendix E.5.

4.3 The Role of Observable Characteristics

The role of observable characteristics is to control for the otherwise unobservable heterogeneity in the means of the underlying subjective distributions, $\mathbb{E}_i[g_i]$, such that subjective heterogeneity and risk are not confounded when assessing risk from the distribution of realized earnings growth, cf. equation (4). In this section we examine how well observable characteristics predict heterogeneity, i.e., $\mathbb{E}_i[g_i]$, and how using observable characteristics as a way to control for heterogeneity affects the estimate of the interdecile range of the distribution of realized earnings growth, $\text{IDR}[h_A^X]$.

To do this we regress subjective means, $\mathbb{E}_i[g_i]$, on observable characteristics, X_i. We do this for a range of different covariate sets and for each of the covariate sets, we calculate the average of interdecile ranges of the cross sectional distribution of realized earnings growth from the administrative data within X cells, $\text{IDR}[h_A^X]$. To assess how well

Table 1: Regressions of $\mathbb{E}_i[g_i]$ on observable characteristics

	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable : $\mathbb{E}_i[g_i]$						
Age group indicators	✓	✓	✓	✓	✓	✓
Earnings percentile indicators		✓	✓	✓	✓	✓
Female			✓	✓	✓	✓
Education				✓	✓	✓
Past earnings growth quintile indicators					✓	✓
Unemployment						✓
Industry						✓
Observation	10,746	10,746	10,746	10,746	10,746	10,746
R-squared	0.015	0.028	0.028	0.028	0.028	0.029
Average $\text{IDR}_i[g_i]$	0.179					
$\text{IDR}[h_A]$	0.630					
Average $\text{IDR}[h_A^X]$	0.621	0.602	0.599	0.593	0.591	0.573

Note: The table shows regressions of $\mathbb{E}_i[g_i]$ on observable characterstics. Age group indicators are dummy variables for 5 years age bin. Earnings percentiles include dummies for earning level percentiles. The percentiles are calculated within age groups. Education is a dummy for having completed a college degree. For past earnings growth, we calculate average earnings growth from 2016-2020 and construct quintiles within age group. Unemployment is a dummy variable, taking the value 1 if workers were without an employer for at least one month during 2020. Additionally, the industry classification captures private and public sector employment. $\text{IDR}[h_A]$ is the interdecile range of h_A, i.e. the pooled cross section of earnings growth. Average $\text{IDR}[h_A^X]$ is the average of interdecile ranges calculated from the cross sectional distributions of realized earnings growth in the administrative data within X cells. A weighted average is used to compute the Average $\text{IDR}[h_A^X]$, taking into account the different sample sizes across each cell of X.

observable characteristics control for heterogeneity, we compare with the interdecile range of the pooled cross sectional distribution of realized earnings, i.e., constructed without controlling for observable characteristics, $\text{IDR}[h_A]$, and with the average of the subjective interdecile ranges, $\text{IDR}_i[g_i]$.

The result is shown in Table 1. In column (1), $\mathbb{E}_i[g_i]$ is regressed on a full set of age group dummies. The R^2 is 0.015 meaning that age group dummies are able to explain only 1.5% of the variation in $\mathbb{E}_i[g_i]$. As a result, $\text{IDR}[h_A^X] = 0.621$ is only slightly lower than $\text{IDR}[h_A] = 0.630$ and much bigger than the average subjective interdecile range, $\text{IDR}_i[g_i^X] = 0.179$. In column (2) we add percentile dummies for the level of earnings. This increases R^2 to 0.028 and reduces $\text{IDR}[h_A^X] = 0.602$. In column (3)-(6) we sequentially add a dummy for being female, a dummy variable for having a college degree, quintile dummies for the average earnings growth during 2016-2020, i.e., the past five years, and in the last column, we incorporate a dummy variable for unemployment, which indicates if workers experienced at least one month of unemployment during 2020, along with industry classification differentiating between private and public sectors. In all cases are the covariates significant in explaining the variation in $\mathbb{E}_i[g_i]$, but collectively the covariates explain only a small fraction of the variation in $\mathbb{E}_i[g_i]$. As a result, for the richest specification in column (6), $\text{IDR}[h_A^X] = 0.573$ which is still only slightly lower than $[h_A] = 0.630$ and much bigger than the average subjective interdecile range, $\text{IDR}_i[g_i^X] = 0.179$. In sum, observable characteristics have a significant but limited ability to capture the heterogeneity in the means of the underlying subjective distributions, and this explains our key finding that risk inferred from the cross-sectional distribution of earnings growth systematically overstates the level of subjective risk that the majority of individuals experience, even when taking into account heterogeneity along observable characteristics. As individuals have approximately mean correct expectations about their income growth, cf. Figure 5, this indicates that individuals have more information than the econometrician who only observes administrative data on realized earnings growth across the population and therefore cannot separate risk from heterogeneity.

5 Job Transitions and Subjective Earnings Risk

In this section, we take advantage of our conditional survey instrument to decompose subjective holistic earnings risk, g_i, according to job transitions and show that such transitions are key in explaining the level and heterogeneity of higher-order moments. To illustrate this we compute not only the average life cycle patterns of the four moments of the subjective holistic earnings growth distributions, g_i, but also the subjective risk arising from staying in the current job, f_i^S. Figure 10 illustrates the results, which confirm the great importance of job transitions for earnings risk.

Panel (a) shows average mean earnings growth across the life cycle. Generally, mean earnings growth decreases as the life cycle progresses and this is the case for both holistic earnings growth and for earnings growth conditional on staying. Holistic earnings growth is, on average, positive up to about age 50 and then turns negative. Fixing earnings risk to the stay branch increases expected earnings growth for all ages, and this happens to a degree where also the oldest workers expect positive earnings growth, i.e., the net contribution of job transitions is to reduce expected earnings growth.

Panel (b) shows how the average subjective risk, which we measure as the average interdecile range, pertaining to earnings growth over the life cycle. Considering risk based on subjective holistic earnings growth expectations we find that risk is highest for young people. Fixing earnings growth risk to come only from the stay branch generates a big drop in risk at all ages, but most dramatically for the young. This shows that risk pertaining to one-year ahead earnings growth is intimately tied to job transitions.

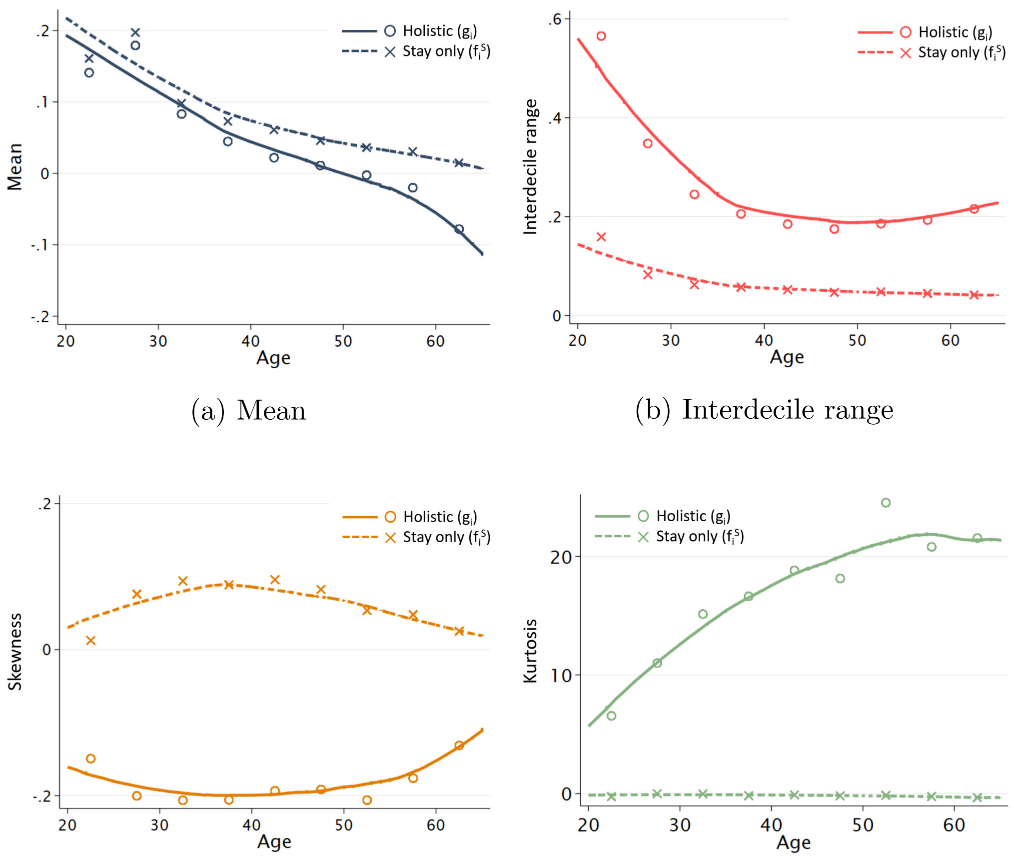
In Panel (c) we consider skewness. For all ages there is, on average, negative skewness in the subjective holistic distributions. Yet when quantifying skewness only from the stay branch, it is close to zero. Negative skewness appears when people expect to disproportionately draw large negative shocks and it indicates that job transitions are, in expectation, responsible for the downside risk that people face.

Finally, in panel (d) we consider kurtosis. According to the holistic measure of subjective earnings growth, kurtosis is significant at a level of about 10-20 and it is increasing in age. When removing risk stemming from job transitions, kurtosis is practically removed. This is consistent with the notion that extreme earnings growth derives from job transitions.

Thus, we find that job transitions are essential in determining life-cycle patterns of higher order moments of subjective holistic one-year ahead earnings risk. This finding is the analog to the differences between job switchers' and job stayers' earnings growth in realizations from administrative data (Guvenen et al., 2021). Our takeaway is that this factor, whether someone expects to make a job transition, also translates over to beliefs about earnings growth.

6 Subjective Earnings Risk in a Search Model

Thus far, we have established several new facts about subjective earnings risk. We found that subjective earnings risk as reported by our survey respondents is about one-third of what could be inferred from administrative data on earnings realizations. The realized distribution of earnings growth confounds workers' true beliefs about their own risk with underlying heterogeneity in expected growth rates. Controlling for a variety of observable characteristics fails to mitigate this discrepancy. This suggests the presence of a form of unobserved heterogeneity that impacts beliefs about earnings risk. We have also learned that within-person, respondents' reported beliefs were broadly in line with their realized earnings. This points to a latent variable related to agents' private information about their own earnings. This would indeed lead an econometrician to overestimate risk when pooling data on realized earnings growth across the population. Additionally, we also learned that job transitions are important shapers of subjective risk. Therefore, this latent variable must also be related to future job transitions.



(a) Mean

(b) Interdecile range

(c) Skewness

(d) Kurtosis

Note: The figure shows the average value of the 1st to 4th quantile based moments (See notes to Figure 3) over the life cycle of holistic earnings risk, g_i, and risk conditional on staying, f_i^S. "o" and "x" represent the empirical mean across 5-years age bins. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights. Online Appendix C.8 show the corresponding figure using standard moments instead of quantile based moments.

Figure 10: Moments of holistic earnings risk, g_i, vs. risk conditional on staying, f_i^S, over the life cycle

In this section, we use a search model to better understand this latent variable that underlies our main results. We focus on the model developed by Menzio et al. (2016).[23] This model is designed to explain the life-cycle profiles of the employment-to-employment (EE), employment-to-unemployment (EU), unemployment-to-employment (UE) rates, and average earnings. An important ingredient is the productivity of firm-worker matches, called match quality. It is not always known to the firm and the worker. Its dynamics result in endogenous age variation in these transition rates, time out of work, and earnings. In contrast, other models include features such as consumption and saving, welfare systems, firm heterogeneity, and/or more elaborate human capital dynamics (for example, Low et al., 2010; Hubmer, 2018; Bagger and Lentz, 2019; Jung and Kuhn, 2019), often as part of an effort to address topics over and above just job transition patterns. While all these models, including Menzio and Shi (2011) and Menzio et al. (2016), have many common features, the latter is particularly well-suited for understanding earnings risk, job transitions, and for the data that we have at hand.

After calibrating the model, we back out the implied subjective beliefs of workers in the model. We then establish that these line up broadly with beliefs from the survey. This result allows us to use the model to better interpret the facts about subjective earnings risk that we discovered in Sections 4 and 5. In particular, we argue that match quality is a plausible candidate for the latent variable that drives subjective beliefs.

6.1 Model Description

We now provide a brief outline, relegating details to Online Appendix F. The model features workers with finite lifespans who can search for work both on and off the job, and face unemployment risk. They acquire experience while employed and aim to find jobs with a high match quality.

The economy is populated by T overlapping generations of risk-neutral workers. In every period, a new set of workers is born who also live for T periods. Workers discount the future at rate $\beta \in (0, 1)$ and maximize their present discounted sum of utility. Workers can either be employed (matched with a firm) or unemployed.

There is a continuum of firms that when matched with a worker, produce output $zg(y)$. The first component, z, is a match quality that is specific to each firm-worker pair. The second component, $g(y)$, is specific to the worker. y represents the worker's experience,

[23] The model in this paper is the life-cycle extension of the well-known directed search model of Menzio and Shi (2011).

which is the cumulative number of periods that they have been employed. The function g maps y into productivity and is increasing and concave.

Workers and firms search for each other within submarkets indexed by (x, y, t): workers of experience level y and of age t choose the level of lifetime utility x that they want to search for. Firms that post vacancies in submarket x must provide that utility to their employees through their employment contract. Each submarket will have an endogenous market tightness, a ratio of vacancies to unemployed, denoted by $\theta_t(x, y)$. The workers' and firms' of x, and therefore which submarkets will have searchers, are determined in equilibrium. Workers can search both on and off the job.

The aggregate state of the economy is $\psi = (n, u, e, \gamma)$. $n(t)$ is the measure of workers of age t in the labor market. $u(y, t)$ is the measure of unemployed workers. $e(z, y, t)$ is the measure of workers of type (y, t) with match quality z. When matches first form, the quality is unknown: $z = z_0$ denotes this case. γ is the measure of newly born workers.

Each period of time consists of five stages, which occur in the following order: 1) entry and exit from the labor market, 2) separation, 3) search, 4) matching, and 5) production.

During the entry and exit stage, non-participating workers of age t enter the labor market with probability μ_t. A fraction ν_t of participating workers permanently leave the labor market where $\nu_{T+1} = 0$, i.e., if a worker reaches age T, they will permanently exit for sure next period.

In the separation stage, workers and firms who remain matched after the previous period decide whether to separate. There are two different types of separations. They can occur exogenously with probability δ. Endogenous separations can also occur: they are determined by age, experience, and the discovery of match quality. The details will be explained further when defining the value functions.

In the search stage, workers get the opportunity to search with probability λ_e and unemployed workers search with probability λ_u. If they do search in that period, they choose a single submarket x where they direct their search. At the same time, firms choose how many vacancies to open in each submarket (taking into account workers' decisions), where k is the cost of posting a vacancy.

In the matching stage, a worker searching in submarket (x, y, t) meets a vacancy with probability $p(\theta_t(x, y))$. p is a matching function that governs how likely workers are to meet a firm as a function of the market tightness (the ratio of vacancies to unemployed). $q(\theta_t(x, y)) = p(\theta_t(x, y))/\theta_t(x, y)$ is the probability that a vacancy meets a worker in submarket (x, y, t). When a firm and a worker meet, the firm offers a contract worth x in lifetime utility. If the worker accepts the offer, then they become a match.[24] At this point, the match quality z is drawn from distribution $f(z)$, but may or may not yet be observable to the firm and worker. In addition, for existing matches, the match quality z is redrawn with probability η from the same distribution. This reflects exogenous changes in productivity that can make this particular match better or worse: the firm implements a new technology, the worker gets better at their tasks in this job, etc.

The last stage is production. Unemployed workers produce and consume b. Employed workers produce $zg(y)$ and consume their earnings w, which is specified by their employment contract (along with the policies for separation rates and which submarket the worker should search in as a function of the history of the match). With probability α, the worker and firm observe z and become a "known quality" match from now on. With probability $1 - \alpha$ they remain as an "unknown quality" match.

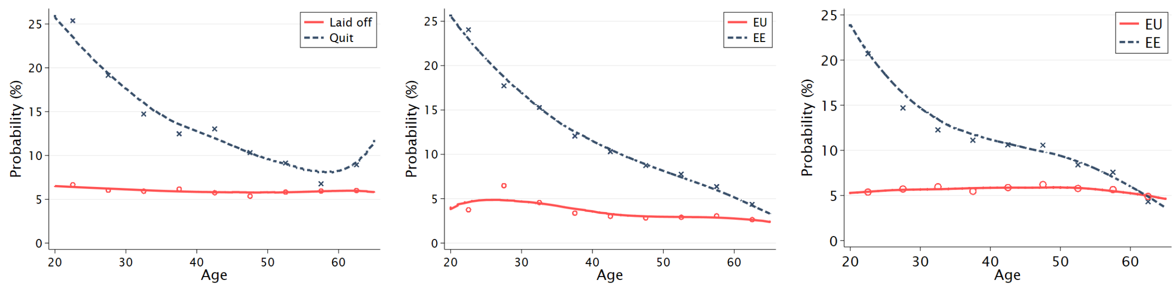
6.2 How the Model Works

This section gives a brief overview of the mechanisms and key forces driving the model. In particular, we highlight how job transitions and earnings risk unfold and where they come from – these are key objects that we will link back to our survey responses and the Danish register. For more details on the value functions and the equilibrium see Online Appendix F.1 and F.2.

Transitions from employment to unemployment are triggered by changes in match quality z. Every match has a reservation match quality $r_t(y)$. If z is below the $r_t(y)$, the match is immediately destroyed. If z is above, it is kept. Updates to z occur in two scenarios: when z is revealed after being unknown in a new match (with probability α) and when it is redrawn (with probability η). Any of these scenarios can result in an EU transition.

Employment-to-employment transitions occur when workers with low enough match quality successfully search on-the-job. Workers optimally choose a single submarket to search in as a function of their current (y, z, t). In equilibrium, workers with lower match quality will choose to search in submarkets where jobs are easier to find. If z becomes high enough, workers do not search on the job at all because it is better not to risk losing their good match to go to a new match with initially unknown quality. As a result, the workers who go through an EE transition will be the ones who have the most to gain

[24] In equilibrium, workers accept all jobs offered to them: they have optimally chosen their submarkets and know exactly the promised lifetime utility x of any job offer.



(a) Survey (b) Registry (c) Model

Note: Population weights are used in Panel (a). The lines show local regression smoothed lines and the scatter plots show the empirical mean of each transition probability in 5-year age bins.

Figure 11: Job transition probabilities

from the switch.

Finally, earnings are linked to human capital (experience) and match quality. Growth in either of these will result in earnings growth. Earnings risk comes from one of two sources. First, a job transition will result in a new match quality, and in some cases, a flattening of experience (if the transition involves going through unemployment). Thus, just as in the survey, job transitions in the model will be closely tied with earnings risk. In addition, earnings risk is also present if the worker stays with their current employer because of the possibility of discovering or resetting their match quality.

6.3 Calibration and Belief Simulation

We calibrate the model using data on employment and earnings outcomes from the registry, where the raw data is measured at the monthly frequency. This approach allows us to learn, using data that is typically available to researchers, whether the model-implied beliefs have similar properties to subjective beliefs we measured. The key moments that we target are the EU and EE rates as a function of tenure, and average earnings as a function of age. For more details on the calibration and model fit, see Online Appendix F.3.

This calibration strategy also has the added benefit of allowing the model to endogenously generate the age profile of the EU and EE rates. This ensures that the model can deliver on its own the correct transition patterns by age. Even though these were not targeted, panels (b) and (c) of Figure 11 show that the model does a remarkable job at matching the registry along these dimensions. This confirms that the model's mechanisms are a good starting point for understanding job transitions over the life cycle, a key ingredient for earnings risk. It is also remarkable that the survey patterns match up as well[25] – this says that on average and within age groups, people are correct about the chances of undergoing either one of the transition types. The alignment of these three objects will be important later for understanding the model's patterns in beliefs. Online Appendix F.3 further elaborates on the model's performance with respect to other untargeted moments.

With this calibration in hand, our next step is to generate the beliefs for a cross-section of model-simulated workers. Specifically, we want to create the model counterparts of g_i, exactly as if agents in the model were respondents to the CLP. To do this, we take a set of workers from the model and simulate their lives forward for one year many times and use these simulations to recover a distribution of one-year ahead earnings growth beliefs.

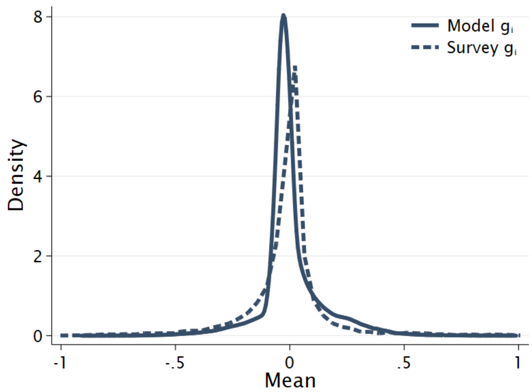
Specifically, we start from a sample of model-simulated workers drawn from the stationary equilibrium, which is a distribution of workers over age, experience, employment status, and match quality. Then for each worker and age in the data, we simulate their life-cycle forward from January to December. We do this 5,000 times per worker and age, in which each simulation begins from the same starting point. Over the next year, they may experience a job separation, change in match quality, etc. Each of these paths represents one realization of the worker's beliefs about all of the outcomes that are possible over the next year. Note that here we are imposing rational expectations as do nearly all models in this literature: in expectation, beliefs are the same as outcomes. On each of these simulated paths, we can then calculate their total earnings and by taking the log difference with their earnings over the previous year, we obtain a distribution analogous to g_i in the survey for each simulated model agent.

6.4 Comparing Survey and Model Beliefs

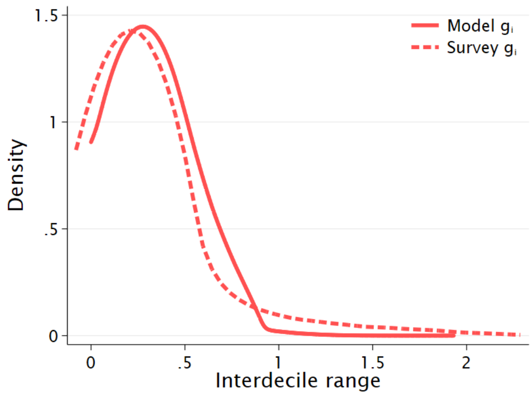
In this section, we compare between the beliefs g_i in the model and survey. We first compare the distributions of means and interdecile ranges of g_i. These are depicted in Figure 12.

The results show that overall, the model replicates the levels and shapes of these distributions remarkably well. In both the survey and the model, most people expect themselves

[25] Note that here, to facilitate our comparison, we are implicitly equating quits in the CLP with EE in the model and layoffs in the CLP with EU in the model. The reason is that from the perspective of the worker, quits and EE transitions are usually interpreted as voluntary and layoffs and EU transitions are seen as involuntary. Of course in most models, these transitions are mutually agreeable and in real life, there can be involuntary EE transitions and voluntary EU transitions.



(a) Mean



(b) Interdecile range

Note: The figures show kernel densities of the mean of g_i (left panel) and the interdecile range of g_i, comparing the model and the survey. The bandwidths are 0.01 for the both mean densities and 0.2 for both interdecile range densities.

Figure 12: Heterogeneity in means and interdecile ranges of g_i in the model and survey

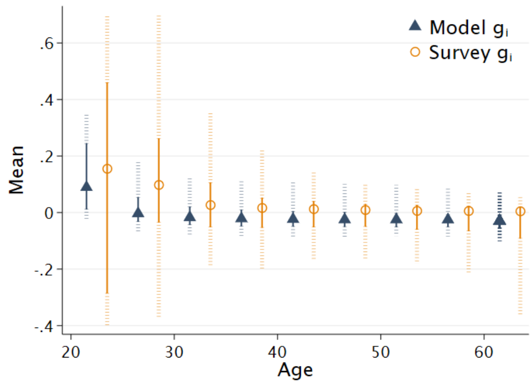
to have low but positive mean earnings growth. There are individuals who are outliers in both directions and the model picks this up as well. The interdecile ranges also have similar shapes. Like in the survey, most agents in the model have low levels of subjective risk, but there is a long right tail of agents with much higher levels. In fact, in both the model and the survey, there are very few people with interdecile ranges above 0.75.

We further explore the belief heterogeneity by comparing these moments across the life cycle in Figure 13. The results here expand on the ones in Figure 10 for the survey. They show means and interdecile ranges of g_i in five-year age bins. In each age bin and for both the mean and interdecile range of g_i, we calculate the median as well as the 10th, 25th, 75th, and 90th percentiles to show some dispersion. Once again, the model is able to replicate the basic patterns in g_i across age. The moments are downward-sloping as a function of age and are of similar levels in the model and the survey.

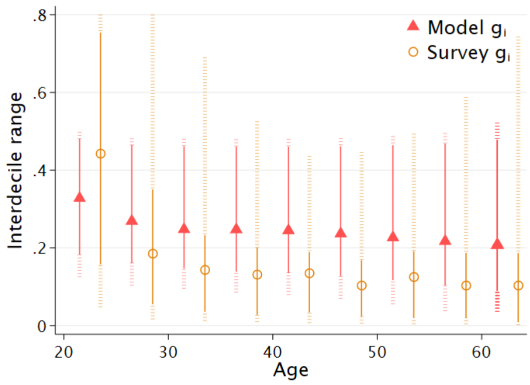
The model can also capture certain features of the heterogeneity within age groups. For the mean, it correctly predicts more heterogeneity for younger workers. Regarding the interdecile range, the dispersion in the model is relatively consistent with that of the survey for middle-aged workers.

Thus we find that all-in-all, this simple, off-the-shelf model is successful at capturing the basic features of workers' beliefs.[26] Why is this the case, despite the fact that the

[26] Note that our quantification here is nearly equivalent to what we would get had we targeted features



(a) Mean



(b) Interdecile range

Note: The markers represent the median within 5-year age bins. The top and bottom of the solid lines represent the 75th and 25th percentiles, respectively. The top and bottom of the dashed lines represent the 90th and the 10th percentiles, respectively.

Figure 13: Life-cycle patterns in means and interdecile ranges of g_i in the model and survey

model was calibrated to administrative data? The answer is that the model is set up to explain and match job transition probabilities over the life cycle. As we showed earlier, the survey and registry data are well-aligned in this aspect. Moreover, even though these are not explicitly targeted in the model, they still match the registry outcomes which means that the model also aligns well with the survey, as shown previously in Figure 11. In addition, Figure 10 revealed that accounting for job transitions was important for recovering the correct patterns in beliefs, g_i. Hence, by accurately modeling beliefs about job transitions, the model at the same time does a good job at accounting for overall beliefs about earnings risks. From this finding we conclude that in order to to capture the beliefs about earnings risk documented from our data, one should explicitly model the search process and account for job transitions, rather than relying on a reduced-form income process identified off of the variance of earnings growth in administrative data (h_A).

6.5 Match Quality and Heterogeneity in Subjective Earnings Risk

Because the model captures well the features and heterogeneity of g_i, we are in a position to use it to interpret the findings from our survey. Broadly, it appears that the mech-

of the subjective beliefs directly.

anisms in the model (which are also found in many other models of search) like match quality, human capital accumulation, and climbing the job ladder go a long way towards rationalizing the beliefs collected in the survey. More specifically, we point out that match quality may be related to the unobserved heterogeneity that our survey suggests the presence of.

First, match quality does not have a data counterpart. It is not observable in any administrative data set and is not captured by earnings because earnings also depend on age and experience. Moreover, it is an abstract concept that would be difficult, if not impossible, to elicit directly in a survey. If a latent variable like match quality is responsible for heterogeneity in subjective expectations, then it would be no surprise that we found very little role for observables in explaining $\mathbb{E}_i[g_i]$ in Section 4.3.

We found in Section 5 that subjective beliefs are connected to job transition beliefs; in the model, match quality is the key motive for transitions. Thus, heterogeneity in match quality leads to heterogeneity in subjective earnings risk. For example, workers with high match quality expect to stay at their jobs for a long time and do not anticipate likely transitions in the short term. Workers with lower match quality search on the job more intensely and therefore expect to make a job-to-job transition soon. In this situation, there is risk coming from this choice due to the fact that the match quality in the new job is not yet known. Finally, workers in matches whose quality is yet to be revealed face the highest probability of transitioning to unemployment in the future.

Finally, match quality varies over the age and earnings distributions in particular ways. This is a key reason why the major patterns in Figure 13 match up well. But of particular importance is the dispersion of the moments of g_i. The dispersion in the mean of g_i declines over the life-cycle: this is connected to the fact that there is less dispersion in match quality as workers age. Although many people take some time to find good matches, some find them relatively quickly, creating dispersion in mean growth prospects for young workers. In contrast, as workers age, more and more of them settle into good matches and reduce their likelihood of making future transitions.[27] In the survey, the departure between subjective earnings risk and its administrative counterpart is largest for young and low-earners. This is exactly what we would expect if match quality is a latent variable that plays into subjective earnings beliefs. In other words, workers have private information about match quality, and therefore their expected job transitions. Because match quality is heterogeneous, especially for young and low-earnings workers, this leads to heterogeneity in $\mathbb{E}_i[g_i]$, which is consistent with our interpretation of the

[27] The same is true for earnings: there is more dispersion in match quality among lower-earning workers.

mixture result in Section 4.

7 Conclusion

We introduce a survey instrument that measures earnings risk. A key feature of our instrument is that it conditions on possible job transitions, i.e., whether people stay in their current job, quit or are laid off. A link with administrative data provides many credibility checks. We show evidence that subjective expectations about earnings and job transitions are consistent with earnings realizations when appropriately aggregated. Remarkably, this is true across the age distribution, even though job transitions and earnings growth vary significantly across age. These findings give confidence in the validity of the survey.

The linked survey and administrative data also reveals *subjective* earnings risk to be two to six times lower than its counterpart estimated from administrative data on realized earnings growth alone. This is because expected earnings growth is heterogeneous and is confounded with risk when inferred from realized earnings growth even when calculated within narrow cells defined by observable characteristics. This finding shows that while observable characteristics, including among other things age, education, the earnings level and past earnings growth, are significant in predicting heterogeneity in earnings growth, such characteristics do not account adequately for the relevant heterogeneity.

The survey data shows that subjective earnings risk primarily relates to possible job transitions. We calibrate a structural model of job search over the life-cycle to match realized job transitions by tenure and the level of earnings and show that this model is broadly able to replicate the distribution of subjective beliefs about earnings growth reported in the survey. The model suggests that match quality is critical for understanding subjective earnings risk. Also, the model rationalizes individuals having superior information about their expected earnings growth. This can explain why subjective earnings risk deviate from risk inferred from administrative data on earnings growth because the latter confounds heterogeneity in expected earnings growth and subjective risk. Combining all our insights, we conclude that in order to capture beliefs about earnings risk, one should explicitly model the search process and account for job transitions, rather than relying on a reduced-form income process identified from the variance of earnings growth in administrative data

More broadly, our findings highlight the value of using survey-based measures of subjective earnings expectations to understand the nature of labor market and earnings risk.

This has implications for understanding and modelling search and savings behavior. For example, people who face a low level of risk need limited precautionary savings and our findings could contribute to explaining why many households hold limited liquidity. Moreover, this may interact with search behavior where people can vary search efforts to reduce the impact of adverse earnings shocks by adjusting search intensity.

Acknowledements: We thank Anmol Bhandari, Andres Blanco, Martin Browning, Charles Manski, Guido Menzio, Christian Moser, Andreas Mueller, Serdar Ozkan, Todd Schoellman, and Basit Zafar, as well as seminar/conference participants at Dartmouth, the St. Louis Fed, University of Copenhagen, University of Michigan, LAEF Labor Markets and Macroeconomic Outcomes Conference, Society for Economic Dynamics, NBER Summer Institute 2023, the 7th Workshop on Subjective Expectations, the 3D-In-Macro workshop at Oslo BI, USC, and the Minneapolis Fed for comments. The work presented in this paper benefited from financial support from Center for Economic Behavior and Inequality, CEBI, a center of excellence at the University of Copenhagen financed by grant DNRF134 from the Danish National Research Foundation. Caplin thanks the Alfred P. Sloan Foundation and the NOMIS Foundation for their support for the research on the cognitive foundations of economic behavior and acknowledges support for this research from Sloan Foundation Grant 2015-14209. The views expressed herein are solely those of the authors and do not necessarily reflect those of the Federal Reserve Bank of St. Louis or the Federal Reserve System.

References

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

Andersen, H. Y. and S. Leth-Petersen (2021). Housing wealth or collateral: How home value shocks drive home equity extraction and spending. *Journal of the European Economic Association 19*(1), 403–440.

Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020). Ex ante returns and occupational choice. *Journal of Political Economy 128*(12), 4475–4522.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review 104*(6), 1551–96.

Bagger, J. and R. Lentz (2019). An empirical model of wage dispersion with sorting. *The Review of Economic Studies 86*(1), 153–190.

Balleer, A., G. Duernecker, S. Forstner, and J. Goensch (2021). The effects of biased labor market expectations on consumption, wealth inequality, and welfare.

Balleer, A., G. Duernecker, S. Forstner, and J. Goensch (2023). Wage bargaining and labor market policy with biased expectations.

Bick, A., A. Blandin, and K. Mertens (2021). Work from home before and after the covid-19 outbreak. *Available at SSRN 3786142*.

Browning, M., M. Gørtz, and S. Leth-Petersen (2013). Housing wealth and consumption: a micro panel study. *The Economic Journal 123*(568), 401–428.

Busch, C., D. Domeij, F. Guvenen, and R. Madera (2022). Skewed idiosyncratic income risk over the business cycle: Sources and insurance. *American Economic Journal: Macroeconomics 14*(2), 207–42.

Campbell, D., A. Carruth, A. Dickerson, and F. Green (2007). Job Insecurity and Wages. *The Economic Journal 117*(518), 544–566.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from denmark. *The Quarterly Journal of Economics 129*(3), 1141–1219.

Conlon, J. J., L. Pilossoph, M. Wiswall, and B. Zafar (2018). Labor market search with imperfect information and learning. Technical report, National Bureau of Economic Research.

De Bruin, W. B., C. F. Manski, G. Topa, and W. van der Klaauw (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics 26*(3), 454–478.

Delavande, A. and S. Rohwedder (2008). Eliciting subjective probabilities in internet surveys. *Public Opinion Quarterly 72*(5), 866–891.

Dominitz, J. (1998). Earnings expectations, revisions, and realizations. *Review of Economics and statistics 80*(3), 374–388.

Dominitz, J. (2001). Estimation of income expectations models using expectations and realization data. *Journal of Econometrics 102*(2), 165–195.

Dominitz, J. and C. F. Manski (1997). Using expectations data to study subjective income expectations. *Journal of the American statistical Association 92*(439), 855–867.

Druedahl, J. and A. Munk-Nielsen (2020). Higher-order income dynamics with linked regression trees. *The Econometrics Journal 23*(3), S25–S58.

Epper, T., E. Fehr, H. Fehr-Duda, C. T. Kreiner, D. D. Lassen, S. Leth-Petersen, and G. N. Rasmussen (2020, April). Time discounting and wealth inequality. *American Economic Review 110*(4), 1177–1205.

Faberman, R. J., A. I. Mueller, A. Şahin, and G. Topa (2022). Job search behavior among the employed and non-employed. *Econometrica 90*(4), 1743–1779.

Goldstein, D. G. and D. Rothschild (2014). Lay understanding of probability distributions. *Judgment & Decision Making 9*(1).

Guiso, L., T. Jappelli, and L. Pistaferri (2002). An empirical analysis of earnings and employment risk. *Journal of Business & Economic Statistics 20*(2), 241–253.

Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic dynamics 12*(1), 58–79.

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021). What do data on millions of us workers reveal about lifecycle earnings dynamics? *Econometrica 89*(5), 2303–2339.

Guvenen, F., L. Pistaferri, and G. L. Violante (2022). Global trends in income inequality and income dynamics: New insights from grid. *Quantitative Economics 13*(4), 1321–1360.

Halvorsen, E., H. A. Holter, S. Ozkan, and K. Storesletten (2023). Dissecting idiosyncratic earnings risk. *Journal of the European Economic Association*.

Hendren, N. (2017). Knowledge of Future Job Loss and Implications for Unemployment Insurance. *American Economic Review 107*(7), 1778–1823.

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

Hvidberg, K. B., C. T. Kreiner, and S. Stantcheva (2023). Social positions and fairness views on inequality. *The Review of Economic Studies 86*(1), 253–269.

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

Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2022). Worker beliefs about outside options. Technical report, National Bureau of Economic Research.

Jung, P. and M. Kuhn (2019). Earnings losses and labor mobility over the life cycle. *Journal of the European Economic Association 17*(3), 678–724.

Karahan, F., S. Ozkan, and J. Song (2022). Anatomy of lifetime earnings inequality: Heterogeneity in job ladder risk vs. human capital. *FRB St. Louis Working Paper* (2022-2).

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica 79*(3), 651–692.

Koşar, G. and H. W. van der Klaauw (2023). Workers' perceptions of earnings growth and employment risk. Working Paper 1056, Federal Reserve Bank of New York.

Leth-Petersen, S. (2010). Intertemporal consumption and credit constraints: Does total expenditure respond to an exogenous shock to credit? *American Economic Review 100*(3), 1080–1103.

Leth-Petersen, S. and J. Sæverud (2022). Inequality and dynamics of earnings and disposable income in denmark 1987–2016. *Quantitative Economics 13*(4), 1493–1526.

Low, H., C. Meghir, and L. Pistaferri (2010). Wage risk and employment risk over the life cycle. *American Economic Review 100*(4), 1432–67.

Manski, C. F. and J. D. Straub (2000). Worker perceptions of job insecurity in the mid-1990s. *The Journal of Human Resources 35*(3), 447–479.

Menzio, G. and S. Shi (2011). Efficient search on the job and the business cycle. *Journal of Political Economy 119*(3), 468–510.

Menzio, G., I. A. Telyukova, and L. Visschers (2016). Directed search over the life cycle. *Review of Economic Dynamics 19*, 38–62.

Moscarini, G. and F. Postel-Vinay (2013). Stochastic search equilibrium. *Review of Economic Studies 80*(4), 1545–1581.

Mueller, A. I., J. Spinnewijn, and G. Topa (2021). Job seekers' perceptions and employment prospects: Heterogeneity, duration dependence, and bias. *American Economic Review 111*(1), 324–63.

Pistaferri, L. (2001). Superior information, income shocks, and the permanent income hypothesis. *Review of Economics and Statistics 83*(3), 465–476.

Pistaferri, L. (2003). Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply. *Journal of Labor Economics 21*(3), 729–754.

Stephens, M. (2004). Job Loss Expectations , Realizations , and Household Consumption Behavior. *The Review of Economics and Statistics 86*(1), 253–269.

Stoltenberg, C. A. and A. Uhlendorff (2022). Consumption choices and earnings expectations: Empirical evidence and structural estimation. Working Paper 15443, IZA.

Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics 107*(2), 439–479.

Von Wachter, T., J. Song, and J. Manchester (2009). Long-term earnings losses due to mass layoffs during the 1982 recession: An analysis using us administrative data from 1974 to 2004. *unpublished paper, Columbia University.*

Wang, T. (2023). Perceived versus calibrated income risks in heterogeneous-agent consumption models.

Wiswall, M. and B. Zafar (2021). Human capital investments and expectations about career and family. *Journal of Political Economy 129*(5), 1361–1424.

Online Appendix for "Subjective Earnings Risk"

Andrew Caplin, New York University
Victoria Gregory, FRB St. Louis
Eungik Lee, New York University
Søren Leth-Petersen, University of Copenhagen
Johan Sæverud, University of Copenhagen

January 3, 2024

A Survey Questionnaire

In this appendix, we introduce the survey questionnaire of Copenhagen Life Panel. To set the stage, we start by asking questions about work status in the past year. In the survey fielded in January 2021, the first question is about labor market status one year earlier, in January 2020. Respondents were asked to classify themselves as one of: employed; self-employed; looking for work; temporarily out of labor force; and permanently out of labor force.

Respondents then report their current work status. They again can choose from among the same five options. If they report there is a change between the past and current status, we further ask about the dynamics. Then they move to the module with expectations questions if they are currently employed. The full questionnaire is as follows:

First, we are interested in your working status in Jan 2020 (last year).

Throughout this survey, we are going to ask you about working for pay. We have in mind work, including work as self-employed, for which you receive regular pay and work at least 10 hours per week.

[Part 2-1: Labor income, Past]

Q_2_1. Think about Jan 2020 (last year).

Were you working for pay?

Variable name: Q_2_1

Variable value: 1 Yes 2 No

[If Q_2_1=Yes] Q_2_1_2. Think about Jan 2020 (last year).

Were you self-employed?

Variable name: Q_2_1_2

Variable value: 1 Yes 2 No

[If Q_2_1=No] Q_2_1_3. Did you look for work in Jan 2020?

Variable name: Q_2_1_3

Variable value: 1 Yes 2 No

[If Q_2_1_3=No] Q_2_1_4. Please pick the most appropriate description of your employment status in Jan 2020 (last year).

Variable name: Q_2_1_4

Variable value: 1 Temporarily out of work, 2 Permanently out of work

[Note: clarification Johan]

[If J_past= Temporarily Out/Looking for work] Q_2_2. Did you have any earned income during 2020?

Variable name: Q_2_2

Variable value: 1 Yes 2 No

[If Q_2_1=Yes or Q_2_2=Yes] Q_2_3. What was your earned income during 2020?

Please report the most accurate amount you think.

Variable name: Q_2_3

Variable value: [0-]

[Part 2-2: Labor income, Dynamics]

[If J_past=E WFP] Q_2_4. From Jan 2020 till now, are you still working for pay with the same employer you had in Jan 2020?

Variable name: Q_2_4

Variable value: 1 Yes 2 No

[if Q_2_4=No] Q_2_4_2. When did you stop working with this employer you had in Jan 2020?

Variable name: Q_2_4_2

Variable value: From Jan-1-2020 – to Jan-2021

[if Q_2_4=No] Q_2_4_3. For what reason, did you stop working with this employer you had in Jan 2020?

Variable name: Q_2_4_2

Variable value: 1.Laid-off, 2 Quit, 3 Other

[if Q_2_4=No] Q_2_4_4. After you stopped working for this employer you had in Jan 2020, did you find other work for pay within a month?

Variable name: Q_2_4_4

Variable value: 1 Yes 2 No

[if Q_2_4_4=No] Q_2_4_5. How many months were you out of work after you stopped working for the employer you had in Jan 2020.

Variable name: Q_2_4_5

Variable value: [1-12, I did not work for pay after that]

[Part 2-2: Labor income, Now]

We now ask about your current employment status.

[if Q_2_4!=Yes, Q_2_5!=Yes] Q_2_7. Are you currently working for pay?

Variable name: Q_2_7

Variable value: 1 Yes 2 No

[If Q_2_7=Yes] Q_2_7_2. Are you currently self-employed?

Variable name: Q_2_7_2

Variable value: 1 Yes 2 No

[If Q_2_7=No] Q_2_7_3. Are you currently looking for work for pay?

Variable name: Q_2_7_3

Variable value: 1 Yes 2 No

[If Q_2_7_3=No] Q_2_7_4. Please pick the most appropriate description of your current employment status

Variable name: Q_2_7_4

Variable value: 1 Temporarily out of work, 2 Permanently out of work

[Part 2-3: Labor income, Future]

[if J_now=E WFP] Q_2_8. Please think about your possible relationship with your current employer in 2021. Assign the probability in each possible case. The sum of the probability should be 100.

1. Staying with the current employer during 2021
2. Laid-off from current employer at some point during 2021
3. Quit from the current employer at some point during 2021
4. Separation for other reason [checkbox activated]

Variable name: Q_2_8_1, Q_2_8_2, Q_2_8_3, Q_2_8_4

Variable value: [0-100]

[Stay branch] [if Q_2_8_1>0] Q_2_9. We are interested in your earned income in 2021 if you stay with the current employer throughout the calendar year 2021.

Variable name: Q_2_9_1, Q_2_9_2, Bins and balls

[if Q_2_8_1>0] Q_2_10. Now, please think about your earned income in 2025 if you stay with the current employer throughout the calendar year 2021.

Please think about all possibilities regardless of whether you stay with the same employer after 2021.

Variable name: Q_2_10_1, Q_2_10_2, Bins and balls

[Laid-off branch] [if Q_2_8_2>0] Q_2_11. Suppose you were to be laid off from the current employer during 2021. What is the probability that you would start working for pay again after your current work terminates?

Variable name: Q_2_11_1 (within 1 month), Q_2_11_2 (within 3 months), Q_2_11_3 (within 1 year), Q_2_11_4 (within 2 years) **[Custom 58]**

[if Q_2_11_4>0] Q_2_12. Suppose you were to be laid off from the current employer during 2021. After then you start to work for pay at some point in 2 years.

Think about the possible earned income from the first 12 months of this new work for pay.

Variable name: Q_2_12_1, Q_2_12_2, Bins and balls

[if Q_2_11_4>0] Q_2_13. Suppose you were to be laid off from the current employer during 2021. After then you start to work for pay at some point in 2 years.

Think about the possible earned income in 2025.

Think about all the possibilities between reemployment and 2025.

Variable name: Q_2_13_1, Q_2_13_2, Bins and balls

[Quit branch] [if Q_2_8_3>0] Q_2_14. Suppose you were to quit from the current employer during 2021. What is the probability that you would start working for pay again after you quit?

Variable name: Q_2_14_1 (within 1 month), Q_2_14_2 (within 3 month), Q_2_14_3 (within 1 year), Q_2_14_4 (within 2 years)

[if Q_2_14_4>0] Q_2_15. Suppose you quit from the current employer during 2021. After then you start to work for pay at some point in 2 years.

Think about the possible earned income from the first 12 months of this new work for pay.

Variable name: Q_2_15_1, Q_2_15_2, Bins and balls

[if Q_2_14_4>0] Q_2_16. Suppose you quit from the current employer during 2021. After then you start to work for pay at some point in 2 years.

Think about the possible earned income in 2025.

Think about all the possibilities between reemployment and 2025.

Variable name: Q_2_16_1, Q_2_16_2, Bins and balls

[Other branch] [if Q_2_8_4>15] Q_2_17. Suppose you were to be separated because of other reasons you are thinking of (not laid-off or quit).

What is the probability that you would start working for pay again after your current work?

Variable name: Q_2_17_1 (within 1 month), Q_2_17_2 (within 3 months), Q_2_17_3 (within 1 year), Q_2_17_4 (within 2 years)

[if Q_2_17_4>0]Q_2_18. We are interested in your earned income in the first-year and 2025 earned income after you are separated for other reasons and then reemployed.

Suppose you were to be separated for other reasons from the current employer during 2021. After then you start to work for pay at some point in 2 years.

Think about the possible earned income from the first 12 months of this new work for pay.

Please report your lowest and highest possible amount.

Variable name: Q_2_18_1, Q_2_18_2, Bins and balls

[if Q_2_17_4>0] Q_2_19. Suppose you were to be separated for other reason from the current employer during 2021. After then you start to work for pay at some point in 2 years.

Think about your earned income in 2025. Think about all the possibilities between reemployment and 2025.

Please report your lowest and highest possible amount.

Variable name: Q_2_19_1, Q_2_19_2, Bins and balls

B Comparison to Administrative Data

Table B.1 compares the demographics of the survey sample with that of the comparable registry population. The first column is for all survey participants and the second column is for employed samples. The third column is for everyone in the registry and the fourth column is for workers with annual earnings greater than 24,000 DKK. We note that the share of females is similar in the survey and the registry. Survey participants are, on average, older, more educated, and have higher earnings than the registry population.

Table B.1: Survey sample demographics

	Survey		Registry	
Working status	All	Employed	All	Employed
N	14,875	10,945	3,429,071	2,700,069
Female	0.48	0.48	0.5	0.48
Age				
Mean	48.00	47.44	42.50	42.56
S.D.	12.54	11.87	13.26	12.77
Distributions				
20-29	0.11	0.10	0.22	0.21
30-39	0.15	0.16	0.20	0.21
40-49	0.22	0.24	0.22	0.24
50-59	0.31	0.33	0.23	0.24
60-65	0.21	0.17	0.12	0.10
Education				
Above college	0.49	0.53	0.34	0.39
Annual Earnings (unit: DKK)				
Mean	421,278	495,783	327,525	418,624
S.D.	382,719	289,681	430,482	442,410

Note: The table compares average demographic characteristics of different subsamples of the Danish population observed in the administrative data in 2020. The column 'Survey, All' includes gross sample of survey participants. The column 'Survey, Employed' includes the subset of survey participants who were employed a the time of the survey (January 2021). The column 'Registry, All' includes all individuals in the Danish population belonging to the cohorts from which the sample is drawn. Survey participants are excluded from this sample. The column 'Registry, Employed' includes the subset of the previous column with earned income of at least 24,000 DKK in 2020. In Jan-2021, the exchange rate for 1 US Dollar was approximately 7 Danish Krone (DKK).

In the analysis we scale all statistics by the relative population weights. To construct population weights, we the Danish population observed in 2020 in the administrative data. We estimate a probit regression with a survey participation dummy as the dependent variable and age, log earnings, education, and gender as explanatory variables. All these characteristics are available in the administrative data. Table B.2 shows the marginal effect on survey participation using probit regression. Our survey respondents are around 0.37% of the Danish population. We find that the selection into the survey is related to various demographics. For instance, as age increases by one unit, the probability of participation increases by 0.012%.

Table B.2: Marginal effect on participation

	Mean of Pr(participation):0.37%		
	$\frac{dy}{dx} \times 100$	z-statistics	p-value
age	0.012	49.47	<0.001
female	-0.028	-4.45	<0.001
log earnings	0.015	19.38	<0.001
above university	0.228	29.9	<0.001
	N: 2,711,011		
	Log-likelihood: -92,294		

Note: The table presents marginal effects from probit regressions where the dependent variable is a dummy variable for survey participation.

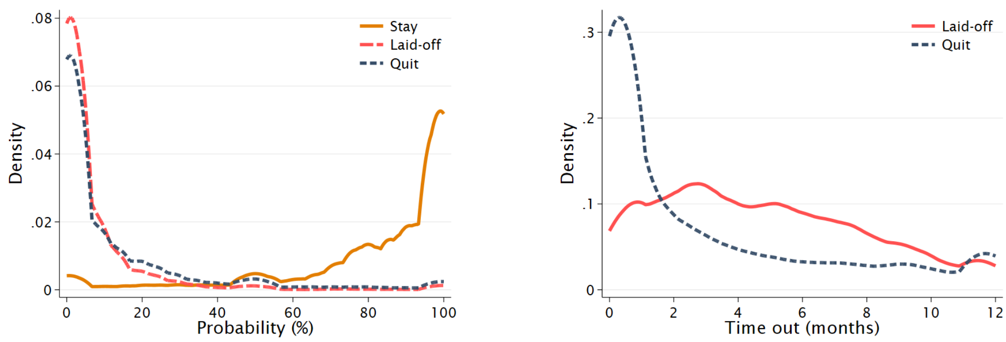
To obtain population weights for the survey, we use the inverse of the predicted probability of participating in the survey. Then we apply these population weights to the analysis and figures in the main text.

C Supplementary Results

C.1 Distributions of Job Transition Probability and Time Out of Labor Force

In this section, we show the distribution of job transition probability and time out of the labor force after the job separation in Fig 4.

Fig C.1 (a) displays the distribution of the probability of stay, laid-off, and quit, respectively. We observe that the probability of quitting is slightly more heterogeneous than the probability of being laid-off, although qualitatively they are quite similar. We also find the probability of stay has the most heterogeneous distribution among the three branches. Fig C.1 (b) illustrates the distribution of time out of the labor force after quitting and being laid off, respectively. Notably, the time out after quitting exhibits peaked distribution where the peak is close to 0. This means most workers expect very short periods of time out of work. In contrast, the time out after being laid off is more similar to a uniform distribution with much less peakedness.



(a) Job transition probability (b) Time out labor force

Note: Panel (a) shows the distribution of the probability of each branch. Panel (b) shows the distribution of time out labor force after quitting and being laid off respectively.

Figure C.1: Distribution of Fig 4.

C.2 Standard Measures of Higher Order Moments

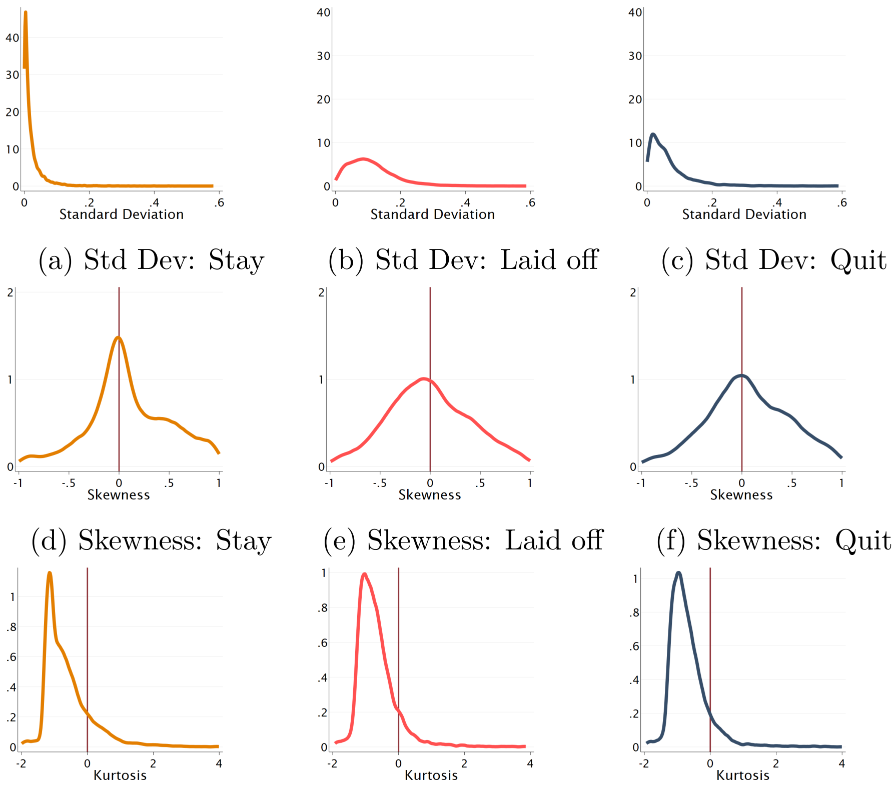
Figure C.2 shows the cross sectional distribution of moments of the subjective distribution calculated using standard measures (standard deviations, skewness, and kurtosis). This figure corresponds to Figure 4 in the paper. The distribution of moments using standard measures are quite similar to the pattern in Figure 4 in the paper which is based on robust measures. Workers tend to have high second moments after being laid-off and quitting, while it is much lower in the stay branch. The skewness is also consistent with the Kelley skewness in Figure 4. Across the branches, it is centered around 0. Lastly, the kurtosis measure which we normalized by 3 (normal distribution) is also consistent with Crow-Siddiqui kurtosis, which is less peaked compared to the normal distribution.

C.3 Job Transitions Based on 2019 Data

Figure C.3 replicate Figure 4 in the main text using data from the 2019 registry. Figure C.3(a) shows the result for the stay probability against age. The pattern is very similar to the pattern presented in Figure 4(a) the main text. Figure C.3(b) replicates Figure 4(b) of the main text but now based on 2018-19 data. We find that the time out work pattern across the life cycle is very consistent.

C.4 Last Year's Earnings

In this section, we compare the subjects' reported earnings for 2020 to actual earnings in 2020 as recorded in the income-tax registry. Figure C.4 shows a binned scatter plot of average earnings reported in the survey (Y-axis) by bins of earnings recorded in the administrative data (X-axis), which is third-party reported by employers directly to the



(a) Std Dev: Stay (b) Std Dev: Laid off (c) Std Dev: Quit

(d) Skewness: Stay (e) Skewness: Laid off (f) Skewness: Quit

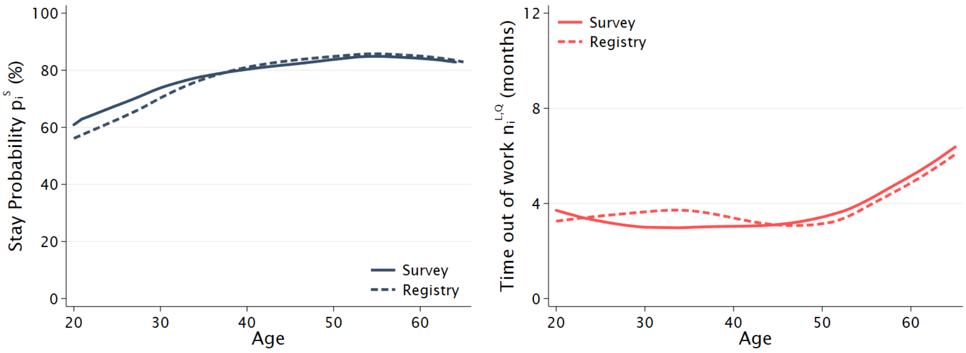
(g) Kurtosis: Stay (h) Kurtosis: Laid off (i) Kurtosis: Quit

Note: Standard moments of the answers to the questions in the conditional survey instrument. Survey results are weighted using population weights.

Figure C.2: Overview of branch-by-branch survey responses: using standard moments

Danish Tax Agency. We find that earnings reported in the survey is very similar to earnings recorded in the administrative data. This suggests that respondents remember their earnings during 2020 well at the point when we surveyed them in January 2021.

We further investigate whether observable characteristics correlated with the deviation between self-reported and registry earnings in 2020. We first construct the log gap between registry earnings and self-reported earnings: $log(Y_{2020}^{\text{registry}}) - log(Y_{2020}^{\text{survey}})$. The mean of this log gap is 0.048 and the standard deviation is 0.26. Table C.1 shows the regression results. In column (1), we include 5-year age bins, with the age 41-45 group as the baseline. Except for the youngest and oldest groups, the coefficients are not significant at the 5% level. Gender and education are also not significant at the 5% level. In column (2), we control for age as a continuous variable and find significant correlations at the 1% level, which means that older workers are relatively over-reporting their current earnings in the survey. When we control for age and the age-squared term in column (3), the coefficients are not individually significant at the 5% level. For all the regressions the R^2 is 0.005, i.e., the observable characteristics explain very little of the deviation. These



(a) Stay probability (b) Time out of work

Note: The figure shows how two components from the survey (solid line) and the administrative data (dashed line) compare over the life-cycle. Panel (a): In the survey, we directly ask about the probability of staying with the same employer for the next 12 months. In the registry, we compute the proportion of workers who stayed from Dec 2018 to Dec 2019 with the same employer. Panel (b): We calculate the expected time out of work after a separation in the survey. In the administrative data we consider job separations that took place during 2018 and follow time spent until reemployment occurs, possibly extending into 2019. In the registry, 1) we exclude the workers going back to the same employer (possibly seasonal work) and 2) we consider the first separation and do not take into account further separations.

Figure C.3: Job separations and time out of work in the survey and in the administrative data for 2019

results show that it is difficult to find systematic under/overreporting in self-reported earnings across the included observable demographics.

C.5 Comparison with 2019 Registry

Figure C.5(b) replicates Figure 5(b) in the paper using administrative data from 2019. The figure shows survey and registry results side by side, and it is very similar to Figure 5 in the main text.

C.6 Registry Earnings Risk using Survey Respondents

Figure C.6 replicates Figures 5(b) in the paper using administrative data for 2020 only for survey respondents. While noisy because it includes fewer observations it compares well to Figure 5(b).

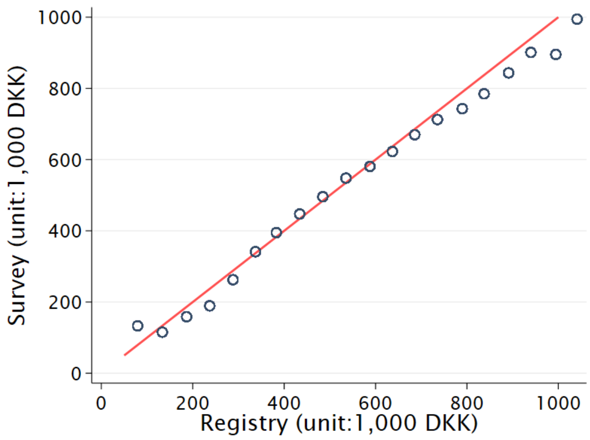
C.7 Standard Moments of h_S^X and h_A^X

Figure C.7 reproduces Figure 7(b)-(d) in the paper using standard deviation, skewness, and kurtosis. Also here we find that the moments are well aligned across the life cycle.

Table C.1: Difference between registry and self-reported survey earnings

Dependent variable:	$log(Y^{r}_{2020}) - log(Y^{s}_{2020})$		
	(1)	(2)	(3)
Age group 20-25	0.046**		
	(0.023)		
Age group 26-30	0.024		
	(0.016)		
Age group 31-35	0.020*		
	(0.012)		
Age group 36-40	0.017		
	(0.012)		
Age group 46-50	0.008		
	(0.010)		
Age group 51-55	-0.010		
	(0.009)		
Age group 56-60	-0.012		
	(0.010)		
Age group 61-65	-0.026***		
	(0.010)		
Age		-0.002***	-0.001365
		(0.000)	(0.002153)
Age-squared			-0.000002
			(0.000023)
Female	0.007	0.007	0.007437
	(0.005)	(0.005)	(0.005282)
University or above	0.008	0.008	0.007668
	(0.005)	(0.005)	(0.005456)
Constant	0.035***	0.106***	0.103**
	(0.011)	(0.015)	(0.0483)
Observation	10,356	10,356	10,356
R^2	0.005	0.005	0.005
F-statistics	5.150	16.48	12.36

Note: Robust standard error in parenthesis. */**/*** represent significance level 10/5/1%. Controls include age, gender, tenure, education, and log earnings in 2020.



Note: Binned scatter-plot comparing earnings observed in the administrative data and self-reported earnings in 2020.

Figure C.4: Self-reported and registry earnings: last year (2020)

C.8 Standard Moments of Subjective Earnings Risks

Figure C.8 reproduces Figure 10(b)-(d) in the paper using standard deviation, skewness, and kurtosis rather than the interdeciel range, Kelley's measure of skewness, and the Crow-Siddiqui measure of excess kurtosis. Figure C.8(a) and (b) are quite similar to Figure 10(b) and (c) in the paper. Figure C.8(c) shows a slightly different pattern than Figure 10(d) in the paper, as older workers exhibit less kurtosis according to the standard measure than according to the Crow Siddiqui measure of kurtosis.

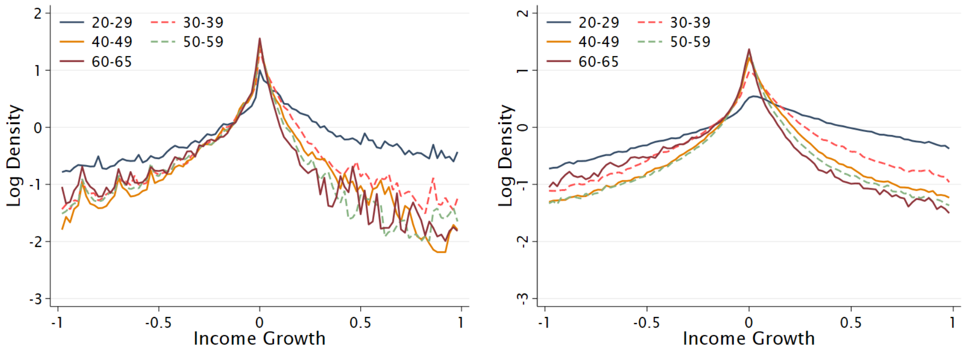
D Method for Simulating Holistic Earnings Risks

This section explains how we simulate the subjective holistic distribution of income growth, g_i, which is constructed by weighting together all the components entering each of the branches, B, for individual i as described in the below equation.

$$\hat{g}_i = (1 - \mathbb{1}[Q_i = 1] - \mathbb{1}[L_i = 1])\hat{f}_i^S + \mathbb{1}[Q_i = 1](1 - \hat{n}_i^Q)\hat{f}_i^Q + \mathbb{1}[L_i = 1](1 - \hat{n}_i^L)\hat{f}_i^L$$

Each simulation produces a value for $\hat{g}_i$. By simulating many times for each respondent, we get a complete empirical probability distribution over one year ahead of subjective holistic income growth. We now explain the details of the simulation. It proceeds in four steps:

1. Simulating job transitions, $\mathbb{1}[Q_i = 1]$, $\mathbb{1}[L_i = 1]$: We simulate 20,000 job transition



(a) Survey, h_S^X (b) 2019 Registry, h_A^X

Note: Panel (a) plots log density for the pooled distribution of expected holistic earnings growth rates[from the survey, h_S^X, where X indicates partitions by age groups. (b) plots the distribution of annual earnings growth from 2018 to 2019 as observed in the administrative data for the full population, h_A^X. For constructing the distribution of earnings growth in the administrative data we dropped observations where the level of annual earnings is less than 24,000 DKK in 2018. Survey results are weighted using population weights.

Figure C.5: Pooled earnings risk and registry earnings risk (data from 2019)

events for each individual based on the stated job transition probabilities (stay, laid off, and quit) in 2021.

2. Simulating time spent out of work, $\hat{n}_i^Q$, $\hat{n}_i^L$: For the job transition events, quit and lay-off from step 1, we simulate time spent out of work. This happens in two steps:

    1. We first simulate the timing (month) of separation during 2021. We assume that the separation happens on day 1 of the month. To simulate the month of separation, we first recover the monthly density of the job transition. Let P_l and P_q be the reported probabilities of being laid off and quitting at some point in 2021. Based on P_l and P_q we can recover p_l and p_q using two simultaneous equations. It has a geometric feature that captures the time flow in one year.

$$P_l = \underbrace{p_l}_{\text{Jan}} + \underbrace{(1-p_l)p_l}_{\text{Feb}} + \underbrace{(1-p_l)^2 p_l}_{\text{Mar}} + \ldots + \underbrace{(1-p_l)^{11} p_l}_{\text{Dec}} \tag{1}$$

$$P_q = \underbrace{p_q}_{\text{Jan}} + \underbrace{(1-p_q)p_q}_{\text{Feb}} + \underbrace{(1-p_q)^2 p_q}_{\text{Mar}} + \ldots + \underbrace{(1-p_q)^{11} p_q}_{\text{Dec}} \tag{2}$$

    We then construct a distribution of job transitions across months and based on this distribution we simulate a month in which the job separation event occur.

    2. Then we simulate the timing of reemployment. We do this using the stated

Note: The figure plots the distribution (log density) of annual earnings growth for 2020 as observed in the administrative data for the survey participants.

Figure C.6: Registry earnings risk (survey respondents only)

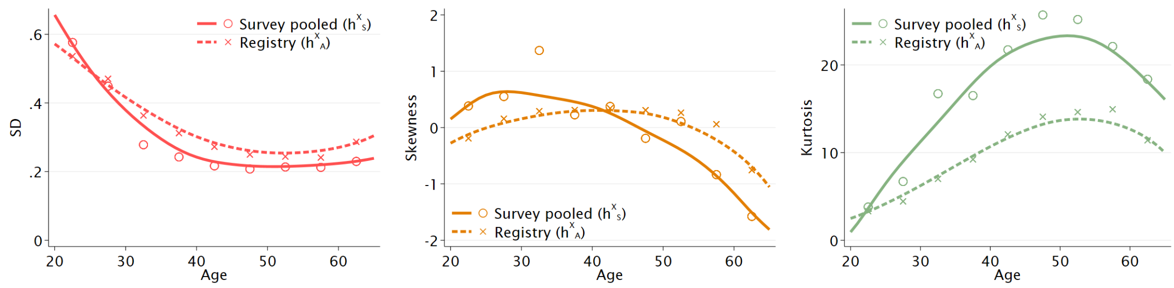
probabilities of being reemployed after 1, 3, and 12 months after the job separation. To do this, we linearly interpolate the probability of reemployment over months and then construct the monthly reemployment distribution. Using this monthly distribution of reemployment, we simulate the duration of the intermediate job search period after the separation. We assume that the reemployment happens at the beginning of the month.

3. Simulating conditional earnings distributions, $\hat{f}_i^S$, $\hat{f}_i^Q$, $\hat{f}_i^L$
   For each given simulated event we draw an income realization from the relevant conditional earnings distribution $\hat{f}_i^B$, $B = \{S, Q, L\}$

4. Weight together the components from steps 1-3

The key assumptions involved in the simulation procedure concerns step 2. Here we assume that uncertainty about time spent out of work is resolved at the beginning of the month. There are three different parts to income in 2021 if a job transition occurs. The first part is income **before** the separation. We assume that workers get a proportional amount of income from the staying branch income.[1] To implement this, we draw earnings from the stay earnings distribution and normalize the annual earnings by the duration of this spell. The second part is **job search period** before the reemployment. We assume zero earned income in this search period. Finally, the third part is **after** reemployment. We draw an annual earnings from their conditional earnings distribution

[1]If they have zero stay probability, we impute their earned income in the last year. However, 98.5% of workers have a positive probability of staying with their current employer.



(a) Standard Deviation (b) Skewness (c) Kurtosis

Note: The figure shows the average value of the 1st to 4th moments of the pooled earnings distribution by age in the survey, h_S^X, and in the administrative data, h_A^X. "o" and "x" represent the empirical mean across 5-years age bins for the survey and the administrative data for 2020, respectively. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights. The figure corresponds to Figure 6 in the paper, but uses standard measures for calculating variance, skewness and kurtosis as opposed to Figure 6, which is based on robust quantile based measures.

Figure C.7: Life cycle patterns in pooled risks moments

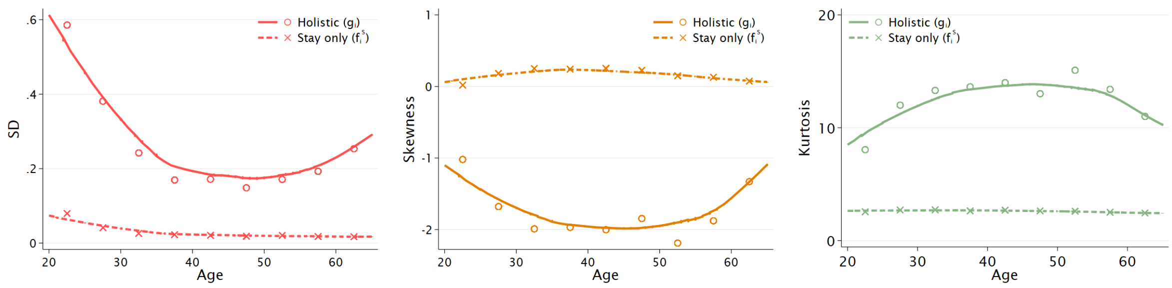
after reemployment in the branch. We adjust the annual earnings from this new employer by the number of months in the new job 2021.

Figure D.1 shows an example of how we simulate the earnings in 2021 following lay-off. When a lay-off occurs in step 1 then we draw a month for the job separation event after solving the equations (1)-(2). In the example, March is selected. In this case, the individual will receive a monthly wage for January and February, and we pick that from the stay branch's earnings distribution f_i^S. For the purpose of illustration, assume that \$ 120 is randomly drawn from f_i^S. We normalize this by multiplying $\frac{2}{12}$ to reflect the fraction of the year. In the example, July is chosen as the reemployment month as a result of simulating the reemployment distribution. Therefore, from March to June this individual has 0 earnings. Lastly, from July to December, the individual will receive a salary from the new employer and we, therefore, draw an earnings realization from her conditional earnings distribution, f_i^L, in this case \$100. We normalize this by $\frac{6}{12}$ to reflect the fraction of the year spent in the new job. Therefore, in this simulation, the aggregated earned income in 2021 is \$70.

E Moments of the Pooled Distribution

E.1 Age Group and Current Earnings Variations

In this section, we report figures corresponding to Figure 8 in the paper for all the groups defined in section 4.1. We note again that we divide the sample into two earnings groups,



(a) Standard Deviation (b) Skewness (c) Kurtosis

Note: The figure shows the average value of the 2nd to 4th standard moments over the life cycle of holistic earnings risk, g_i, and risk conditional on staying, f_i^S. "o" and "x" represent the empirical mean across 5-years age bins. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights.

Figure C.8: Moments of holistic earnings risk, g_i, vs. risk conditional on staying, f_i^S, over the life cycle: using standard moments

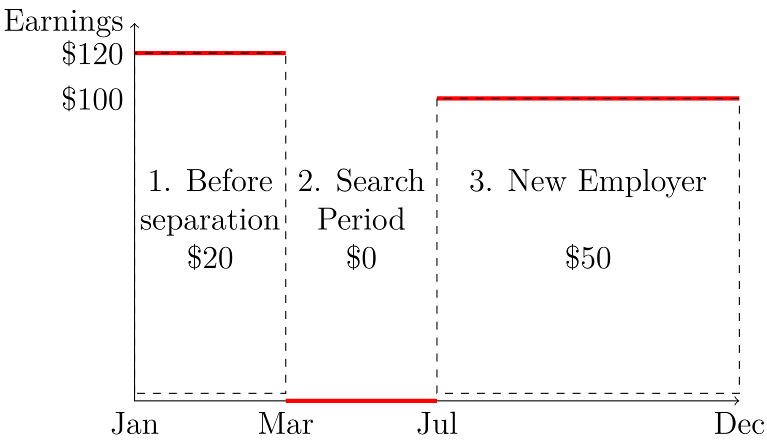
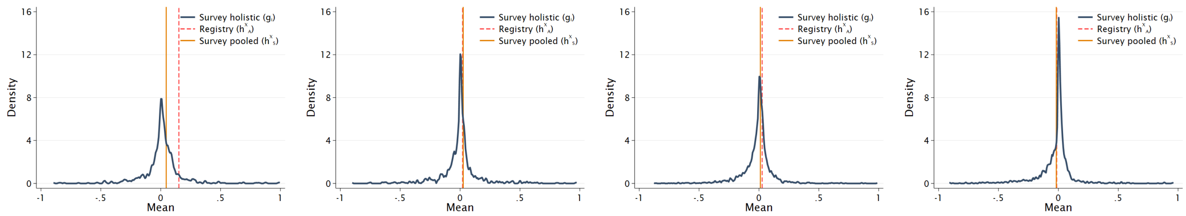


Figure D.1: Holistic earnings risk simulation

above-median (high) and below-median (low), and three broad age groups (20-34, 35-49, 50-65). Figure E.1 shows the results. The number on the label shows their age group and High/Low means their earnings level in the age group. Across all age groups and earnings levels, we find a pattern consistent with the pattern presented in Figure 8 in the paper. We confirm that the administrative based measure, h_X^A, and survey based measure, h_X^S, are well aligned. On the other hand, the subjective measure, g_i, is heterogeneous while being centered around 0.

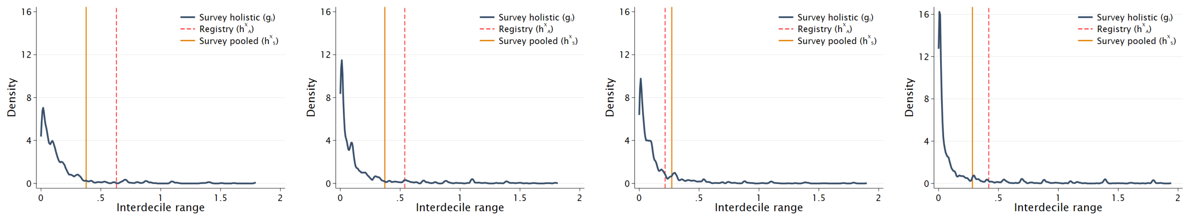
Figure E.2 show the distribution of the interdecile range, $p_{90} - p_{10}$, across groups. We again find that the subjective g_i is systematically lower compared to the registry and survey pooled moments.



(a) 20-34:*High* (b) 35-49:*Low* (c) 35-49:*High* (d) 50-65:*Low*

Note: Estimates of the mean for h_S^X, h_A^X, and the distribution of g_i for four subgroups in the data.

Figure E.1: Mean of h_S^X and h_A^X, and the distributions of individual means of g_i for four selected subgroups



(a) 20-34:*High* (b) 35-49:*Low* (c) 35-49:*High* (d) 50-65:*Low*

Note: Estimates of the interdecile range for h_S^X, h_A^X, and the distribution of g_i for four subgroups in the data.

Figure E.2: Interdecile range of h_S^X and h_A^X, and the distributions of individual interdecile range of g_i for four selected subgroups

E.2 Derivations

Referring back to equation (4) in the main text, let $m^{(k)}$ be the k^{th} moment of h_S^X and let $m_i^{(k)}$ be the k^{th} moment of g_i, then each moment is of a mixture distribution of N equally weighted distributions is given by the convex combination

$$m^{(k)} = \frac{1}{N}\sum_{i}^{N} m_i^{(k)} \tag{3}$$

and the variance (the centralized second moment) is given by

$$
\begin{aligned}
\operatorname{Var}(h) &= m^{(2)} - \left(m^{(1)}\right)^2 \\
&= \frac{1}{N}\sum_i^N m_i^{(2)} - \left(\frac{1}{N}\sum_i^N \mu_i\right)^2 \\
&= \frac{1}{N}\sum_i^N \left(\sigma_i^2 + \mu_i^2\right) - \left(\frac{1}{N}\sum_i^N \mu_i\right)^2 \\
&= \frac{1}{N}\sum_i^N \sigma_i^2 + \frac{1}{N}\sum_i^N \mu_i^2 - \left(\frac{1}{N}\sum_i^N \mu_i\right)^2
\end{aligned} \tag{4}
$$

The sum of the last two terms is always non-negative by Jensen's Inequality and increasing in dispersion in the μ_i's. This means that differences in the individual μ_i's increases the variance of h.

The skewness (the centralized third moment) is given by

$$
\begin{aligned}
\operatorname{Skew}(h) &= m^{(3)} - 3m^{(1)}m^{(2)} + 2\left(m^{(1)}\right)^3 \\
&= \frac{1}{N}\sum_i^N m_i^{(3)} - 3\frac{1}{N}\sum_i^N m_i^{(1)}\frac{1}{N}\sum_i^N m_i^{(2)} + 2\left(\frac{1}{N}\sum_i^N m_i^{(1)}\right)^3 \\
&= \frac{1}{N}\sum_i^N \left(\gamma_i^3 + 3\sigma_i^2\mu_i + 3\mu_i^3\right) - \frac{3}{N^2}\sum_i^N \mu_i \sum_i^N \left(\sigma_i^2 + \mu_i^2\right) + \frac{2}{N^3}\left(\sum_i^N \mu_i\right)^3
\end{aligned} \tag{5}
$$

where γ_i^3 is the skewness of g_i and we use equation (3) to convert $m^{(k)}$ to $m_i^{(k)}$ that are given by

$$
\begin{aligned}
m_i^{(1)} &= \mu_i \\
m_i^{(2)} &= \sigma_i^2 + \mu_i^2 \\
m_i^{(3)} &= \gamma_i^3 + 3\mu_i\left(\sigma_i^2 + \mu_i^2\right) - 2\mu_i^3 \\
&= \gamma_i^3 + 3\sigma_i^2\mu_i + 3\mu_i^3
\end{aligned} \tag{6}
$$

From Equation (5) we see that Skew(h) is the average of the individual skewness (γ_i^3, first term of first sum) in addition to an ambiguous dependence on the individual means and variances.

Kurtosis (the centralized fourth moment) is given by

$$\begin{aligned}
\text{Kurt}(h) &= m^{(4)} - 4m^{(1)}m^{(3)} + 6\left(m^{(1)}\right)^2 m^{(2)} - 3\left(m^{(1)}\right)^4 \\
&= \frac{1}{N}\sum_i^N m_i^{(4)} - 4\frac{1}{N}\sum_i^N m_i^{(1)} \frac{1}{N}\sum_i^N m_i^{(3)} + \\
&\quad 6\left(\frac{1}{N}\sum_i^N m_i^{(1)}\right)^2 \frac{1}{N}\sum_i^N m_i^{(2)} + 3\left(\frac{1}{N}\sum_i^N m_i^{(1)}\right)^4 \\
&= \frac{1}{N}\sum_i^N \left(\kappa_i^4 + 4\gamma_i^3\mu_i + 6\sigma_i^2\mu_i^2 + \mu_i^4\right) - \frac{4}{N^2}\sum_i^N \left(\gamma_i^3 + 2\sigma_i^2\mu_i + \mu_i^3\right) + \\
&\quad \frac{6}{N^3}\left(\sum_i^N \mu_i\right)^2 \left(\sum_i^N \sigma_i^2 + \mu_i^2\right) + \frac{3}{N^4}\sum_i^N (\mu_i)^4
\end{aligned} \tag{7}$$

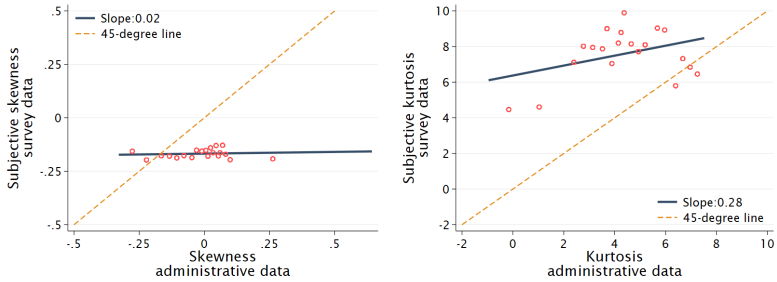
where κ_i^4 is the kurtosis of g_i and in addition to equation (6) we use that

$$\begin{aligned}
m_i^{(4)} &= \kappa_i^4 + 4m_i^{(1)}m_i^{(3)} - 6\left(m_i^{(1)}\right)^2 m_i^{(2)} + 3\left(m_i^{(1)}\right)^4 \\
&= \kappa_i^4 + 4\mu_i\left(\gamma_i^3 + 3\mu_i\left(\sigma_i^2 + \mu_i^2\right) - 2\mu_i^3\right) - 6\mu_i^2\left(\sigma_i^2 + \mu_i^2\right) + 3\mu_i^4 \\
&= \kappa_i^4 + 4\gamma_i^3\mu_i + 6\sigma_i^2\mu_i^2 + \mu_i^4
\end{aligned} \tag{8}$$

From Equation (7) we see that Kurt(h) is the average of the individual kurtosis (κ_i^4, first term of first sum) in addition to an ambiguous dependence on the individual means, variances and skewnesses.

E.3 Subjective and Registry Earnings Risk, Skewness and Kurtosis

Figure E.3 (a) shows the relationship between skewness of the h_A^X, i.e., skewness inferred from the administrative data in cell X, and average skewness of the subjective holistic distributions g_i. As shown above, the relationship between these two measures is ambiguous. In practice the two measures turn out to be unrelated. The correlation is -0.05 and insignificant at the 10% level of significance. Panel (b) shows the relationship between kurtosis of h_A^X, i.e., kurtosis inferred from the administrative data in cell X and average kurtosis of the subjective holistic distributions g_i in the same X cells. These two measures turn out to be weakly related with an estimated slope parameter of 0.11 which is not significant at the 10 % level. This is consistent with the theoretical prediction derived in the previous subsection showing that the relationship between the two measures is ambiguous.



(a) Skewness (b) Kurtosis

Note: The figure compares average skewness and kurtosis of subjective holistic income expectations, g_i, to those calculated from administrative data. Both are calculated within 300 cells divided by age group and within age group earnings percentiles, h_A^X, i.e, the same partition applied in the construction of Figure 9 in the paper. The panels show binned scatterplots (red circles) where the bins represent vigintiles of the interdecile range calculated from the administrative data. A regression line based on the 300 data points is overlaid.

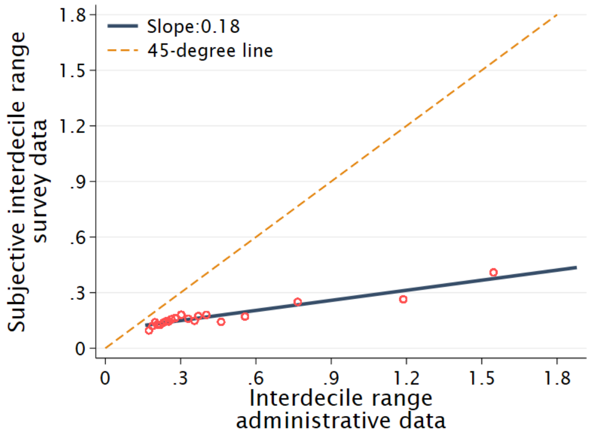
Figure E.3: Comparing skewness and kurtosis calculated from subjective expectations and from administrative data

E.4 Subjective and Registry Earnings Risk, Finer X-cells

In this subsection, we construct registry based earnings risk using a more detailed partition based on demographics. We partition the administrative data based on age, gender, education, and earnings deciles and calculate registry based moments within each of these cells. In total, we have 1,800 cells, i.e., a substantially more detailed stratification than the one applied in the paper which divides the data into 300 cells. Based on this we do a similar exercise as in Figure 9 of the main text. Figure E.4 shows the result. The pattern is very similar to the pattern shown in Figure 9 in the paper.

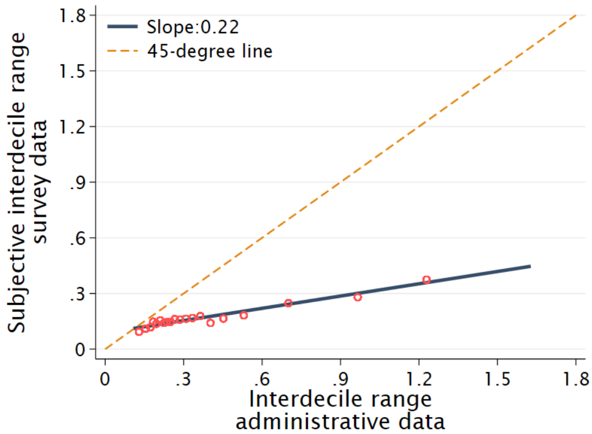
E.5 Subjective and Registry Earnings Risk, Heterogeneous Income Profiles (HIP)

The literature examining heterogeneous income processes (HIP) assumes that individual-level heterogeneity in earnings growth is present. To address this issue we calculate average individual level earnings growth over the last 5 years (2016-2020). We then divide the sample into cells by age, gender, education, current earnings quintiles, and last 5 years' average earnings growth quintiles, resulting in 4,500 cells over which we conduct an exercise similar to the one presented in Figure 9 in the paper. Again, the results are practically identical to the results presented in Figure E.5.



Note: The figure compares average interdecile range, $p_{90}-p_{10}$, of subjective holistic earnings expectations, g_i, to interdecile range calculated from administrative data. Both are calculated within 1,800 cells divided by age, gender, education, and earnings deciles, h_A^X. The panel shows a binned scatterplot (red circles) where the bins represent vigintiles of the interdecile range calculated from the administrative data. A regression line based on the 1,800 data points is overlaid.

Figure E.4: Comparing interdecile ranges calculated from subjective expectations and from administrative data: finer X-cells



Note: The figure compares average interdecile range, $p_{90}-p_{10}$, of subjective holistic earnings expectations, g_i, to interdecile range calculated from administrative data. Both are calculated within 4,500 cells divided by age, gender, education, current earnings quintiles, and last 5 years' average earnings growth quintiles, h_A^X. The panel shows a binned scatterplot (red circles) where the bins represent vigintiles of the interdecile range calculated from the administrative data. A regression line based on the 4,500 data points is overlaid.

Figure E.5: Comparing interdecile ranges calculated from subjective expectations and from administrative data: HIP

F Search Model: Details and Calibration

F.1 Value Functions

Recall that ψ denotes the aggregate state of the economy: $\psi = (n(y,t), u(y,t), e(z,y,t), \gamma)$. $n(y,t)$ denotes the measure of workers employed in matches of unknown quality with experience y and age t. $u(y,t)$ denotes the measure of unemployed workers with experience y and age t. $e(z,y,t)$ is the measure of employed workers with match quality z, experience y, and age t. γ is the measure of newly-born workers. The value function for an unemployed worker with experience y and age t is as follows:

$$U_t(y,\psi) = b + \beta \mathbb{E}_{\psi'} \left[U_{t+1}(y,\psi') + \lambda_u \max_x \left\{ p(\theta_{t+1}(x,y,\psi')) \left(x - U_{t+1}(y,\psi') \right) \right\} \right] \tag{9}$$

This worker earns b as home production. If they do not get the opportunity to search, with probability $1-\lambda_u$, their continuation value is $U_{t+1}(y,\psi')$. With probability λ_u, they do get the opportunity to search. In this case, they choose the submarket x they want to search in that maximizes the probability that they get a job offer there, $p(\theta_{t+1}(x,y,\psi'))$, times the value they get from that job, $(x - U_{t+1}(y,\psi'))$.

Next is the joint value function for a worker and a firm (which sums up the worker and firm value functions) in a match with known quality z:

$$\begin{aligned} V_t(z,y,\psi) &= zg(y) \\ &\quad + \beta \mathbb{E}_{\psi'} \left[\max_{d \in [\delta,1]} dU_{t+1}(y+1,\psi') + (1-d) \left(\mathbb{E}_{z'} V_{t+1}(z',y+1,\psi') + \lambda_e S_{t+1}(z,y+1,\psi') \right) \right] \end{aligned} \tag{10}$$

where

$$\begin{aligned} S_{t+1}(z,y+1,\psi') &= \max_x \left\{ p(\theta_{t+1}(x,y+1,\psi')) \left(x - \mathbb{E}_{z'} V_{t+1}(z',y+1,\psi') \right) \right\} \\ \mathbb{E}_{z'} V_{t+1}(z',y+1,\psi') &= \eta V_{t+1}(z_0,y+1,\psi') + (1-\eta) V_{t+1}(z,y+1,\psi') \end{aligned}$$

The flow value is just the match output (the firm's profit is $zg(y) - w$, but the worker earns w). Next period, the worker gains 1 unit of experience y, but the worker and firm separate with probability d, which is an optimally chosen policy (discussed further in the next section). If they separate, the worker goes to unemployment and gets value $U_{t+1}(y+1,\psi')$. Otherwise, with probability λ_e, the worker gets the opportunity to search on the job, with value encompassed in $S_{t+1}(z, y+1, \psi')$. In this case, the worker (like the unemployed worker above) chooses the optimal submarket x to search in which maximizes the probability they find a job there times the additional value they get from the new job relative to their old one. If the worker's search is unsuccessful or the worker does not get the opportunity to search, they get $\mathbb{E}_{z'} V_{t+1}(z', y+1, \psi')$ and stay with their existing job. In this case, the match quality may be reset to z_0. This happens with probability η and the worker becomes a match of unknown quality with value $V_{t+1}(z_0, y+1, \psi')$. Otherwise, with probability $1-\eta$, the match continues as is, $V_{t+1}(z, y+1, \psi')$.

Lastly, the value function for a worker with unknown match quality is as follows:

$$
\begin{aligned}
V_t(z_0, y, \psi) = \alpha \sum_z V_t(z, y, \psi) f(z) + (1-\alpha) \sum_z z g(y) f(z) \\
+ \beta(1-\alpha)\mathbb{E}_{\psi'} \left[\max_{d \in [\delta, 1]} dU_{t+1}(y+1, \psi') + (1-d)\left[V_{t+1}(z_0, y+1, \psi') + \lambda_e S_{t+1}(z, y+1, \psi')\right]\right]
\end{aligned}
\tag{11}
$$

With probability α, the match quality is discovered, drawn from $f(z)$, and the match immediately gets the value of being of quality z, $V_t(z, y, \psi)$. Otherwise, the output is the expected productivity of the match. Next period, the set of possible events is the same as the analogous branch for the match of known quality as above.

F.2 Equilibrium

To characterize the optimal policies, note first that the following will hold in each submarket:

$$
k \geq q(\theta_t(x, y, \psi))[V_t(z_0, y, \psi) - x] \tag{12}
$$

The left-hand side is the cost of opening a vacancy in any given submarket and the right-hand side is the expected benefit to the firm of opening a vacancy in submarket x. This is the probability that a firm will meet a worker in that submarket $(q(\theta_t(x, y, \psi)))$ times the expected value of employing a new worker: the value of a match of unknown quality minus the value the firm delivers to the worker, x. If the condition holds with equality, the submarket x will be open (i.e., have searchers and vacancies). Otherwise, if the cost exceeds the benefit the submarket will be closed.

When workers search for jobs, their preferences are given by:

$$p(\theta)(x - v)$$

This says that workers prefer to search for jobs that are easier to find ($p(\theta)$ and that offer lifetime values x above what they are getting in their current employment state, v. In equilibrium, there will be a trade-off between these two job attributes. Workers in better employment states will prefer to only search for jobs that are both harder to get but that offer higher values. Combining this with the complementary slackness condition (12) leads to the following search problem for the worker:

$$\max_{\theta \geq 0} p(\theta) \left[V_t(z_0, y, \psi) - \mathbb{E}_{z'} V_t(z', y, \psi) \right] - k\theta$$

The solution characterizes the worker's optimal choice of submarket when they are searching while already employed. The expression says that workers choose their submarket to maximize the probability that they find a job times the additional value they get from taking that job, net of the cost of creating the vacancies. There is an analagous one for unemployed workers, where the outside option is the value of unemployment:

$$\max_{\theta \geq 0} p(\theta) \left[V_t(z_0, y, \psi) - U_t(y, \psi) \right] - k\theta$$

Turning to the separation policies, these are made on the basis of comparing the value of keeping the match with the value of breaking up. The match will separate with probability 1 if:

$$U_{t+1}(y+1, \psi') > (1 - \lambda_e)\mathbb{E}_{z'} V_{t+1}(z', y+1, \psi') + \lambda_e S_{t+1}(z, y+1, \psi')$$

The left-hand side is the value of unemployment. The right-hand side is the value of keeping the match which consists of the value of the match continuing to next period plus the value of search to the worker. There is an analogous expression for workers in matches of unknown quality.

If the inequality is reversed, the match only separates exogenously with probability δ. In sum, the separation policies can be described by thresholds for match quality, in which the match is destroyed if it is below it, and kept if it is above the threshold.

Everything is now in place to define the equilibrium.

Definition: A *Block Recursive Equilibrium* consists of:

- Value functions for the unemployed, employed in a match of known quality, and employed in a match of unknown quality: $U_t(y, \psi), V_t(z_0, y, \psi), V_t(z, y, \psi)$, respectively

- Policy functions that determine which submarkets the employed search in and the match quality thresholds for destruction of the match
- A market tightness $\theta_t(x, y, \psi)$ (ratio of vacancies to unemployment) for each submarket

such that

- The value, policy, and market tightness functions do not depend on the aggregate state ψ
- The market tightnesses satisfy (12)
- The policy functions solve (9), (10), and (11)

Menzio et al. (2016) show that there is a unique Block Recursive Equilibrium in this setup and the equilibrium is socially efficient.

F.3 Calibration Details

F.3.1 The Registry Data

We calibrate the model to data from a Danish administrative register called *eIncome*. These data cover the entire Danish population (around 6 million people) and are recorded monthly. See Kreiner et al. (2016) for more details. We limit our sample to workers between the ages of 20 and 65.

We use four different data sets. The first is the employer-employee matched data. For each worker, we get their total labor earnings and firm identifiers for each job they hold within a month. We define a person's main job in each month as the job that provided the highest amount of earnings.

Second, we use the database of unemployment insurance claims – 80% of Danish workers are covered by the unemployment insurance fund, so for the vast majority of workers we are able to see if they made any unemployment claims each month.

The third database we use covers retirement and disability benefits. Using these data, we can also identify who is out of the labor force or retiring and how much they claimed each month.

Finally, we use tax claim data to identify the self-employed (they are not observed in the matched employer-employee data). These claims provide a self-reported level of self-employed earnings at the annual level. We exclude workers with self-employment income amounting to more than 30,000 DKK per year in order to make sure that we are considering workers who are employed. The proportion of the sample who reported more than 30,000 DKK in self-employment income is around 4% of the population in 2020.

A key step for generating the moments we need for the calibration is to classify each person's labor market status each month. To do this, we find the highest value among the earnings from the main job, UI benefits, retirement benefits, and disability benefits. If the highest value is labor earnings, we classify the worker as employed. If it is UI claims, we classify the worker as unemployed. If it is retirement or disability benefits, we classify the worker as permanently out of work. Finally, if the highest value is less than 2,000 DKK, we classify the worker as temporarily out of work.

These classifications now allow us to locate the following types of job transitions. EE transitions are when the worker is employed at different employers between months $t-1$ and t. UE transitions occur when the worker is unemployed in $t-1$ and employed in month t. EU transitions are when the worker is employed in month $t-1$ and unemployed in month t.

F.3.2 Calibration Strategy and Targeted Moments

We set the discount factor, β, externally to correspond to an annual interest rate of 4%. The rest of the parameters, described here, are calibrated internally to match moments in the data.[2] b is the amount of home production produced by the unemployed. $f(z)$ is the distribution of match quality. Following Menzio et al. (2016) it is parameterized as a Weibull distribution with mean 1, scale σ, and shape ϕ, approximated on a 100-point grid. This distribution is flexible enough to accommodate many possible shapes. α is the probability that the match quality is discovered. η is the probability that the match quality is reset. $g(y)$ is the function that determines how experience is mapped to output. Again, following Menzio et al. (2016), it is parameterized as $g(y) = (1-\rho_1) + \rho_1 y^{\rho_2}$. ρ_1 determines the level and ρ_2 determines the curvature. The scalar parameters of the matching process are the vacancy cost k, the search probability of unemployed λ_u (normalized to 1 without loss of generality), the search probability of employed λ_e, and the exogenous firing probability δ. In addition, we parameterize the matching function as $p(\theta) = \min\{\theta^{1/2}, 1\}$.[3]

Next, we briefly describe the moments we target and how they are identified by the

[2] We use an adaptive grid search method to arrive at the set of parameter values that best match the data moments. To evaluate the fit, we use a minimum distance metric, which is the sum of squared differences of the model's vs. the data's moments, where each moment is given equal weight.

[3] This is set up to automatically give an elasticity of the job finding probability with respect to the market tightness of 0.5, roughly the value estimated by Menzio and Shi (2011). The minimum ensures that the job-finding probability never goes above 1.

parameters of the model.[4] We target the overall monthly UE, EU, and EE rates. The UE transition rate is identified by the vacancy cost k because it determines how many vacancies will open and therefore the job-finding probability of the unemployed. The δ informs the overall EU rate because this separation probability applies equally to all matches. λ_e impacts the overall EE rate because the more often workers get the chance to search, the more EE transitions that will take place. We use b to target a ratio of home production to earnings of 0.7.[5]

We use the tenure profiles of the EE and EU rates to parameterize the parts of the model that pertain to match quality. When the quality is known, the likelihood that a match reaches a particular tenure depends on the search policies of workers: workers with lower quality z will search for new jobs that are easier to get. The more low z's there are in the distribution, the more short tenure jobs that end in another employment spell there will be. Similarly, the more low z's the more that will be destroyed upon discovering the quality, and the more short tenure jobs that end in an unemployment spell there will be. The rates at which match quality is discovered or reset also impact these tenure profiles because they trigger changes in match quality which in turn determines how likely they are to go through a transition.

Finally, we use the average earnings profile as a function of age to inform the parameters of the human capital accumulation function. Since we assume that earnings are a constant fraction of output,[6] the $g(y)$ function will be strongly tied to earnings. The functional form that we choose is flexible enough to accommodate the concave shape of the earnings profile.

Figure F.1 shows the model fit for the key targeted moments: overall job transition rates, job transition rates as a function of tenure, and earnings as a function of age. All the model-generated moments match their empirical counterparts well. Table F.1 summarizes our calibrated parameter values.

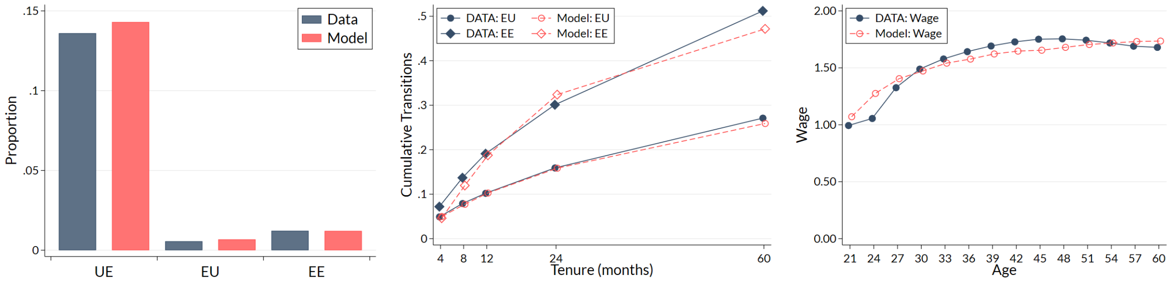
F.3.3 Untargeted Moments

Next, we present some of the untargeted moments produced by the model. We focus on the ones that are closest to moments we are most interested from the survey and the registry: the life-cycle patterns of job transition rates and the distribution of annual

[4]In general, each parameter controls many moments but some moments are particularly informative about certain parameters.

[5]We get this value by calculating the ratio of average unemployment insurance to total earnings in the 2019 registry.

[6]Earnings are not pinned down in this model because there are many protocols that can deliver the required value to the worker. Thus, assumptions need to be made about how they are set. In practice, there is little impact of this choice on the implications of calibrated directed search models, so most go with the piece-rate assumption (see Menzio et al. (2016) and Gregory et al. (2021)).



(a) Job transition rates (b) Transition rates by tenure (c) Earnings by age

Figure F.1: Targeted moments

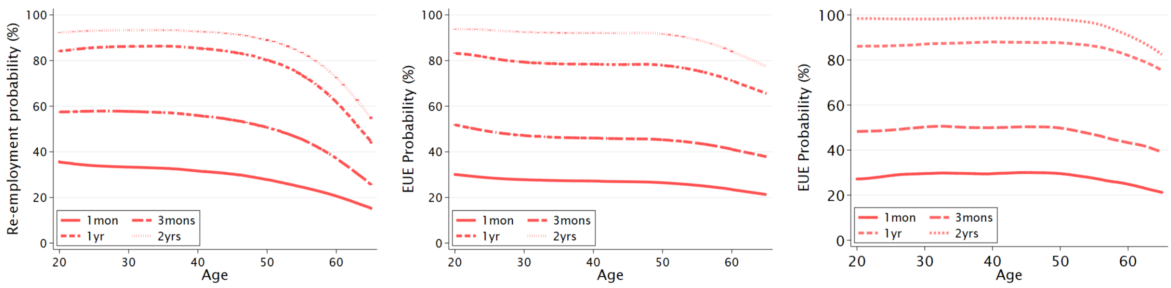
Table F.1: Calibrated parameters

	Description	Value
β	discount factor (externally set)	0.9967
b	home production	2.083
k	vacancy cost	54.067
λ_u	job search probability: unemployed	1
λ_e	job search probability: employed	0.991
δ	exogenous destruction probability	0.0007
σ	scale of match quality distribution	12.085
ϕ	shape of match quality distribution	2.704
α	probability of match quality discovery	0.124
η	probability of match quality resetting	0.094
ρ_1	human capital accumulation: level	3.407
ρ_2	human capital accumulation: curvature	0.090

earnings growth. As explained in the main part of the paper, we equate EEs in the model and register with quits in the survey and EUs in the model and register with layoffs in the survey. We discuss the age profile of the EE and EU rates in the main part of the paper, so we focus on other moments here.

Figure F.2 compares the patterns in time-to-reemployment after a layoff (in the survey) or the length of the U spell in an EUE transition (in the registry and model). Each line shows the probability of being reemployed within 1, 3, 12, or 24 months as a function of age. Again, the model matches the registry very well, despite not being targeted. The pattern in the survey is also quite close, except for those above age 55 or so.

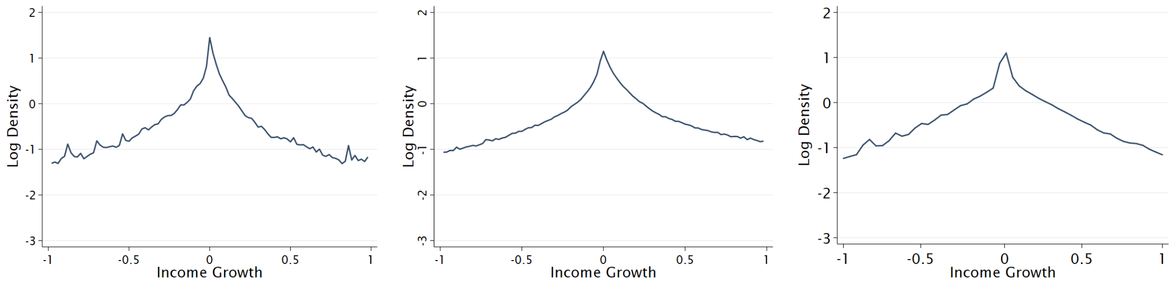
Lastly, Figure F.3 compares the densities of annual earnings growth in the survey, registry, and model. Again, the distribution generated by the model is similar to that of the registry.



(a) Survey (b) Registry (c) Model

Note: Population weights are used in Panel (a). The lines show local regression smoothed lines.

Figure F.2: Reemployment probabilities



(a) Survey (b) Registry (c) Model

Note: Population weights are used in Panel (a).

Figure F.3: Distributions of annual log earnings growth

References

Gregory, V., G. Menzio, and D. G. Wiczer (2021). The alpha beta gamma of the labor market. Technical report, National Bureau of Economic Research.

Kreiner, C. T., S. Leth-Petersen, and P. E. Skov (2016). Tax reforms and intertemporal shifting of wage income: Evidence from danish monthly payroll records. *American Economic Journal: Economic Policy 8*(3), 233–257.

Menzio, G. and S. Shi (2011). Efficient search on the job and the business cycle. *Journal of Political Economy 119*(3), 468–510.

Menzio, G., I. A. Telyukova, and L. Visschers (2016). Directed search over the life cycle. *Review of Economic Dynamics 19*, 38–62.