NBER WORKING PAPER SERIES

SUBJECTIVE EARNINGS RISK

Andrew Caplin
Victoria Gregory
Eungik Lee
Søren Leth-Petersen
Johan Sæverud

ABSTRACT

Earnings risk is central to economic analysis. While this risk is essentially subjective, it is typically inferred from administrative data. Following the lead of Dominitz and Manski, 1997, we introduce a survey instrument to measure subjective earnings risk. We pay particular attention to the expected impact of job transitions on earnings. A link with administrative data provides multiple credibility checks. It also shows subjective earnings risk to be far lower than its administratively-estimated counterpart. This divergence arises because expected earnings growth is heterogeneous, even within narrow demographic and earnings cells. We calibrate a life-cycle model of search and matching to administrative data and compare the model-implied expectations with our survey instrument. This calibration produces far higher estimates of individual earnings risk than do our subjective expectations, regardless of age, earnings, and whether or not workers switch jobs. This divergence highlights the need for survey-based measures of subjective earnings risk.

Andrew Caplin
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10012
and NBER
andrew.caplin@nyu.edu

Victoria Gregory
Federal Reserve Bank of St. Louis
1 Federal Reserve Bank Plaza
St. Louis, MO 63102
vgregory757@gmail.com

Eungik Lee
Department of Economics
New York University
19 W. 4th Street, 6th Floor
New York, NY 10012
lee.eungik@gmail.com

Søren Leth-Petersen
Department of Economics and CEBI
University of Copenhagen
Oster Farimagsgade 5
Building 26
DK-1353 Copenhagen K
Denmark
Soren.Leth-Petersen@econ.ku.dk

Johan Sæverud
Department of Economics and CEBI
University of Copenhagen
Osterfarimagsgade 5
Building 26
Copenhagen DK-1353
Denmark
js@econ.ku.dk

A data appendix is available at http://www.nber.org/data-appendix/w31019

1 Introduction

Earnings risk is central to economic analysis due to its impact on labor supply (Abowd and Card, 1987), job search (Low et al., 2010), consumption and savings decisions (Deaton et al., 1992), inequality (Gottschalk and Moffitt, 1994), etc. While this risk is essentially subjective, it is typically estimated from administrative data. As Dominitz and Manski (1997) noted, this is only as valid as are the underlying assumptions of homogeneity and full information rational expectations. Reflecting the importance of more accurately assessing earnings risk, they designed and implemented a survey instrument to measure subjective income risk. Their work started a field studying subjective probabilistic expectations in a range of domains as highlighted in two Econometric Society Presidential Addresses (Manski, 2004; Almås et al., 2023) and the recent *Handbook of Economic Expectations* (Bachmann et al., 2022).

In a recent paper that dives deeply into earnings risk using administrative data, Guvenen et al. (2021) characterize the distribution of earnings growth in the US. They document that higher-order moments, skewness and kurtosis, in addition to mean and variance, are important for describing the distribution of earnings growth in the population. Grouping observations, they further show how these moments vary with age and the level of earnings to help characterize labor market risks that workers face. Similar analyses have been implemented in many countries in the Global Repository of Income Dynamics (GRID) project, and findings are remarkably homogeneous across countries (Guvenen et al., 2022). Yet inferring earnings risk from administrative data comes with assumptions that are hard to test without subjective expectations data. Hence, the questions raised by Dominitz and Manski (1997) on how risk is inferred from administrative data are no less vital today than they were 25 years ago.

Although there is an existing branch of research focused on subjective earnings expectations (Pistaferri, 2001, 2003; Guiso et al., 2002; Dominitz, 1998, 2001), measuring it has proven challenging. One challenge relates to possible job transitions and time out of the labor force. Guvenen et al. (2021) document the prominent role of such job transitions for earnings risk, in particular for higher order moments. A second challenge relates to credibility. Almås et al. (2023) emphasize that it is essential to assess the credibility of subjective earnings expectations since they are neither standard behavioral data nor factual administrative data and as such relatively unfamiliar to economists. Given the limitations of the survey architectures in which they launched their pioneering instrument, Dominitz and Manski (1997) were in a position only to check consistency of

survey-measured beliefs about future income with basic principles of probability rather than with patterns in administratively-measured income.

In this paper we revisit the thesis of Dominitz and Manski (1997) that administratively estimated earnings risk may differ significantly from its subjective survey-estimated counterpart. One key distinction is that we have access to a richer and more comprehensive measurement infrastructure. A second is that we address the recent findings of Guvenen et al. (2021) on the importance of time out of the labor force by conditioning our expectations instrument on possible job transitions. To address their findings on the importance of higher order moments, the full probability distribution over next year's earnings is measured with and without job transitions. To address the credibility challenge, responses are linked to administrative data with third-party reported records of earnings and job transitions (Andersen and Leth-Petersen, 2021). We show last year's survey-reported earnings match closely with their administrative counterpart. Average survey-reported probabilities of switching jobs in the next year tightly match historical averages as does the average time between jobs. When we suitably aggregate survey-reported earnings variability to the population level, it replicates key patterns in the administrative data. Finally, we find a match between life cycle patterns of skewness and kurtosis in addition to mean and variance, so that subjective data mirror standard findings in administrative data.

Our data on survey-based subjective earnings risk pinpoint a major limitation of standard methods of inference from administrative data. Administratively-estimated earnings risk is many times higher than its survey-based counterpart. Figure 1 illustrates this pattern of overestimation when we divide the population into cells according to age and earnings, as in Guvenen et al. (2021). The figure shows a binned scatter plot of average survey-based subjective earnings risk against the corresponding levels of risk inferred from administrative data. The figure shows administratively-estimated earnings risk to be between two and six times higher than its survey-based counterpart. The main source of this difference is that, even within narrow sub-groups, there is significant variation in mean survey-measured subjective earnings growth. These differences in mean growth rates raise administratively-estimated earnings risk, as ex ante heterogeneity is erroneously assigned to differences in luck. In confirmation of this channel, the gap between subjective risk and its administrative counterpart is particularly high for groups with highly heterogeneous expected growth rates in earnings, such as younger workers.

As Dominitz and Manski (1997) noted, a key reason for gathering subjective earnings data is to discipline models of search. We follow up on their proposal by calibrating a
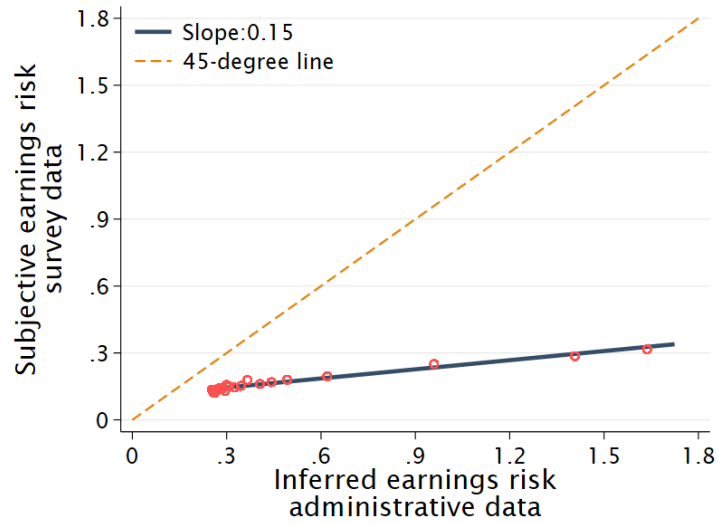
Note: The figure shows a binned scatter plot of subjective earnings risk (vertical axis) against earnings risk inferred from the administrative data (horizontal axis). We measure risk as the interdecile range, $p_{90} - p_{10}$, of the distribution of earnings growth rates. The administrative data have been partitioned into 300 cells defined by age groups (20-34, 35-49, 50-65) and earnings percentiles. Within each cell, the interdecile range is calculated from the administrative data, and the average of subjective interdecile ranges from the survey data is calculated. 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 300 data points is overlaid. The result is explained in detail in Section 4.

Figure 1: Comparing subjective and registry inferred earnings risk

canonical model of search over the life cycle to administrative data on job transitions, deriving the model-implied expectations, and comparing them to the subjective risks identified in our survey. We focus on the model of search over the life cycle by Menzio et al. (2016).[1] The model is an appealing benchmark for us because it emphasizes job mobility, which our survey points out is critical for earnings risk, and it has just the right amount of detail to endogenously generate variation in mobility over the life cycle.[2]

Standard practice in the literature is to calibrate models to administrative data on job transitions. Little is known about whether the *implied* expectations of workers in such calibrated models are actually consistent with the subjective expectations data. To learn more, we follow the standard practice of calibrating the model to administrative data. This allows us to back out the implied beliefs about earnings risk in the face of any job transition, setting up a direct comparison with our survey-measured subjective expectations. Specifically, we start from a sample of model-simulated workers drawn from the stationary equilibrium. Beliefs about the probability of job transitions derive from simulating each worker's states forward many times, and for the branch-specific beliefs, putting each worker on each branch and simulating forward from there. Imposing rational expectations as is standard in the literature, these paths represent the worker's beliefs about all of the outcomes that are possible over the next year.

As might be expected, the key point of contrast is that the model calibrated to administrative data produces far higher estimates of individual earnings risk than do our subjective expectations measures. This is true even when conditioning on job transitions. Whether workers stay in their current employment or make a job transition, they subjectively perceive earnings risk to be far lower than the model implies. Furthermore, while the model has the potential to generate variation in risk around job transitions as a function of earnings and age, the level and patterns of risk in the model-implied beliefs miss those of the subjective data. This finding stems from common features in search models: these include going all the way to the bottom of the job ladder after a separation,

---

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

[2] 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 on the life cycle of earnings risk. In such models, like the one we focus on in this paper, people search on the job and face the risk of job separation. As workers gain more experience, they gradually move into higher-paying jobs, which they tend to remain at longer compared to the lower-paying, more short-lived jobs they held at the start of their lives. The model features that generate these dynamics can rationalize life-cycle variation in earnings risk.

and the firm and the worker initially not knowing the productivity of the match. Our results highlight the value of using survey-based measures of subjective earnings risk in modeling labor market transitions.

Our paper builds on recent advances in our understanding of earnings dynamics and earnings risk measured in large administrative data (e.g., Guvenen et al., 2021).[3] These studies showed the importance of higher-order moments in the distribution of 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.[4] 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 Dominitz (1998, 2001), Pistaferri (2001, 2003), and Guiso et al. (2002). The branch of that literature on which we build relates to measurement of conditional expectations (e.g. Arcidiacono et al., 2020, Wiswall and Zafar, 2021). Our work is also related to recent literature using subjective expectations data to inform and discipline structural models of labor market dynamics (Conlon et al., 2018; Bick et al., 2021; Mueller et al., 2021; Faberman et al., 2022; Jäger et al., 2022).

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 the administrative data and compares beliefs implied by the model to subjective expectations from the survey. 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.[5] We then introduce the *Copenhagen Life Panel* in which it was implemented

---

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

[4] The importance of job transitions for earnings is also revealed in separate studies of layoffs and quits (Topel and Ward, 1992; Jacobson et al., 1993; Von Wachter et al., 2009).

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

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 2 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.

$$
\begin{array}{ccc}
\text{JOB} & \text{TIME} & \text{EARNINGS} \\
\text{TRANSITION} & \text{OUT} & \text{GROWTH RATES}
\end{array}
$$

STAY $\quad p_i^S \rule{10em}{0.4pt} f_i^S$

LAY-OFF $\quad p_i^L \rule{4em}{0.4pt} n_i^L \rule{4em}{0.4pt} f_i^L$

QUIT $\quad p_i^Q \rule{4em}{0.4pt} n_i^Q \rule{4em}{0.4pt} f_i^Q$
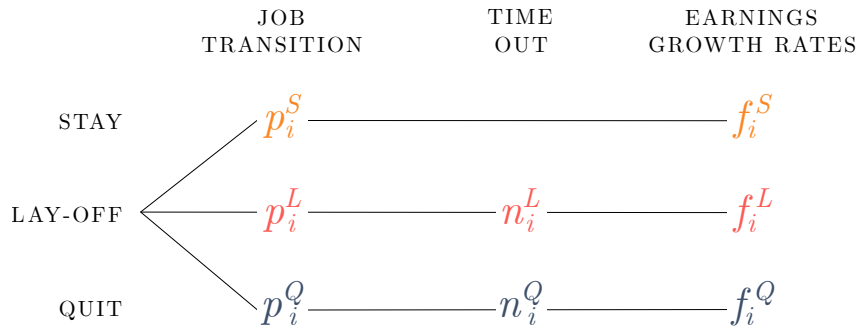
Note: The survey instrument consists of three branches, each representing a job transition (Stay, Layoff, Quit), and three domains for each branch: for each individual i we elicit job transition probabilities, p_i^B, time out of work, n_i^B, and distributions of conditional earnings growth rates, f_i^B, where $B \in \{S, L, Q\}$.

Figure 2: Survey instrument overview

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*

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.

- *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.[6] 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

---

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

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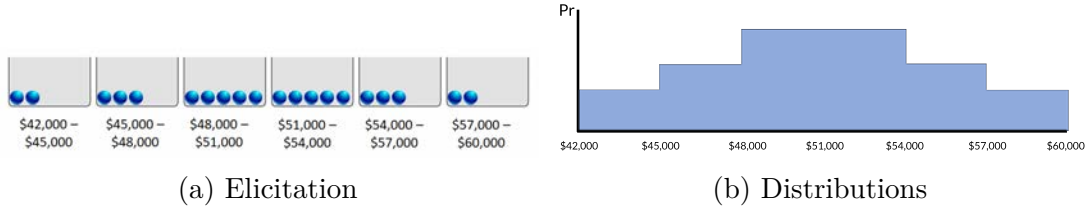
Note: Panel (a) shows the sample screen for the elicitation and Panel (b) shows how we interpret the distribution of the answer in Panel (a) as a mixture of uniform distributions.

Figure 3: Balls in bins

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 3(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 3(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 3(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.

In the survey we ask respondents about last year's earnings. 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* [7] (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

---

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

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.[8]

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 data are collected for the entire Danish population and they are longitudinal by nature.

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 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.[9]

2.3 Job Transitions

In Figure 4 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, $\bar{p}^L = 6\%$.

---

[8] 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).

[9] 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.
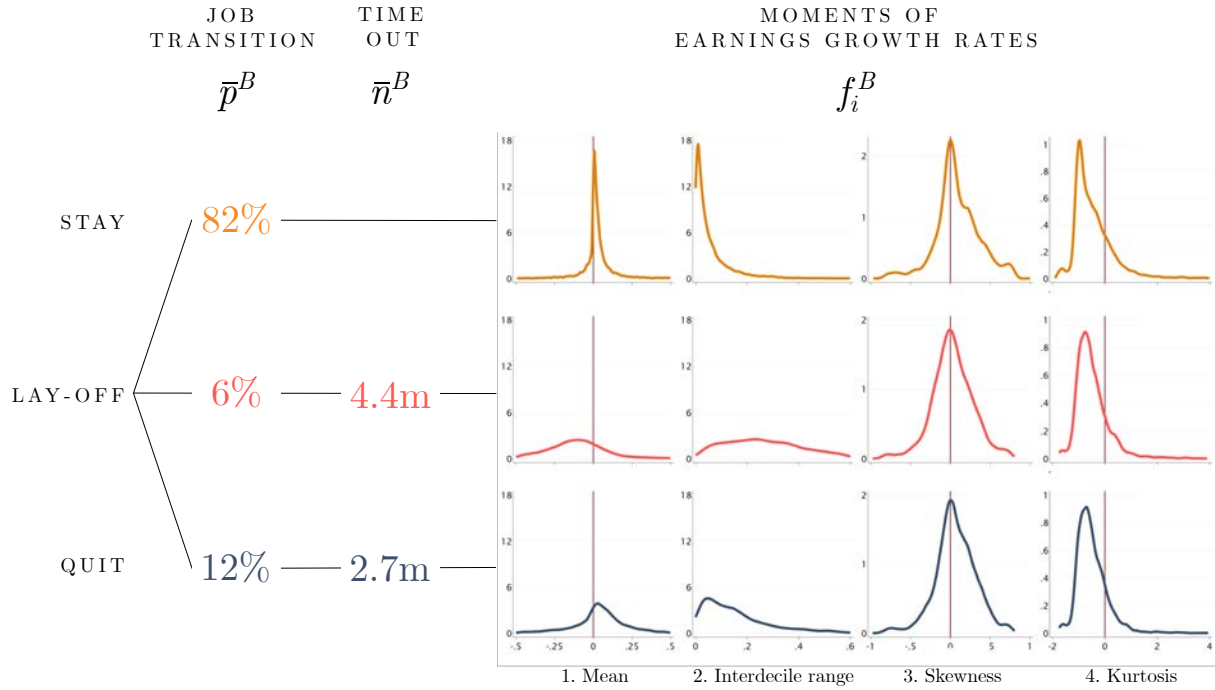
Note: The figure shows answers to the questions in the conditional survey instrument, where the rows correspond to the branches "Stay", "Layoff", and "Quit". The first column shows the average probabilities of each branch, $\bar{p}^B$. The second column shows the average of the expected reemployment 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. 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.

Figure 4: Overview of branch-by-branch survey responses

Moving to the right in Figure 4, 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 $(n_{i,1}^L, n_{i,3}^L, n_{i,12}^L, n_{i,24}^L)$ 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:

$$n_i^L = 2(n_{i,3}^L - n_{i,1}^L) + 7.5(n_{i,12}^L - n_{i,3}^L) + 12(100 - n_{i,12}^L)$$

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

The numbers under "Time out" in Figure 4 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.

The fact that many survey respondents expect to spend a short time out of work following 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. Suggestive evidence for this channel is found when 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 (Zeldes, 1989; 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. In Online Appendix C.1, we show the full results of this analysis.

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 3, 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, $\hat{f}_i^B$.[10]

The last four columns of Figure 4 show the cross sectional distribution of the first four moments of subjective earnings growth distributions.[11] 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 heterogeneity 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 uncertain people are about their earnings prospects in each branch. As might be expected, the results show that people tend to be most certain about their earnings growth in the stay branch and least certain 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 uncertainty.

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.

---

[10] 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).

[11] 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.

Lastly, we turn to the distribution of excess kurtosis. The final column of Figure 4 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.

Figure 4 provides an overview of the subjective distributions in the sample. Being laid off leads to the worst outcomes, on average, and respondents are most uncertain about 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 uncertainty 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.[12]

3 Comparing Survey and Administrative Data

Almås et al. (2023) emphasize the importance of establishing the credibility of new measures such as our subjective earnings expectations instrument to confirm that there are no first-order discrepancies between what the instrument is intended to measure and what it actually measures.

In our case the Danish research data infrastructure allows us to directly compare measures elicited in the survey with administrative data. 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 2020 currently available for research. 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

---

[12] In Online Appendix C.2 and C.3, we present a life cycle version of the graph and a comparable graph using standard measures of the moments.

use a standard battery of administrative data compiled by Statistics Denmark.

In comparing the survey data with administrative counterparts, we open 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 Registry. Our baseline comparison will be based on administrative data for 2020.[13] 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 two-thirds of this small loss had been recovered.

3.1 Job Transitions

In the survey we ask about the probability of a job separation and the associated time out of work before reemployment. Actual job transitions and time out of work following a job separation are directly observed in the administrative data. As a first-order check on the survey answers to these items, we compare these objects directly.

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 5, 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 2020 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

The comparison of time out of work following a job separation involves one fine point. While in the survey we ask about expectations concerning two separate types of job separations, quits and layoffs, in the administrative data we only observe whether a job

---

[13] Administrative data for 2021 has not yet been made available for research.

separation has occurred but not the reason for it. 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 2020, we consider time spent out of work following job separations that took place in 2019 such that we can follow periods out of work that extend into 2020. 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 5, 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 Simulating Earnings Risk from the Survey

We now compare the cross sectional distribution of expected earnings growth in the survey and the cross sectional distribution of earnings growth in the administrative data. The first characterizes the expected earnings growth variability in the population and the second characterizes the realized earnings growth variability in the population.

In order to arrive at a cross sectional distribution of expected earnings growth, we aggregate the conditional answers in two steps. First we aggregate the conditional answers for each respondent into one distribution that characterizes overall subjective earnings risk, which we denote subjective expected *holistic* earnings growth. This object summarizes the overall subjective probability distribution over future earnings taking into account all the different contingencies that we asked about. Second, we pool the distributions of subjective expected holistic earnings growth for all 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.

Step 1: Individual Measure To construct the subjective holistic expected earnings growth distribution we weight together each of the branches, $B = \{S, Q, L\}$, for individual

16

(a) Stay probability   (b) Time out of work
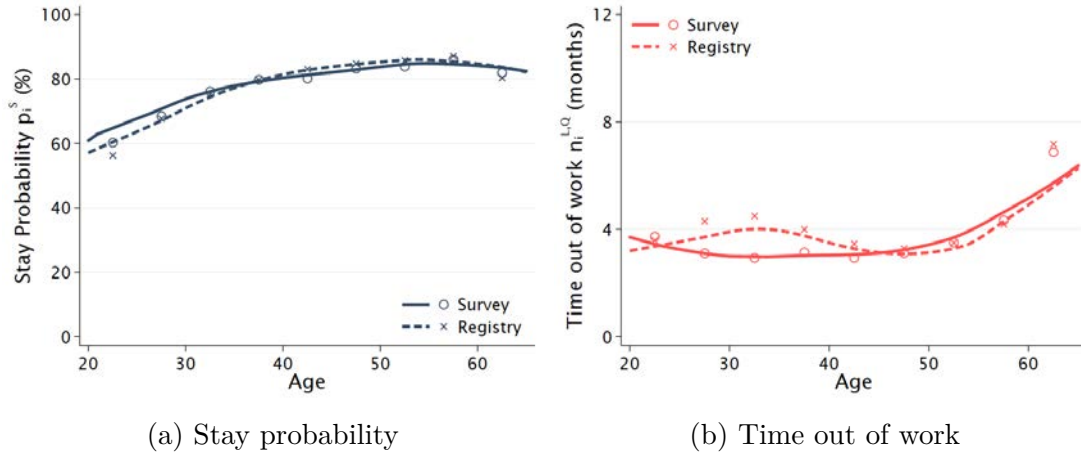
Note: The figure shows how two components from the survey (solid line) and the administrative data (dashed line) compare over the life-cycle. Panel (a): In the survey, we directly ask about the probability of staying with the same employer for the next 12 months. In the registry, we compute the proportion of workers who stayed from Dec 2019 to Dec 2020 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 2019 and follow time spent until reemployment occurs, possibly extending into 2020. 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.4 shows the corresponding figures for 2019.

Figure 5: Job separations and time out of work in the survey and in the administrative data

17

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.

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 $\hat{n}_i^B$,[14] which are then weighted together according to Equation (2). In practice, 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 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.[15] In the following we refer to g_i but in practice we use $\hat{g}_i$.

Step 2: Aggregate Measure   We combine the subjective holistic earnings growth distributions, g_i, for all N individuals in our survey into a pooled distribution. We pool within cells in which individuals share the same observable 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

---

[14] 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.

[15] 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.

(a) Survey, h_S^X (b) Registry, h_A^X
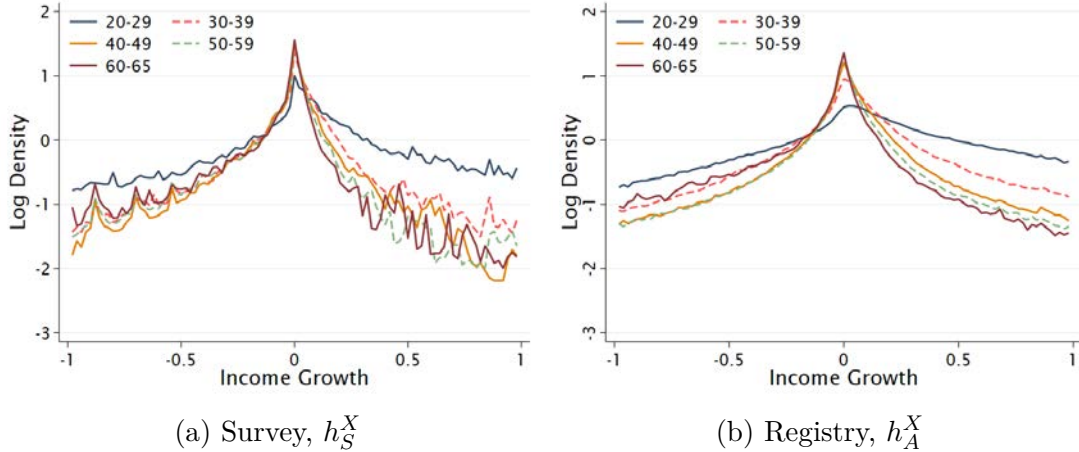
Note: Panel (a) plots log density for the pooled distribution of expected holistic earnings growth rates from the survey, h_S^X, where X indicates partitions by age groups. (b) plots the distribution of annual earnings growth from 2019 to 2020 as observed in the administrative data for the full population, 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 2019. Survey results are weighted using population weights. Online Appendix C.6 shows the corresponding figures using administrative data for 2019.

Figure 6: Pooled earnings risk and registry earnings risk

who have similar observable characteristics, X.

3.4 Comparing the Distribution of Earnings Growth

We now compare the pooled distribution of subjective holistic expected earnings growth distributions with the distribution of realized earnings growth observed in in population wide administrative data. 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.[16]

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,

---

[16] 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.5 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.

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.[17] 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.7, 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.

3.5 Four Moments of Earnings Risk over the Life Cycle

In Figure 7 we examine how the moments of the distributions h_S^X and h_A^X evolve over 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 be relatively more uncertain about their earnings growth. 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.[18]

In Online Appendix C.9, we further divide the survey and the registry data into three broad age groups and earnings deciles within each age group. Again we find very similar patterns between the survey and the registry. The comparison is based on administrative data for 2020. Once again, the low impact of COVID is confirmed since we find very similar patterns when we compare survey results to administrative data in 2019 (see

---

[17] 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.

[18] In Online Appendix C.8, we document that also the standard moment measures (standard deviation, skewness, and kurtosis) show similar patterns in the registry and the survey.
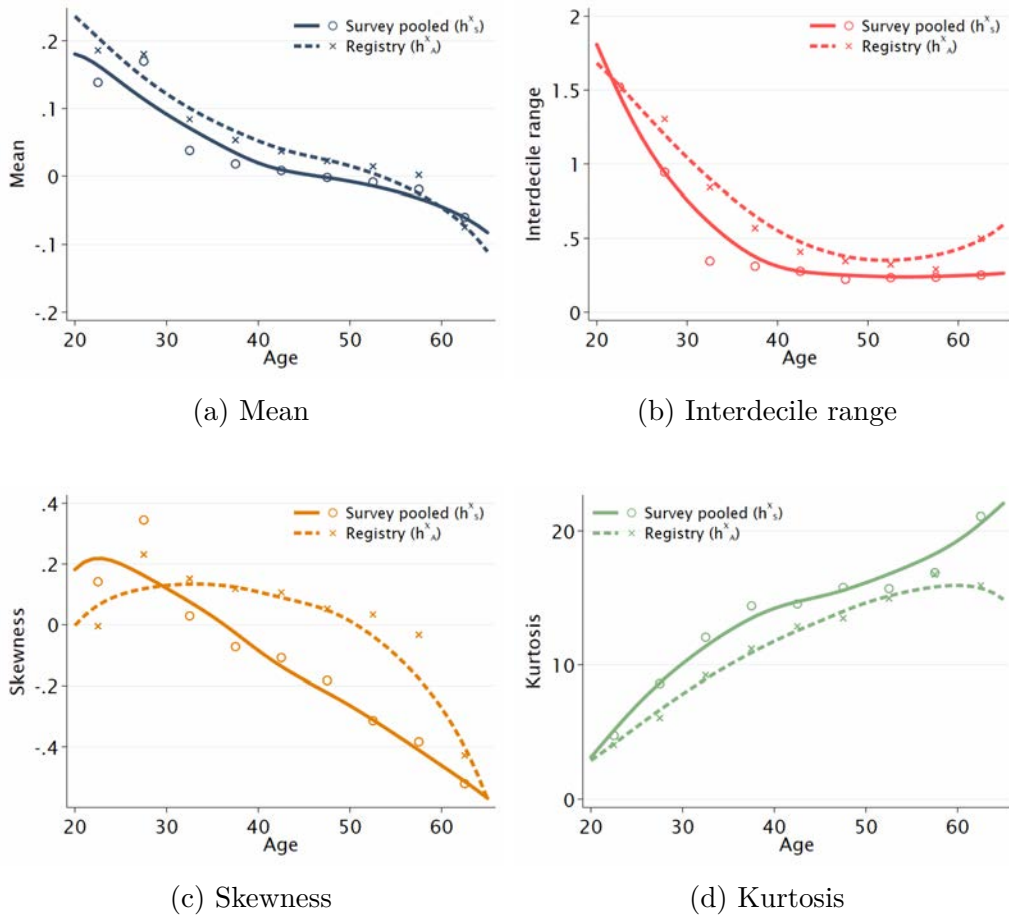
(a) Mean

(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 4) 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 2020, respectively. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights. Online Appendix C.8 shows the corresponding figures using standard moments.

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

Online Appendix C.6). 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.

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, groups 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 these 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.

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 uncertainty inferred from the administrative data tends to be larger than uncertainty 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:

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

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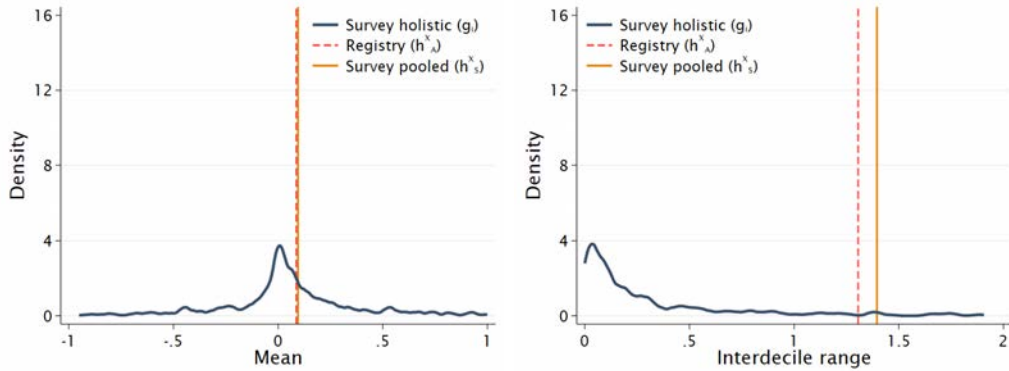
(a) 20-34:*Low*, Mean

(b) 20-34:*Low*, Interdecile range
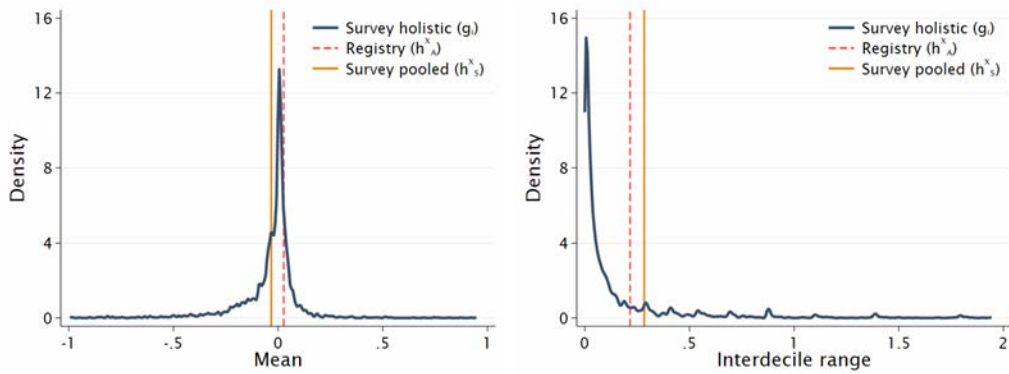
(c) 50-65:*High*, Mean

(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

24

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 is confounded with heterogeneity.[19]

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 which reproduces Figure 1.[20]

We find that the average of the subjective interdecile ranges, $\frac{1}{N^X} \sum_i^{N^X} \mathrm{IDR}_i[g_i^X]$, within each cell is much smaller than the interdecile range calculated from the administrative data within the same cell, $\mathrm{IDR}[h_A^X]$. Consistent with this, 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.[21]

---

[19] 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.

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

[21] 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 1 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 on growth rates. The resulting figure is practically identical to Figure 1. These results are
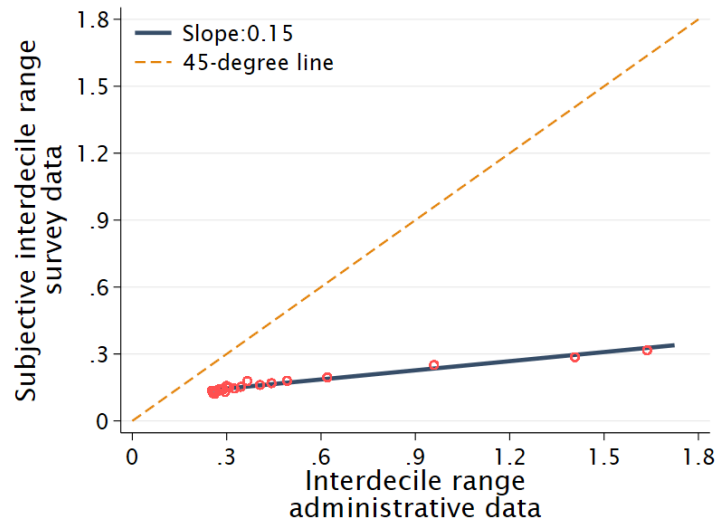
Note: The figure compares average interdecile ranges of subjective holistic earnings expectations, $\frac{1}{N^X}\sum_i^{N^X} \mathrm{IDR}_i[g_i^X]$, to interdecile ranges calculated from administrative data, $\mathrm{IDR}[h_A^X]$, within 300 cells divided by age groups (20-34, 35-49, 50-65) and earnings percentiles. The panel shows a binned scatterplot (red circles) of $\frac{1}{N^X}\sum_i^{N^X} \mathrm{IDR}_i[g_i^X]$ by vigintiles of $\mathrm{IDR}[h_A^X]$. A regression line based on the 300 data points is overlaid.

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

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 uncertainty, which we measure as the average interdecile range, pertaining to earnings growth over the life cycle. Considering uncertainty based on subjective holistic earnings growth expectations we find that uncertainty is highest for young people. Fixing earnings growth uncertainty to come only from the stay branch generates a big drop in uncertainty at all ages, but most dramatically for the young. This shows that uncertainty 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.

Overall, job transitions are essential in determining life cycle patterns of higher order mo-

---

reported in Online Appendix E.5.

ments of subjective holistic one-year ahead earnings risk. This raises an obvious question of how well our findings on subjective earnings risk match the implications of state of the art models of job transitions over the life cycle, to which we turn in the next section.

6 Subjective Earnings Risk in a Search Model

As Dominitz and Manski (1997) noted, a key use of subjective earnings risk is to discipline models of job search. The *Copenhagen Life Panel* was designed with this in mind, given its focus on job transitions and earnings risk. In this section, we consider how well a state of the art life cycle model of the labor market developed by Menzio et al. (2016)[22] fits our data when estimated in the standard manner. This model is designed to explain the life-cycle profiles of the employment-to-employment (EE), employment-to-unemployment (EU), and unemployment-to-employment (UE) rates as well as average wages. As a result, it can endogenously generate age variation in these rates, as well as time out of work.[23] In our survey, we measure these expectations directly and show that they have important life-cycle patterns that feed into earnings risk. A further motivation for this model class is that, given our focus on transitions, it gives us the exact level of detail needed for explaining them. 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 how earnings risk is related to job transitions and for the data that we have at hand.

Regardless of the specifics, models in this literature are typically calibrated to match observational data on job transitions. Consequently, expectations are treated as unobserved and are inferred through assumptions. Little is known about whether the *implied* expectations of workers in such calibrated models are actually consistent with the subjective expectations data. To learn more, we calibrate the model to Danish administrative

---

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

[23] Papers in the style of Bagger et al. (2014) often impose that unemployed workers accept all jobs and/or a constant exogenous job destruction rate. Although the arrival and destruction rates can simply be made age-dependent or targeted by age, like in Karahan et al. (2022), we wanted to use a model that has the potential to generate the right patterns on its own. This is important for our exercise because we want to identify which features of the model (that have economic interpretations) can or cannot generate beliefs in line with the CLP, without adding too much complexity.
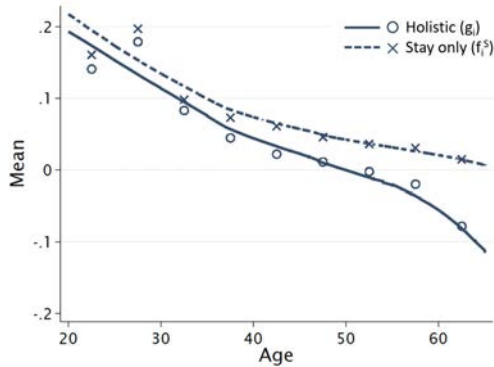
(a) Mean
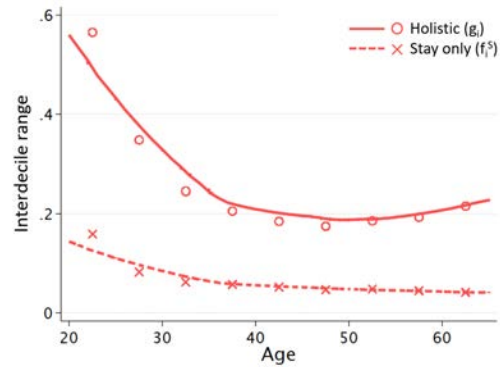
(b) Interdecile range
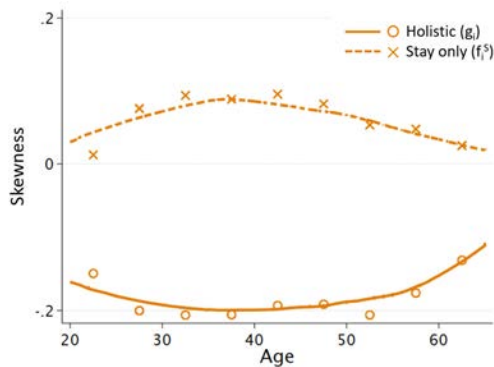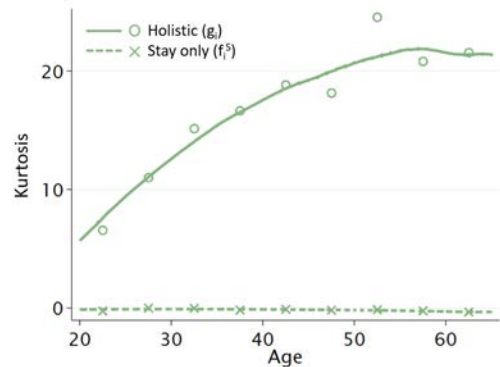
(c) Skewness

(d) Kurtosis

Note: The figure shows the average value of the 1st to 4th quantile based moments (See notes to Figure 4) 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.10 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

data, back out the implied beliefs about earnings risk, and examine both the consistencies and departures from the subjective expectations that we measure in the *Copenhagen Life Panel*.

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, 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 wage 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.

---

[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.

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 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 Belief Simulation in the Search Framework

We calibrate the model in the standard manner using data on employment and wage outcomes from the registry, where the raw data is measured at the monthly frequency. The key moments that we target are the EU and EE rates as a function of tenure, and wages as a function of age. This strategy allows 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. If the model's mechanisms can explain realizations

in the registry, then they are reasonable starting points for exploring whether they are also relevant for beliefs. The details on the calibration and model fit for both targeted and untargeted moments are in Online Appendix F.3.

With this calibration in hand, our next step is to generate the beliefs for a cross-section of model-simulated workers. The beliefs we are interested in are the subjective distributions of earnings conditional on staying, quitting, and being laid off, the probabilities of these events, and the associated duration of out-of-work times – exactly as if they were respondents to the CLP.

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. Our beliefs about the probability of job transitions come from simulating each worker's states forward many times, and for the branch-specific beliefs, putting each worker on each branch and simulating forward from there. These paths represent 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.

We first do one set of simulations to recover workers' beliefs about job transitions: the probability of staying with their employer (p_i^S), undergoing an EE transition (p_i^Q), and undergoing an EU transition (p_i^L) at some point within the next year. For each of these, the proportion of paths in which the transition occurs is the worker's belief about the likelihood of the transition, which map to the first set of CLP questions. Note that here, to facilitate our comparison, we are equating quits in the CLP with EE in the model and layoffs in the CLP with EU in the model.[25]

To generate beliefs for periods until re-employment and earnings growth, we perform another set of simulations branch by branch. On the stay branch, we shut down exogenous job destruction and on-the-job search to obtain scenarios in which the worker stays with their employer. Calculating their average monthly earnings in each scenario enables us to obtain the distribution f_i^S.[26] To simulate each worker's EE branch, we put them in

---

[25] 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. We found signs in the CLP that suggested that these may be commonly-held beliefs, such as the presence of positive time-out-of-work after a quit. We think that the CLP can be helpful for understanding these scenarios and filling in any gaps, and plan to explore this further in future work.

[26] On each branch, the beliefs about conditional earnings growth come from the log difference between

a new job (with no intervening time out of work) of unknown quality and allow each path to evolve for 12 months according to the data-generating process of the model.[27] Their earnings on each path contribute to the distribution f_i^Q. For the EU branch, we send people to unemployment and simulate their paths forward as they eventually find new jobs. We simulate for a longer time – 3 years – to ensure that we have a longer series of earnings once they get to their new employer. Their earnings at their new employer contribute to the distribution of f_i^L. The model-simulated beliefs about time to re-employment, n_i^L, are represented by the average length of the unemployment period on the EU branch.[28] For additional details on the belief simulation, see Online Appendix F.4.

6.4 Results: Survey vs. Model Beliefs

In this section, we present our comparison between the beliefs in the model and survey. To summarize them, we create a figure comparable to Figure 4. The main findings are in Figure 11.

The left side compares the average branching probabilities and average length of the re-employment periods, with the numbers from the survey in the parentheses. We find that the model's implications for these beliefs align well with those of the CLP respondents. This is consistent with our earlier results comparing these outcomes between the survey and register. Since these matched well and the calibration was targeted to the registry, it follows that the model is closely in line with the survey.
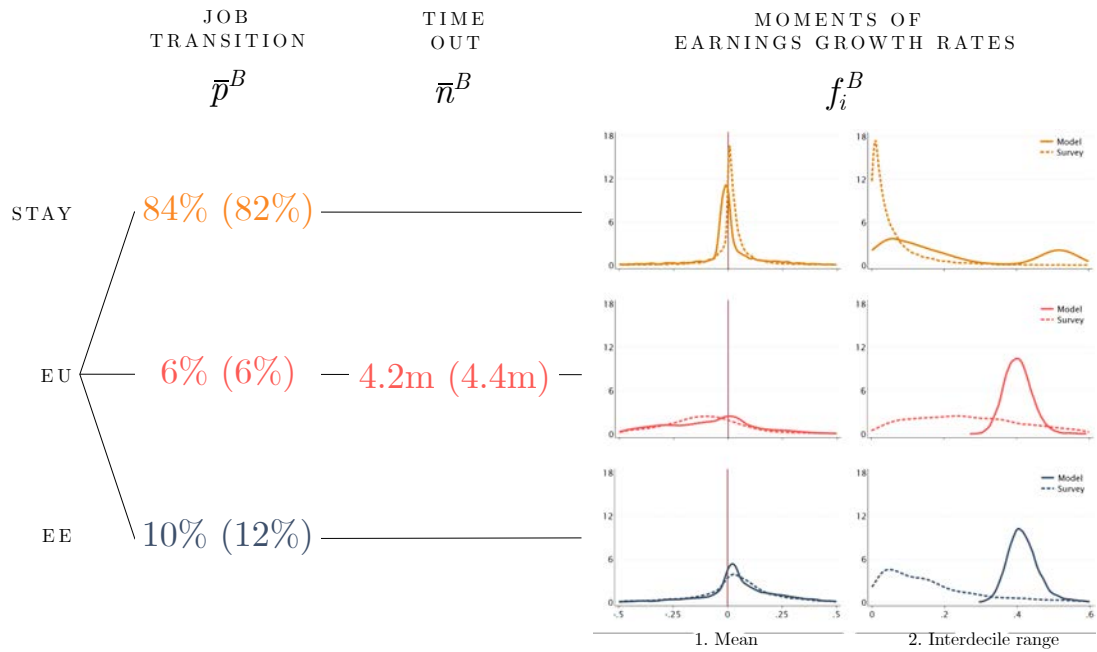
Figure 11 also compares the distribution of the moments of conditional earnings growth. On the stay branch, the distribution of means in the model is similar to the survey, but with more small negative values. These distributions are also quite similar on the other branches. The model can correctly capture a lot of negative values on the EU/laid off branch because transitions to unemployment involve a pause in human capital growth and often lead to matches that have lower quality than the previous one. On the EE/quit branch, the model correctly picks up many high earnings growth realizations. This is because the workers who search while employed have low enough match qualities and

---

average monthly first-year earnings in the new job and the monthly earnings in the old job, exactly as we measured it when analyzing the survey responses.

[27] Note that in the model there is a subset of workers who have 0% probability of an EE transition because they have high enough match quality. We exclude them from this branch, which is consistent with the survey – respondents in the CLP who reported a 0% chance of quitting were not asked any further questions about this branch.

[28] Note again that since we are mapping EEs to quits, we do not have a model counterpart of n_i^Q.

	JOB TRANSITION $\bar{p}^B$	TIME OUT $\bar{n}^B$	MOMENTS OF EARNINGS GROWTH RATES f_i^B
STAY	84% (82%)		
EU	6% (6%)	4.2m (4.4m)	
EE	10% (12%)		

1. Mean   2. Interdecile range

Note: The figure shows conditional values implied by the model, where the rows correspond to the branches "Stay", "Employment-to-Uuemployment", and "Employment-to-Employment". 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 and 2nd moment of the model implied conditional earnings distributions, f_i^B. (See also notes to Figure 4). Numbers in parentheses as well as dashed lines in the distributions are from the survey results for comparison. Survey results are weighted using population weights.

Figure 11: Overview of belief comparison between survey and model

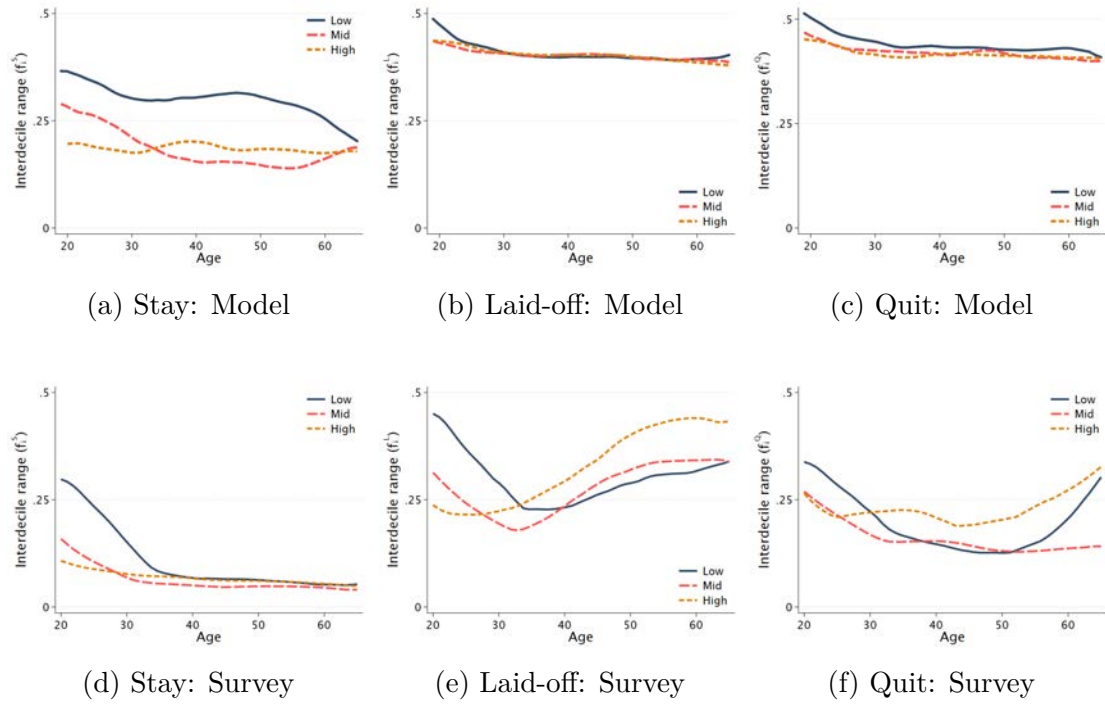know that they should be able to gain once they move jobs.

The departures become evident when examining the variance of earnings growth in each of the three events. Recall that in the survey, people viewed the stay branch as having very little risk. In contrast, agents on the stay branch in the model have two sources of earnings changes, leading to the double-peaked distribution in Figure 11. The lower peak comes from the resetting of match quality for workers with known quality; the upper peak comes from the initial realization of quality for workers of unknown quality.

On both the quit and layoff branches, the interdecile range is much higher in the model compared to the survey. Moreover, there is less heterogeneity in the model compared to the survey. This happens because all workers of the same age and experience level face the exact same job search environment after separating from their jobs: they all draw from the same match quality distribution and everybody "starts from the bottom" of the job ladder after going into an unemployment spell. As a result, the model delivers higher and less heterogeneous levels of risk than does the survey.

This means people believe they have less risk than the equivalent agent of their type in the model, or that agents in the model overestimate this risk. This finding echoes the results found in Figure 9 which compares the subjective and registry inferred risk. This is to be expected given that the model was calibrated to the registry, but these results shed light on why search models may miss these features of beliefs. Compared to the model, this finding suggests that workers have more information about what kinds of jobs may be available to them, or that their own match quality is not that unpredictable.

In the survey, there is also a clear pattern where workers view quits and layoffs as different events with different levels of risk. This can be seen by noting the disparities between the distributions of interdecile range in the survey. However, in the model, workers have similar interdecile ranges regardless of whether an EE or EU occurs. This is another major place where the model beliefs depart from the survey beliefs: the model gets right that EEs typically offer higher returns than EUs, but does not capture the fact that in the survey, people say that EEs are generally less risky. The reason is related to the logic above: when someone enters a new job in the model, whether they came straight from employment or unemployment does not matter for how they fare in the new job. In either case, they start off as unknown and face the same distribution of match quality when it is revealed.

To further explore the difference between the model and the survey, we examine the second moment as a function of current earnings and age. Figure 12, panels (a) - (c)

(a) Stay: Model   (b) Laid-off: Model   (c) Quit: Model

(d) Stay: Survey   (e) Laid-off: Survey   (f) Quit: Survey

Note: The figures show interdecile ranges of conditional earnings risk for three terciles of earnings (Low, Mid, High) over the life cycle. The top row shows model implied risk, the bottom row shows subjective risk. The lines are local linear polynomials, calculated using a bandwidth of 4 years. Survey results are weighted using population weights.

Figure 12: Heterogeneity in interdecile range of f_i^B by transition status, age and earnings level

show interdecile range of conditional earnings growth in the model across the life cycle for 3 earnings groups: low, medium, and high. Panels (d) - (f) show the corresponding expectations elicited in the survey and they document different patterns of heterogeity than what is inferred from the model.

On the layoff and quit branches, it is clear that the model cannot generate enough variation in risk across the age and earnings distributions. For nearly all groups, the model implies an overestimation of risk, with the highest risks being for young and low earning workers on both the quit and layoff branches. It cannot account for the more complex patterns seen in the survey, such as the opposite patterns in layoff risk for low- and high-earning workers, and the increase in risk around quits for older workers. On the other hand, the model can generate more age and earnings variation in risk on the stay branch, albeit with different patterns compared to the survey. However, as we have shown with the survey data, job transitions are very important for overall risk. Since the model cannot account for the risks associated with them, the model's mechanisms may not be so useful for understanding where these survey expectations come from.

These results are informative for identifying how models of the labor market can better align themselves with subjective beliefs collected from surveys like the CLP. Although we compare our data with beliefs implied by Menzio et al. (2016), the features of that give rise to the discrepancies are not unique to this model. Uncertainty about productivity at the start of new matches and/or forcing all workers to start from the bottom of the job ladder after going through unemployment are commonly seen attributes. Moreover, to address the overestimation of risk, the use of expectations rather than outcomes data is crucial. The data from surveys like the CLP can be used in future work to address these gaps.

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. It also reveals subjective earnings risk to be significantly lower than its administratively-estimated counterpart, since expected earnings growth is heterogeneous even within narrow demographic and earnings cells. We also show possible job transitions to be central determinants of subjective earnings risk. We calibrate a life-cycle model of search and matching in the standard manner based on administrative data. We show that the calibrated model produces far higher estimates of individual earnings risk than

do our subjective expectations whether or not workers switch jobs. Our results highlight the value of using survey-based measures of subjective earnings risk in modeling labor market transitions and other key choices impacted by earnings risk, such as savings and portfolio decisions.

References

Abowd, J. M. and D. Card (1987). Intertemporal labor supply and long-term employment contracts. *The American Economic Review*, 50–68.

Almås, I., O. Attanasio, and P. Jervis (2023). Economics and measurement: New measures to model decision making. Technical report, National Bureau of Economic Research.

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.

Bachmann, R., G. Topa, and W. van der Klaauw (2022). *Handbook of Economic Expectations*. Elsevier.

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.

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.

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.

Deaton, A. et al. (1992). *Understanding consumption.* Oxford University Press.

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.

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).

Gottschalk, P. and R. A. Moffitt (1994). Welfare dependence: Concepts, measures, and trends. *The American Economic Review 84*(2), 38–42.

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.

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

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.

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. (2004). Measuring expectations. *Econometrica 72*(5), 1329–1376.

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.

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.*

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

Zeldes, S. P. (1989). Consumption and liquidity constraints: an empirical investigation. *Journal of Political Economy 97*(2), 305–346.