

NBER WORKING PAPER SERIES

SUBJECTIVE EARNINGS RISK

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

Working Paper 31019  
<http://www.nber.org/papers/w31019>

NATIONAL BUREAU OF ECONOMIC RESEARCH  
1050 Massachusetts Avenue  
Cambridge, MA 02138  
March 2023, Revised January 2023

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

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2023 by Andrew Caplin, Victoria Gregory, Eungik Lee, Søren Leth-Petersen, and Johan Sæverud. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

## Subjective Earnings Risk

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

NBER Working Paper No. 31019

March 2023, Revised January 2023

JEL No. J30

### **ABSTRACT**

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

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

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

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

---

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

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

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

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

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

---

<sup>2</sup> This model is a life-cycle extension of the directed search model of [Menzio and Shi \(2011\)](#).

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

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

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

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

---

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

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

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

---

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

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

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

## 2 The Conditional Earnings Instrument

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

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

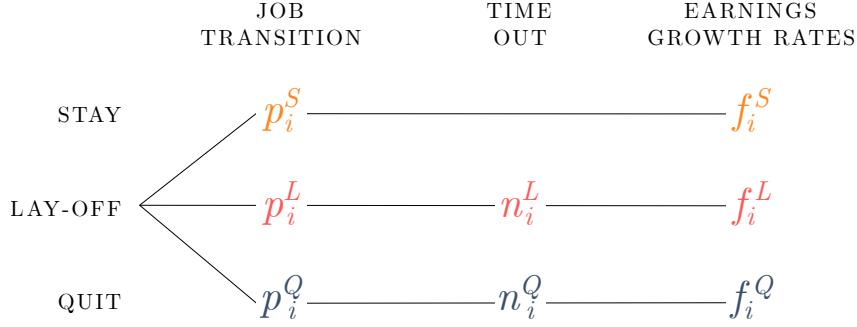
### 2.1 Job transitions

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

- *Please think about your possible relationship with your current employer in 2021. Assign the probability in each possible case. The sum of the probabilities should be 100.*
  1. *Staying with your current employer during 2021*
  2. *Being laid off from your current employer at some point during 2021*
  3. *Quitting from your current employer at some point during 2021*

---

<sup>6</sup> The questionnaire is reported in Online Appendix A.



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

Figure 1: Survey instrument overview

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

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

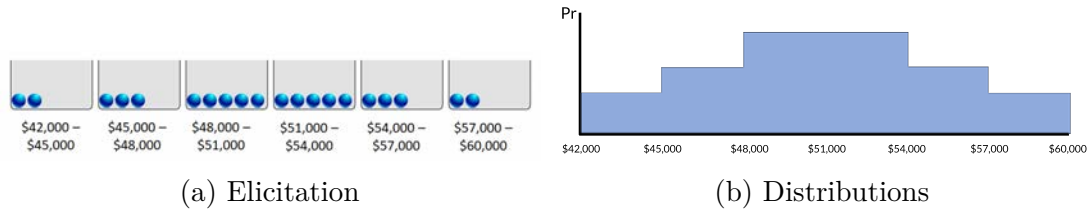
For those who report a positive layoff probability, we follow up by asking about how long they expect to be out of work, and we do this by asking the likelihood of being re-employed within four different horizons: 1, 3, 12, and 24 months:

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

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

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





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

Figure 2: Balls in bins

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

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

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

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

In the survey we ask respondents about last year’s earnings.<sup>8</sup> 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*<sup>9</sup> (CLP) which is an online panel survey implemented in Denmark. We invite a random selection of individuals, who are recorded in the Danish population registry and aged between 20 and 70, to participate in the survey. The population registry is a complete registry of all persons who are born or have ever had an address in Denmark. It contains a personal identifier (CPR-number) applied universally to record any contact an individual has with the public sector. Invitations to participate were sent out using an official email account, called *e-boks*, which all Danes are equipped with. For the purpose of this paper we consider questions about earnings expectations and job transitions that were included in CLP issued in January 2021.<sup>10</sup>

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

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

---

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

<sup>9</sup> The *Copenhagen Life Panel* is an ongoing survey that was initiated in 2020 and is issued every year in January.

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

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

## 2.3 Job Transitions

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

Moving to the right in Figure 3, we report the average expected time out of work upon quitting or being laid off,  $\bar{n}^B$ . To arrive at a summary measure of the duration out of work we aggregate over the likelihood of being out of work for the four horizons that we ask about for the quit/layoff branches. Focusing on time out of work following a layoff, respondents report  $(n_{i,1}^L, n_{i,3}^L, n_{i,12}^L, n_{i,24}^L)$  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:

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

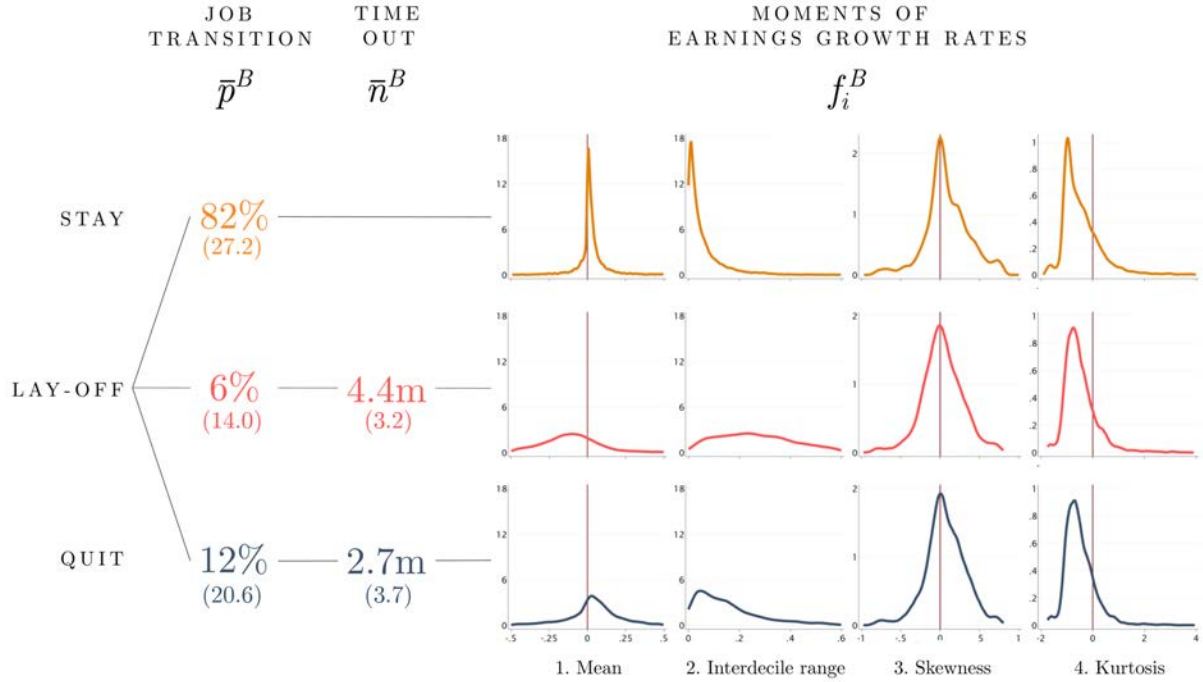
We use the same procedure for the re-employment period after a quit,  $n_i^Q$ .

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

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

---

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



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

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

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

## 2.4 Conditional Earnings Risk

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

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

<sup>12</sup> 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).

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

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

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

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

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

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

### 3 Comparing Survey and Administrative Data

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

---

<sup>14</sup>In Online Appendix C.2, we reproduce Figure 3 using standard measures of the moments.

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

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

### **3.1 Job Transitions**

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

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

### 3.2 Time Out of the Labor Force

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

$$n_i^{L,Q} = \frac{p_i^Q n_i^Q + p_i^L n_i^L}{p_i^Q + p_i^L} \quad (1)$$

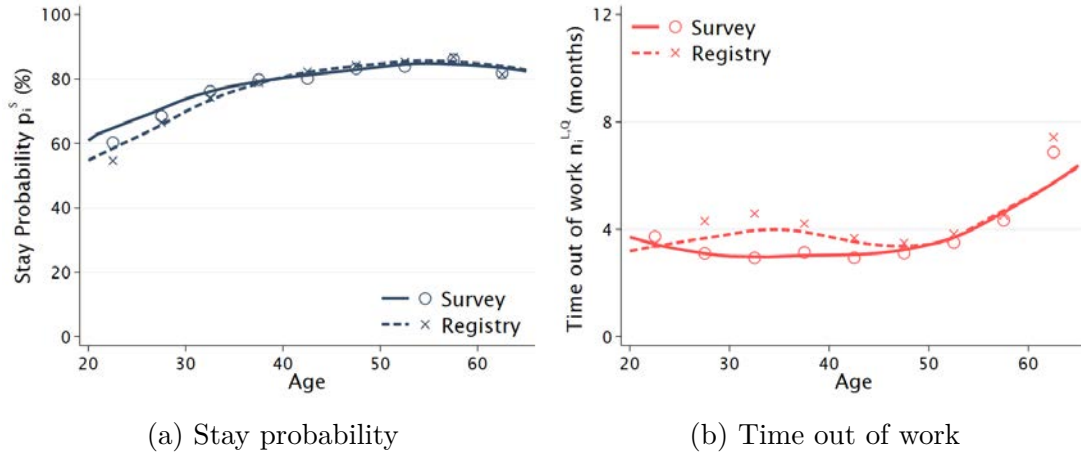
The result of the comparison is shown in Figure 4, panel (b). According to the survey (solid line), the average expected time spent out of work following a job separation is about 3.5 months for people aged less than 50 and the expected time out of work increases dramatically for workers aged 50+. The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

### 3.3 Overall Subjective Earnings Expectations

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

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





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

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

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

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

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

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

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

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

---

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

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

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

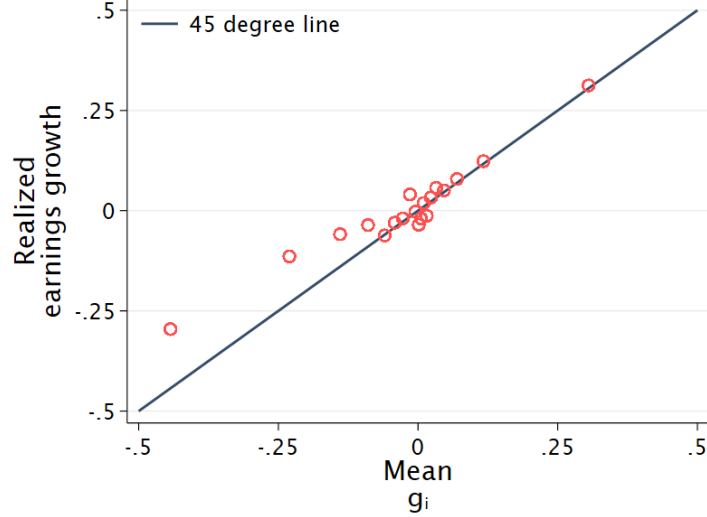
### 3.4 The Cross-Sectional Distribution of Earnings Growth

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

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

---

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



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

Figure 5: Comparing mean  $g_i$  and realized earnings growth

characteristics  $X$ :

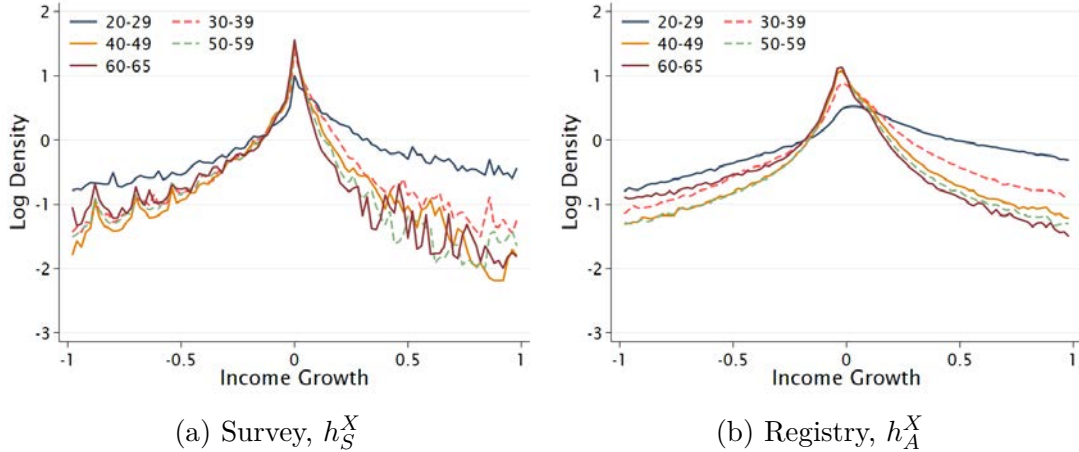
$$h_S^X = \frac{1}{N^X} \sum_{i=1}^{N^X} g_i^X \quad (3)$$

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

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

---

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



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

Figure 6: Pooled earnings risk and registry earnings risk

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

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

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

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

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.<sup>20</sup>

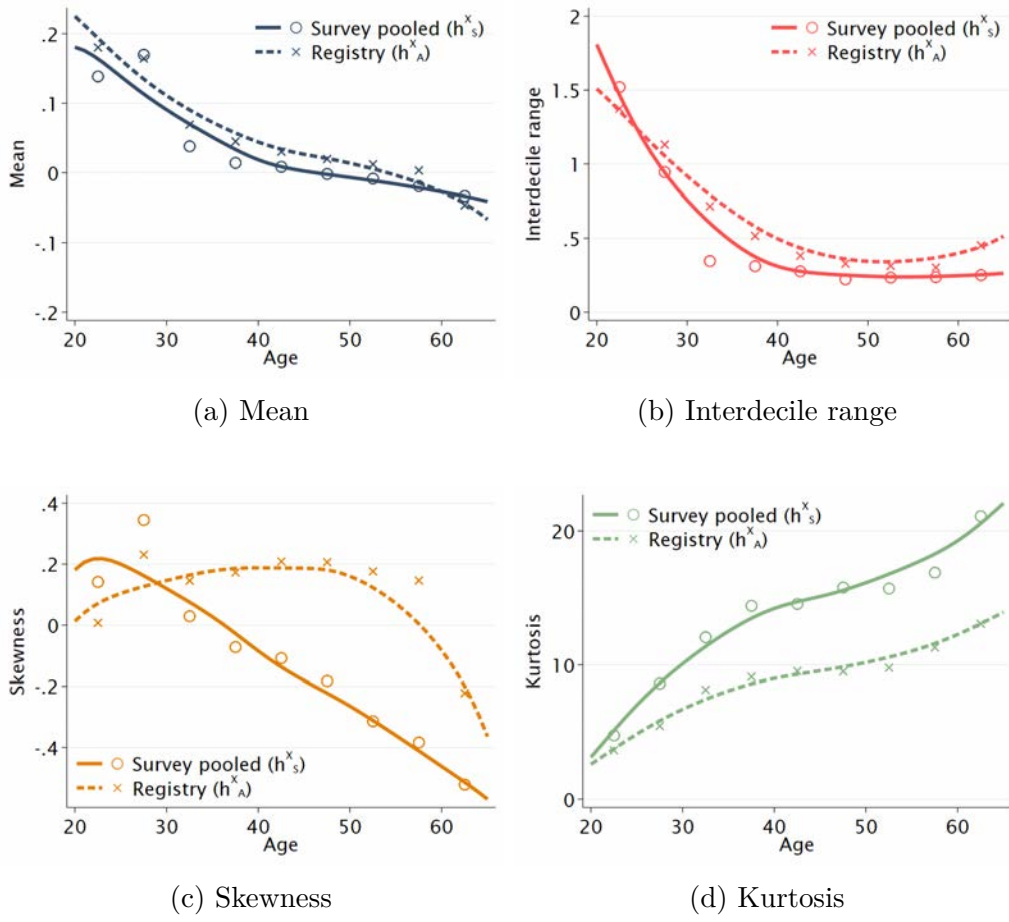
## 4 Administrative versus Survey-Based Earnings Risk

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

---

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

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



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

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

## 4.1 Coarse Stratification

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

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

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

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



distribution of subjective interdecile ranges is smaller than for the 20-34:*Low* group.

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

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

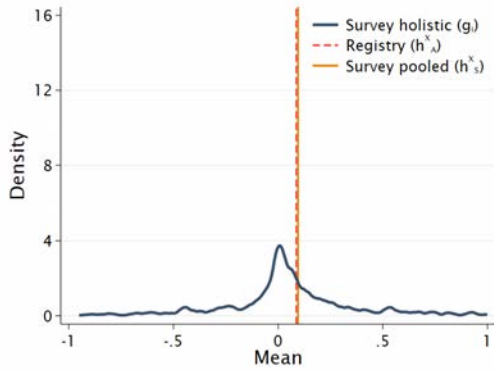
The variance of the mixture distribution is the mixture of the variances of the subjective distributions plus a non-negative term reflecting the differences in means between the subjective distributions. By Jensen's Inequality the average squared mean is weakly greater than the squared average mean,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

## 4.2 Refining the Stratification

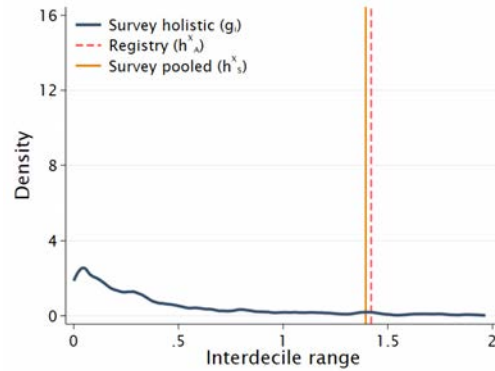
The logic above suggests that the gap between subjective and administratively estimated risk will be lower the more we refine the stratification of the population. To pursue this we now consider a finer stratification. Specifically we make use of the administrative data and partition the distribution 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,  $\text{IDR}[h_A^X]$ , and the average of the subjective interdecile ranges within each of these cells,  $\frac{1}{N^X} \sum_i \text{IDR}_i[g_i^X]$ . The result is shown in Figure 9.<sup>22</sup>

<sup>21</sup> 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.

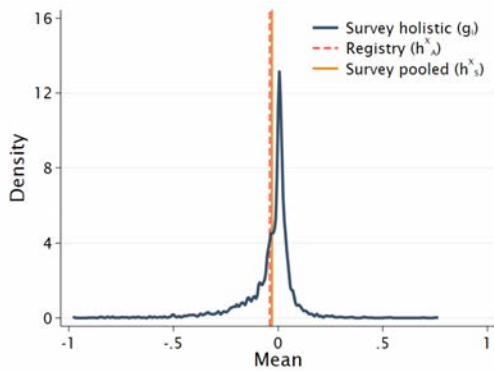
<sup>22</sup> in Online Appendix E.3, we report results for skewness and kurtosis.



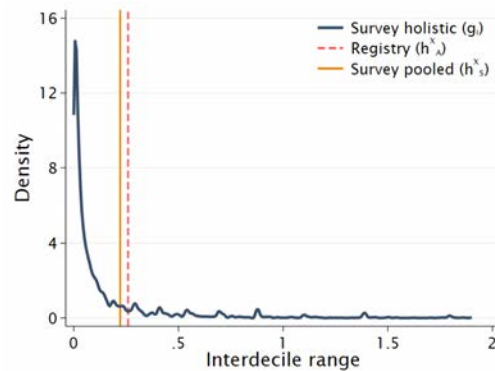
(a) 20-34:Low, Mean



(b) 20-34:Low, Interdecile range



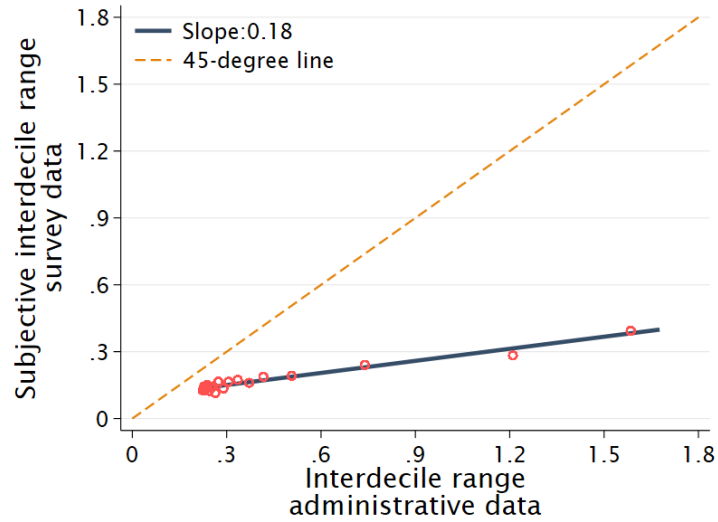
(c) 50-65:High, Mean



(d) 50-65:High, Interdecile range

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

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



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

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

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

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

### 4.3 The Role of Observable Characteristics

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

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

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

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

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

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

The result is shown in Table 1. In column (1),  $\mathbb{E}_i[g_i]$  is regressed on a full set of age group dummies. The  $R^2$  is 0.015 meaning that age group dummies are able to explain only 1.5% of the variation in  $\mathbb{E}_i[g_i]$ . As a result,  $\text{IDR}[h_A^X] = 0.621$  is only slightly lower than  $\text{IDR}[h_A] = 0.630$  and much bigger than the average subjective interdecile range,  $\text{IDR}_i[g_i^X] = 0.179$ . In column (2) we add percentile dummies for the level of earnings. This increases  $R^2$  to 0.028 and reduces  $\text{IDR}[h_A^X] = 0.602$ . In column (3)-(6) we sequentially add a dummy for being female, a dummy variable for having a college degree, quintile dummies for the average earnings growth during 2016-2020, i.e., the past

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

## 5 Job Transitions and Subjective Earnings Risk

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

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

Panel (b) shows how the average subjective risk, which we measure as the average interdecile range, pertaining to earnings growth over the life cycle. Considering risk based on subjective holistic earnings growth expectations we find that risk is highest for young peo-

ple. Fixing earnings growth risk to come only from the stay branch generates a big drop in risk at all ages, but most dramatically for the young. This shows that risk pertaining to one-year ahead earnings growth is intimately tied to job transitions.

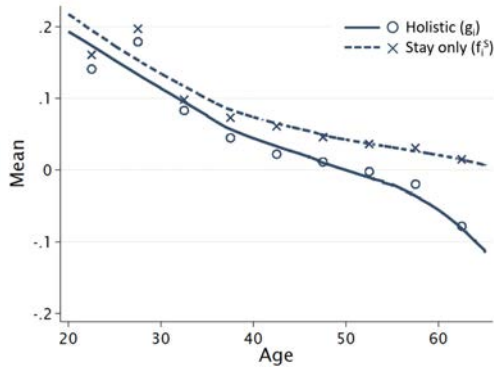
In Panel (c) we consider skewness. For all ages there is, on average, negative skewness in the subjective holistic distributions. Yet when quantifying skewness only from the stay branch, it is close to zero. Negative skewness appears when people expect to disproportionately draw large negative shocks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

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

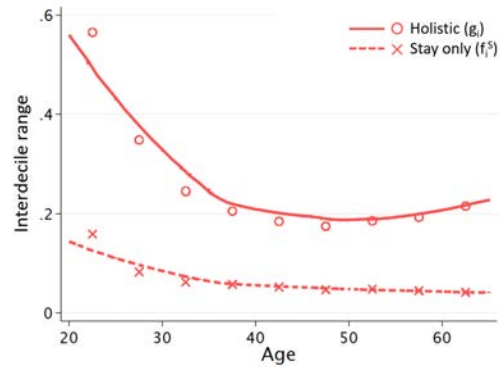
Thus, we find that job transitions are essential in determining life-cycle patterns of higher order moments of subjective holistic one-year ahead earnings risk. This finding is the analog to the differences between job switchers' and job stayers' earnings growth in realizations from administrative data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

## 6 Subjective Earnings Risk in a Search Model

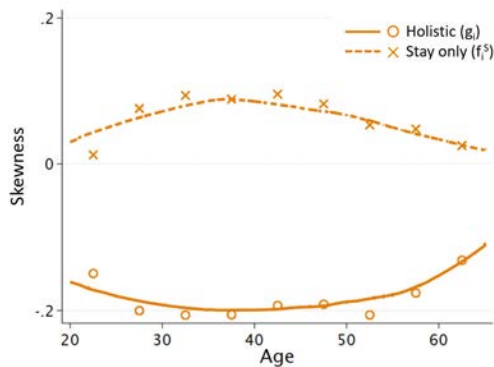
Thus far, we have established several new facts about subjective earnings risk. We found that subjective earnings risk as reported by our survey respondents is about one-third of what could be inferred from administrative data on earnings realizations. The realized distribution of earnings growth confounds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



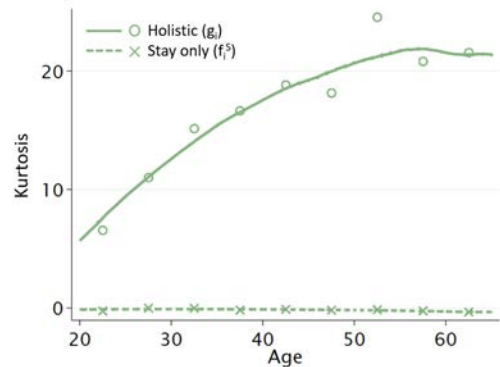
(a) Mean



(b) Interdecile range



(c) Skewness



(d) Kurtosis

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

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



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

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

## 6.1 Model Description

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

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

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

---

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

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

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

The aggregate state of the economy is  $\psi = (n, u, e, \gamma)$ .  $n(t)$  is the measure of workers of age  $t$  in the labor market.  $u(y, t)$  is the measure of unemployed workers.  $e(z, y, t)$  is the measure of workers of type  $(y, t)$  with match quality  $z$ . When matches first form, the quality is unknown:  $z = z_0$  denotes this case.  $\gamma$  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  $\mu_t$ . A fraction  $\nu_t$  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  $\delta$ . 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  $\lambda_e$  and unemployed workers search with probability  $\lambda_u$ . 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.<sup>24</sup> 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  $\eta$  from the same distribution. This reflects exogenous changes in productivity that can make this particular match better or worse: the firm implements a new technology, the worker gets better at their tasks in this job, etc.

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

## 6.2 How the Model Works

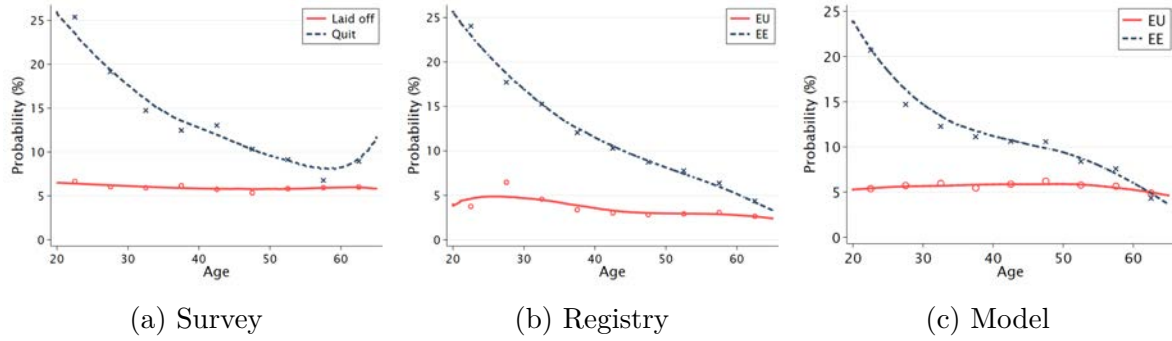
This section gives a brief overview of the mechanisms and key forces driving the model. In particular, we highlight how job transitions and earnings risk unfold and where they come from – these are key objects that we will link back to our survey responses and the Danish register. For more details on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

Transitions from employment to unemployment are triggered by changes in match quality  $z$ . Every match has a reservation match quality  $r_t(y)$ . If  $z$  is below the  $r_t(y)$ , the match is immediately destroyed. If  $z$  is above, it is kept. Updates to  $z$  occur in two scenarios: when  $z$  is revealed after being unknown in a new match (with probability  $\alpha$ ) and when it is redrawn (with probability  $\eta$ ). 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

---

<sup>24</sup> 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.



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

Figure 11: Job transition probabilities

from the switch.

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

### 6.3 Calibration and Belief Simulation

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

This calibration strategy also has the added benefit of allowing the model to endogenously generate the age profile of the EU and EE rates. This ensures that the model can deliver on its own the correct transition patterns by age. Even though these were not targeted, panels (b) and (c) of Figure 11 show that the model does a remarkable job at matching the registry along these dimensions. This confirms that the model’s mechanisms are a

good starting point for understanding job transitions over the life cycle, a key ingredient for earnings risk. It is also remarkable that the survey patterns match up as well<sup>25</sup> – this says that on average and within age groups, people are correct about the chances of undergoing either one of the transition types. The alignment of these three objects will be important later for understanding the model’s patterns in beliefs. Online Appendix F.3 further elaborates on the model’s performance with respect to other untargeted moments.

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

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

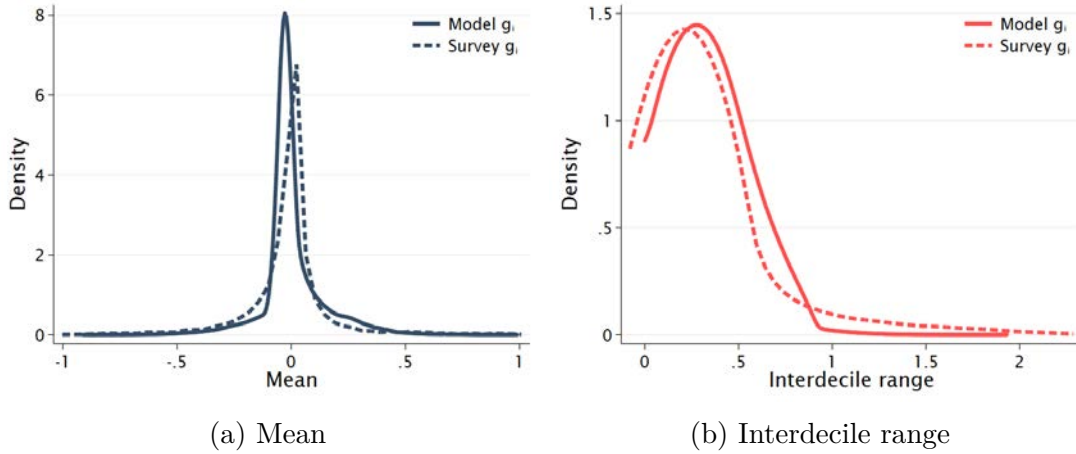
## 6.4 Comparing Survey and Model Beliefs

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

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

---

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



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

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

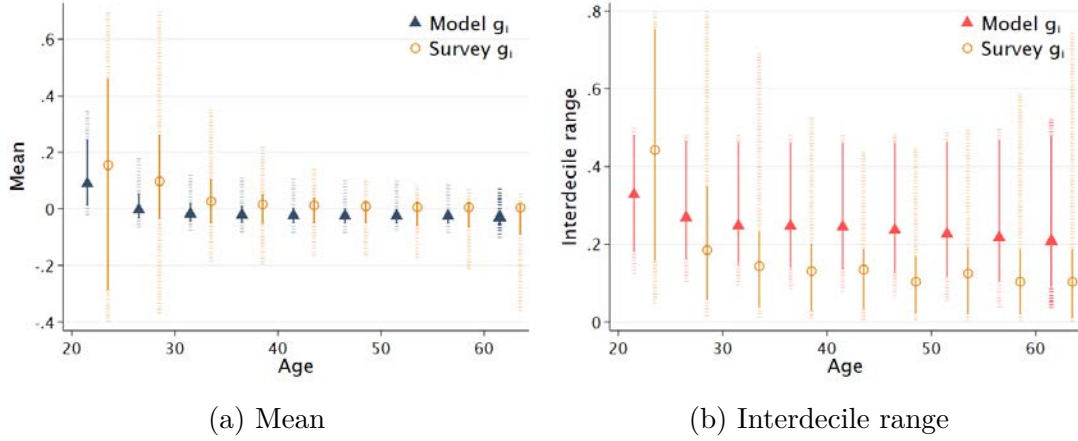
to have low but positive mean earnings growth. There are individuals who are outliers in both directions and the model picks this up as well. The interdecile ranges also have similar shapes. Like in the survey, most agents in the model have low levels of subjective risk, but there is a long right tail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

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

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

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

<sup>26</sup> Note that our quantification here is nearly equivalent to what we would get had we targeted features



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

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

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

## 6.5 Match Quality and Heterogeneity in Subjective Earnings Risk

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

---

of the subjective beliefs directly.

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

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

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

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

---

<sup>27</sup> The same is true for earnings: there is more dispersion in match quality among lower-earning workers.



mixture result in Section 4.

## 7 Conclusion

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

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

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

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

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

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

## References

- Altonji, J. G., A. A. Smith Jr, and I. Vidangos (2013). Modeling earnings dynamics. *Econometrica* 81(4), 1395–1454.
- Andersen, H. Y. and S. Leth-Petersen (2021). Housing wealth or collateral: How home value shocks drive home equity extraction and spending. *Journal of the European Economic Association* 19(1), 403–440.
- Arcidiacono, P., V. J. Hotz, A. Maurel, and T. Romano (2020). Ex ante returns and occupational choice. *Journal of Political Economy* 128(12), 4475–4522.
- Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review* 104(6), 1551–96.
- Bagger, J. and R. Lentz (2019). An empirical model of wage dispersion with sorting. *The Review of Economic Studies* 86(1), 153–190.
- Balleer, A., G. Duernecker, S. Forstner, and J. Goensch (2021). The effects of biased labor market expectations on consumption, wealth inequality, and welfare.
- Balleer, A., G. Duernecker, S. Forstner, and J. Goensch (2023). Wage bargaining and labor market policy with biased expectations.
- Bick, A., A. Blandin, and K. Mertens (2021). Work from home before and after the covid-19 outbreak. *Available at SSRN 3786142*.
- Browning, M., M. Gørtz, and S. Leth-Petersen (2013). Housing wealth and consumption: a micro panel study. *The Economic Journal* 123(568), 401–428.
- Busch, C., D. Domeij, F. Guvenen, and R. Madera (2022). Skewed idiosyncratic income risk over the business cycle: Sources and insurance. *American Economic Journal: Macroeconomics* 14(2), 207–42.
- Campbell, D., A. Carruth, A. Dickerson, and F. Green (2007). Job Insecurity and Wages. *The Economic Journal* 117(518), 544–566.
- Chetty, R., J. N. Friedman, S. Leth-Petersen, T. H. Nielsen, and T. Olsen (2014). Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from denmark. *The Quarterly Journal of Economics* 129(3), 1141–1219.
- Conlon, J. J., L. Pilossoph, M. Wiswall, and B. Zafar (2018). Labor market search with

- imperfect information and learning. Technical report, National Bureau of Economic Research.
- De Bruin, W. B., C. F. Manski, G. Topa, and W. van der Klaauw (2011). Measuring consumer uncertainty about future inflation. *Journal of Applied Econometrics* 26(3), 454–478.
- Delavande, A. and S. Rohwedder (2008). Eliciting subjective probabilities in internet surveys. *Public Opinion Quarterly* 72(5), 866–891.
- Dominitz, J. (1998). Earnings expectations, revisions, and realizations. *Review of Economics and Statistics* 80(3), 374–388.
- Dominitz, J. (2001). Estimation of income expectations models using expectations and realization data. *Journal of Econometrics* 102(2), 165–195.
- Dominitz, J. and C. F. Manski (1997). Using expectations data to study subjective income expectations. *Journal of the American statistical Association* 92(439), 855–867.
- Druehl, J. and A. Munk-Nielsen (2020). Higher-order income dynamics with linked regression trees. *The Econometrics Journal* 23(3), S25–S58.
- Epper, T., E. Fehr, H. Fehr-Duda, C. T. Kreiner, D. D. Lassen, S. Leth-Petersen, and G. N. Rasmussen (2020, April). Time discounting and wealth inequality. *American Economic Review* 110(4), 1177–1205.
- Faberman, R. J., A. I. Mueller, A. Şahin, and G. Topa (2022). Job search behavior among the employed and non-employed. *Econometrica* 90(4), 1743–1779.
- Goldstein, D. G. and D. Rothschild (2014). Lay understanding of probability distributions. *Judgment & Decision Making* 9(1).
- Guiso, L., T. Jappelli, and L. Pistaferri (2002). An empirical analysis of earnings and employment risk. *Journal of Business & Economic Statistics* 20(2), 241–253.
- Guvenen, F. (2009). An empirical investigation of labor income processes. *Review of Economic Dynamics* 12(1), 58–79.
- Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021). What do data on millions of us workers reveal about lifecycle earnings dynamics? *Econometrica* 89(5), 2303–2339.
- Guvenen, F., L. Pistaferri, and G. L. Violante (2022). Global trends in income inequality

- and income dynamics: New insights from grid. *Quantitative Economics* 13(4), 1321–1360.
- Halvorsen, E., H. A. Holter, S. Ozkan, and K. Storesletten (2023). Dissecting idiosyncratic earnings risk. *Journal of the European Economic Association*.
- Hendren, N. (2017). Knowledge of Future Job Loss and Implications for Unemployment Insurance. *American Economic Review* 107(7), 1778–1823.
- Hubmer, J. (2018). The job ladder and its implications for earnings risk. *Review of Economic Dynamics* 29, 172–194.
- Hvidberg, K. B., C. T. Kreiner, and S. Stantcheva (2023). Social positions and fairness views on inequality. *The Review of Economic Studies* 86(1), 253–269.
- Jacobson, L. S., R. J. LaLonde, and D. G. Sullivan (1993). Earnings losses of displaced workers. *The American Economic Review*, 685–709.
- Jäger, S., C. Roth, N. Roussille, and B. Schoefer (2022). Worker beliefs about outside options. Technical report, National Bureau of Economic Research.
- Jung, P. and M. Kuhn (2019). Earnings losses and labor mobility over the life cycle. *Journal of the European Economic Association* 17(3), 678–724.
- Karahan, F., S. Ozkan, and J. Song (2022). Anatomy of lifetime earnings inequality: Heterogeneity in job ladder risk vs. human capital. *FRB St. Louis Working Paper* (2022-2).
- Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? evidence from a tax audit experiment in denmark. *Econometrica* 79(3), 651–692.
- Koşar, G. and H. W. van der Klaauw (2023). Workers’ perceptions of earnings growth and employment risk. Working Paper 1056, Federal Reserve Bank of New York.
- Leth-Petersen, S. (2010). Intertemporal consumption and credit constraints: Does total expenditure respond to an exogenous shock to credit? *American Economic Review* 100(3), 1080–1103.
- Leth-Petersen, S. and J. Sæverud (2022). Inequality and dynamics of earnings and disposable income in denmark 1987–2016. *Quantitative Economics* 13(4), 1493–1526.
- Low, H., C. Meghir, and L. Pistaferri (2010). Wage risk and employment risk over the life cycle. *American Economic Review* 100(4), 1432–67.

- Manski, C. F. and J. D. Straub (2000). Worker perceptions of job insecurity in the mid-1990s. *The Journal of Human Resources* 35(3), 447–479.
- Menzio, G. and S. Shi (2011). Efficient search on the job and the business cycle. *Journal of Political Economy* 119(3), 468–510.
- Menzio, G., I. A. Telyukova, and L. Visschers (2016). Directed search over the life cycle. *Review of Economic Dynamics* 19, 38–62.
- Moscarini, G. and F. Postel-Vinay (2013). Stochastic search equilibrium. *Review of Economic Studies* 80(4), 1545–1581.
- Mueller, A. I., J. Spinnewijn, and G. Topa (2021). Job seekers’ perceptions and employment prospects: Heterogeneity, duration dependence, and bias. *American Economic Review* 111(1), 324–63.
- Pistaferri, L. (2001). Superior information, income shocks, and the permanent income hypothesis. *Review of Economics and Statistics* 83(3), 465–476.
- Pistaferri, L. (2003). Anticipated and unanticipated wage changes, wage risk, and intertemporal labor supply. *Journal of Labor Economics* 21(3), 729–754.
- Stephens, M. (2004). Job Loss Expectations , Realizations , and Household Consumption Behavior. *The Review of Economics and Statistics* 86(1), 253–269.
- Stoltenberg, C. A. and A. Uhlendorff (2022). Consumption choices and earnings expectations: Empirical evidence and structural estimation. Working Paper 15443, IZA.
- Topel, R. H. and M. P. Ward (1992). Job mobility and the careers of young men. *The Quarterly Journal of Economics* 107(2), 439–479.
- Von Wachter, T., J. Song, and J. Manchester (2009). Long-term earnings losses due to mass layoffs during the 1982 recession: An analysis using us administrative data from 1974 to 2004. *unpublished paper, Columbia University*.
- Wang, T. (2023). Perceived versus calibrated income risks in heterogeneous-agent consumption models.
- Wiswall, M. and B. Zafar (2021). Human capital investments and expectations about career and family. *Journal of Political Economy* 129(5), 1361–1424.