Online Appendix for "Subjective Earnings Risk"

Andrew Caplin   Victoria Gregory   Eungik Lee

New York University   FRB St.Louis   New York University

Søren Leth-Petersen   Johan Sæverud

University of Copenhagen   University of Copenhagen

March 3, 2023

A   Survey Questionnaire

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

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

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

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

[Part 2-1: Labor income, Past]

Q_2_1. Think about Jan 2020 (last year).

Were you working for pay?

Variable name: Q_2_1

Variable value: 1 Yes 2 No

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

Were you self-employed?

Variable name: Q_2_1_2

Variable value: 1 Yes 2 No

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

Variable name: Q_2_1_3

Variable value: 1 Yes 2 No

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

Variable name: Q_2_1_4

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

[Note: clarification Johan]

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

Variable name: Q_2_2

Variable value: 1 Yes 2 No

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

Please report the most accurate amount you think.

Variable name: Q_2_3

Variable value: [0-]

[Part 2-2: Labor income, Dynamics]

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

Variable name: Q_2_4

Variable value: 1 Yes 2 No

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

Variable name: Q_2_4_2

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

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

Variable name: Q_2_4_2

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

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

Variable name: Q_2_4_4

Variable value: 1 Yes 2 No

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

Variable name: Q_2_4_5

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

[Part 2-2: Labor income, Now]

We now ask about your current employment status.

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

Variable name: Q_2_7

Variable value: 1 Yes 2 No

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

Variable name: Q_2_7_2

Variable value: 1 Yes 2 No

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

Variable name: Q_2_7_3

Variable value: 1 Yes 2 No

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

Variable name: Q_2_7_4

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

[Part 2-3: Labor income, Future]

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

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

Variable name: Q_2_8_1, Q_2_8_2, Q_2_8_3, Q_2_8_4

Variable value: [0-100]

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

Variable name: Q_2_9_1, Q_2_9_2, Bins and balls

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

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

Variable name: Q_2_10_1, Q_2_10_2, Bins and balls

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

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

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

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

Variable name: Q_2_12_1, Q_2_12_2, Bins and balls

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

Think about the possible earned income in 2025.

Think about all the possibilities between reemployment and 2025.

Variable name: Q_2_13_1, Q_2_13_2, Bins and balls

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

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

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

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

Variable name: Q_2_15_1, Q_2_15_2, Bins and balls

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

Think about the possible earned income in 2025.

Think about all the possibilities between reemployment and 2025.

Variable name: Q_2_16_1, Q_2_16_2, Bins and balls

[Other branch]

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

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

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

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

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

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

Please report your lowest and highest possible amount.

Variable name: Q_2_18_1, Q_2_18_2, Bins and balls

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

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

Please report your lowest and highest possible amount.

Variable name: Q_2_19_1, Q_2_19_2, Bins and balls

B Comparison to Administrative Data

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

Table B.1: Survey sample demographics

	Survey		Registry	
Working status	All	Employed	All	Employed
N	14,875	10,942	3,429,071	2,700,069
Female	0.48	0.48	0.5	0.48
Age				
Mean	48.00	47.44	42.50	42.56
S.D.	12.54	11.87	13.26	12.77
Distributions				
20-29	0.11	0.10	0.22	0.21
30-39	0.15	0.16	0.20	0.21
40-49	0.22	0.24	0.22	0.24
50-59	0.31	0.33	0.23	0.24
60-65	0.21	0.17	0.12	0.10
Education				
Above college	0.49	0.53	0.34	0.39
Annual Earnings (unit: DKK)				
Mean	421,278	495,783	327,525	418,624
S.D.	382,719	289,681	430,482	442,410

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

In the analysis we scale all statistics by the relative population weights. To construct population weights, we the Danish population observed in 2020 in the administrative data. We estimate a probit regression with a survey participation dummy as the dependent variable and age, log earnings, education, and gender as explanatory variables. All these

characteristics are available in the administrative data. Table B.2 shows the marginal effect on survey participation using probit regression. Our survey respondents are around 0.37% of the Danish population. We find that the selection into the survey is related to various demographics. For instance, as age increases by one unit, the probability of participation increases by 0.012%.

Table B.2: Marginal effect on participation

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

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

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

C   Supplementary Results

C.1   Reemployment Periods after Quit and Liquid Wealth

In this section we investigate the correlation between the reemployment periods and the relative amount of liquid wealth to disposable income. Liquid wealth is defined as the sum of bank deposits and financial assets including shares, bonds, and assets held in mutual funds summarized by 2019 Dec-31st. Disposable income is all income recorded for 2019 less total taxes paid for the tax year 2019. Using these two variables, we define the liquid ratio variable as

$$\text{liquidity ratio}_{19} = \frac{\text{Liquid asset}_{19}}{\text{Disposable income}_{19}}$$

Figure C.1 shows the correlation between the liquidity ratio and reemployment after quitting. Panel (a) shows the pooled correlation and (b) shows the correlation across three different age groups. We find the positive correlation is robust within each age group.

(a) Pooled   (b) Life cycle
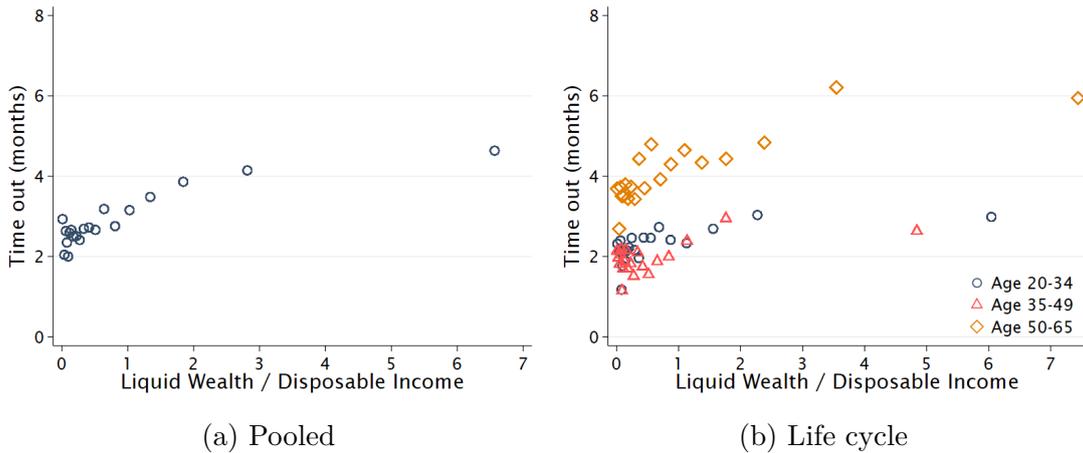
Notes: The figure shows the correlation between the liquidity ratio $\frac{\text{Liquid asset}_{19}}{\text{Disposable income}_{19}}$ and time spent out of work after quitting. The liquidity ratio is calculated from values of disposable income and liquid asset observed for 2019. Panel (a) shows for the pooled case and panel (b) shows for different age groups.

Figure C.1: Liquid wealth ratio and time out of work after quit

We confirm this finding with regressions with various demographics. We additionally control for log earnings, tenure, education, gender, and age. Table C.1 shows the regression results. The dependent variable is $\text{Reemp}_{\text{quit}}$, the expected time out of work following a quit, in all columns. Column (1) is for the pooled sample, column (2) is for the age 20-34 group, column (3) is for the age 35-49 group, and lastly column (4) shows the age 50-65 group. In all cases, except for the youngest subgroup, we confirm the positive

Table C.1: Liquid ratio and time out after quit

Dependent variable:	Time out after Quit (months)			
	(1)	(2)	(3)	(4)
Sample	All	Age 20-34	Age 35-49	Age 50-65
Liquid Ratio	0.325^{***}	0.078	0.213^{***}	0.252^{***}
	(0.032)	(0.071)	(0.067)	(0.055)
Controls	Y	Y	Y	Y
Observation	5,137	1,221	1,957	1,959
R^2	0.113	0.060	0.017	0.227

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

correlation between the liquid ratio and the expected time out of work following a quit.

The results show that workers with relatively higher liquid wealth are likely to spend more time out of work following a quit.

C.2 Life cycle Patterns of survey responses

In this subsection, we examine the life-cycle patterns of the survey responses. We divide the sample into three age groups (20-34, 35-49, 50-65) and re-construct the distributions shown in Figure 4 in the paper. Figure C.2 shows the overview for each age group. We omit the skewness and kurtosis because these moments do not exhibit much variation across ages.

Starting from the probabilities of staying, quitting or being laid-off, we find that younger workers are the least likely to stay with the same employer: they report a 74% chance on average of staying compared to an 86% chance in the 50-65 age group. It turns out that this difference comes from the younger group having the highest probability of quitting, at 20%. We also note that the layoff probability is similar across age groups around 6%. These findings are consistent with the well-known fact that there is more job turn-over among young people. Interestingly, in our survey answers, it appears that workers expect this process to unfold via voluntary quits rather than layoffs. 33% (80%) of respondents report strictly (weakly) higher probability of quitting than being laid-off.

Expected time out of work following a job separation is longest for the oldest age group, age 50-65. People in the middle age group, 35-49, expect to find new jobs the quickest. The result that the average duration is longer for layoffs compared to quits also holds within age groups.

In the next column, we examine the distribution of the mean of earnings growth. In the stay branch, the location of the distribution is similar across age groups, but there is more heterogeneity within the youngest group. For all age groups, most workers expect earnings to grow if they stay with their current employer (most of the mass is to the right of zero).

In the layoff branch, there is a clear life cycle pattern: it appears that the expectation of earnings declines after a layoff is much more prevalent among workers over 35. In contrast, 69% of workers within the youngest group actually expect their earnings to grow after a layoff.

Lastly, the distribution of mean earnings growth rates after quits is mostly above zero for all age groups; however, the youngest group has the highest mean and 90% of respondents among the youngest group expect earnings to grow.

Finally, the $p_{90}-p_{10}$ column summarizes uncertainty for each age group in each branch. On the stay branch, the two oldest groups are most certain, whereas the youngest group exhibits more heterogeneity. The opposite pattern appears on the layoff branch. The youngest group of workers is the least uncertain on average, and the two oldest groups have more respondents who report higher levels of uncertainty. The patterns in the quit

branch reveal little differences in uncertainty across age groups.
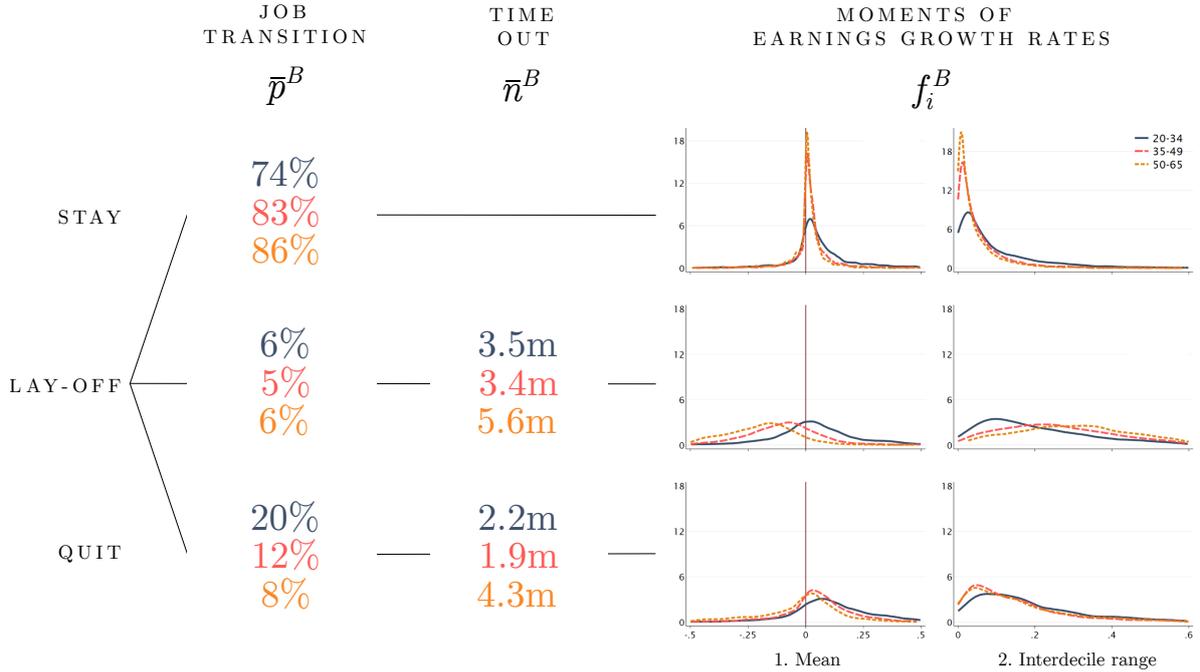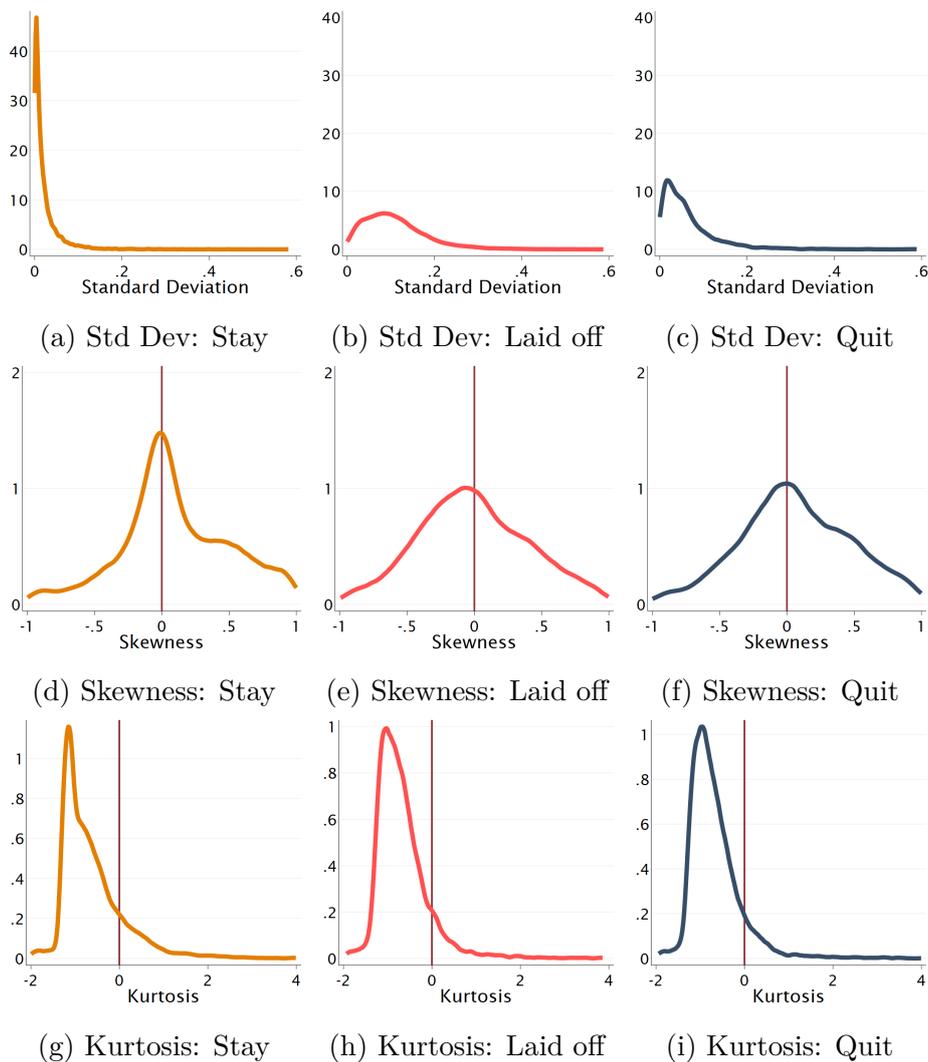

Note: The figure shows answers to the questions in the conditional survey instrument, where the rows correspond to the branches "Stay", "Layoff", and "Quit". The first column shows the average probabilities of each branch, $\bar{p}^B$. The second column shows the average of the expected reemployment period in each branch, $\bar{n}^B$, in months. The distributions show the cross-sectional distribution of the 1st to 4th moment of the subjective conditional earnings distributions, f_i^B. We measure the 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 $K_{CS} = \frac{(p_{97.5}-p_{2.5})}{(p_{75}-p_{25})} - 2.91$. Colors denote age group - Blue: age 20-34, Red: age 35-49, Orange: 50-65. Survey results are weighted using population weights.

Figure C.2: Overview of branch-by-branch survey responses: by age group

C.3 Standard Measures of Higher Order Moments

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

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

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

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

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

Figure C.3: Overview of branch-by-branch survey responses: using standard moments

C.4   Job Transitions

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


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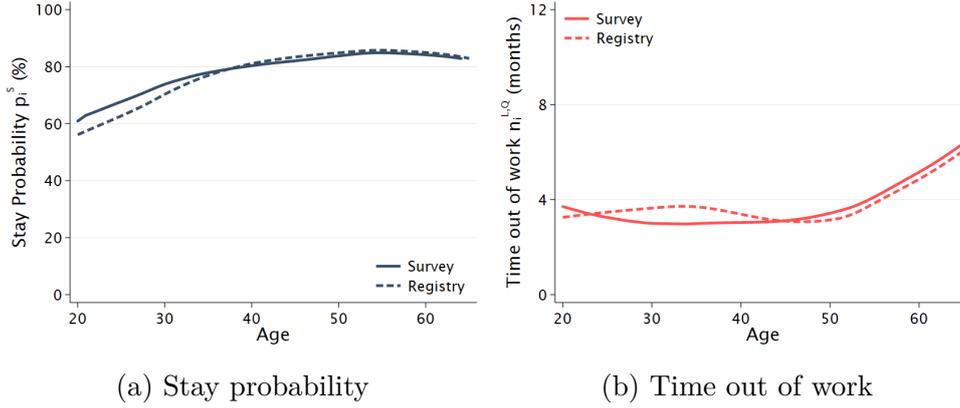
Note: The figure shows how two components from the survey (solid line) and the administrative data (dashed line) compare over the life-cycle. Panel (a): In the survey, we directly ask about the probability of staying with the same employer for the next 12 months. In the registry, we compute the proportion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

Figure C.4: Job separations and time out of work in the survey and in the administrative data for 2019

C.5 Last Year's Earnings

In this section, we compare the subjects' reported earnings for 2020 to actual earnings in 2020 as recorded in the income-tax registry. Figure C.5 shows a binned scatter plot of average earnings reported in the survey (Y-axis) by bins of earnings recorded in the administrative data (X-axis), which is third-party reported by employers directly to the Danish Tax Agency. We find that earnings reported in the survey is very similar to earnings recorded in the administrative data. This suggests that respondents remember their earnings during 2020 well at the point when we surveyed them in January 2021.
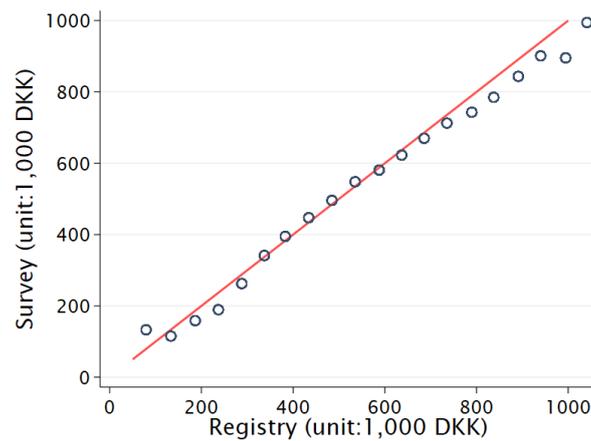

Note: Binned scatter-plot comparing earnings observed in the administrative data and self-reported earnings.

Figure C.5: Self-reported and registry earnings: last year (2020)

C.6 Comparison with 2019 Registry

In this section, we replicate our main registry finding using the 2019 data from the administrative registry.

We first replicate Figure 5 from the paper, which illustrates the log density of annual earnings growth by age groups. The result is presented in Figure C.6, which shows survey and registry results side by side. The figure is very similar to Figure 5 in the main text.


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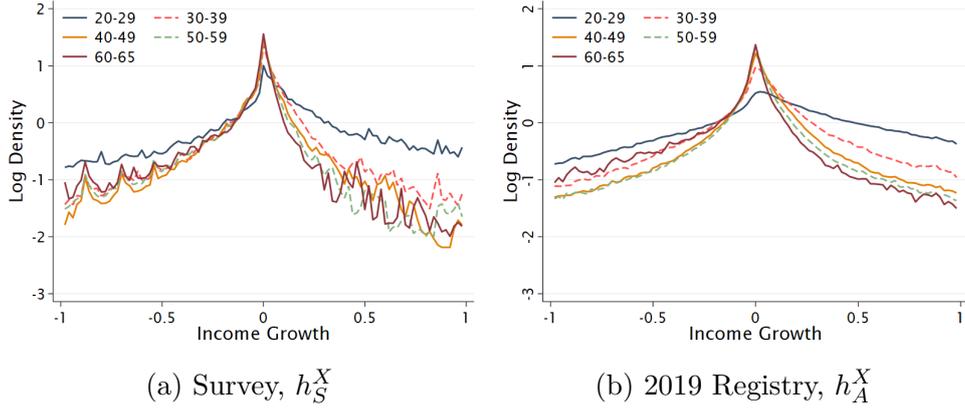
Note:
Panel (a) plots log density for the pooled distribution of expected holistic earnings growth rates from the survey, h_S^X, where X indicates partitions by age groups. (b) plots the distribution of annual earnings growth from 2018 to 2019 as observed in the administrative data for the full population,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

Figure C.6: Pooled earnings risk and registry earnings risk (data from 2019)

C.7 Registry Earnings Risk using Survey Respondents

In this section, we replicate Figure 5(b) using only survey respondents and their 2020 earnings growth. Figure C.7 shows the results. While noisy because it includes fewer observations compared to Figure 5(b) in the paper, it compares well to Figure 5(b).
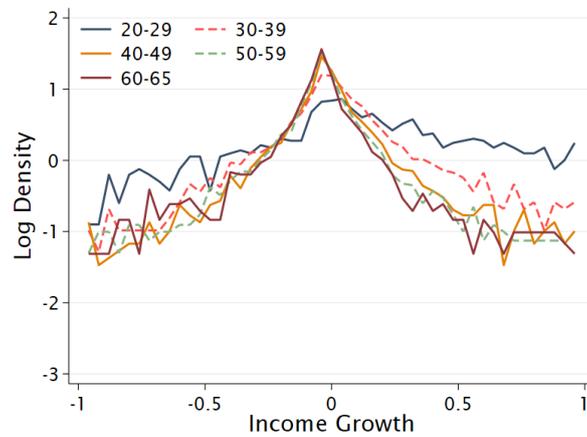

Note: The figure plots the distribution (log density) of annual earnings growth for 2020 as observed in the administrative data for the survey participants.

Figure C.7: Registry earnings risk (survey respondents only)

C.8 Standard Moments of h_S^X and h_A^X

In this section, we show the pooled moments of standard measures for standard deviation, skewness, and kurtosis. The below Figure C.8 shows the changes in pooled moments across life cycle. In the standard measures, we again find the moments are well aligned across the life cycle.
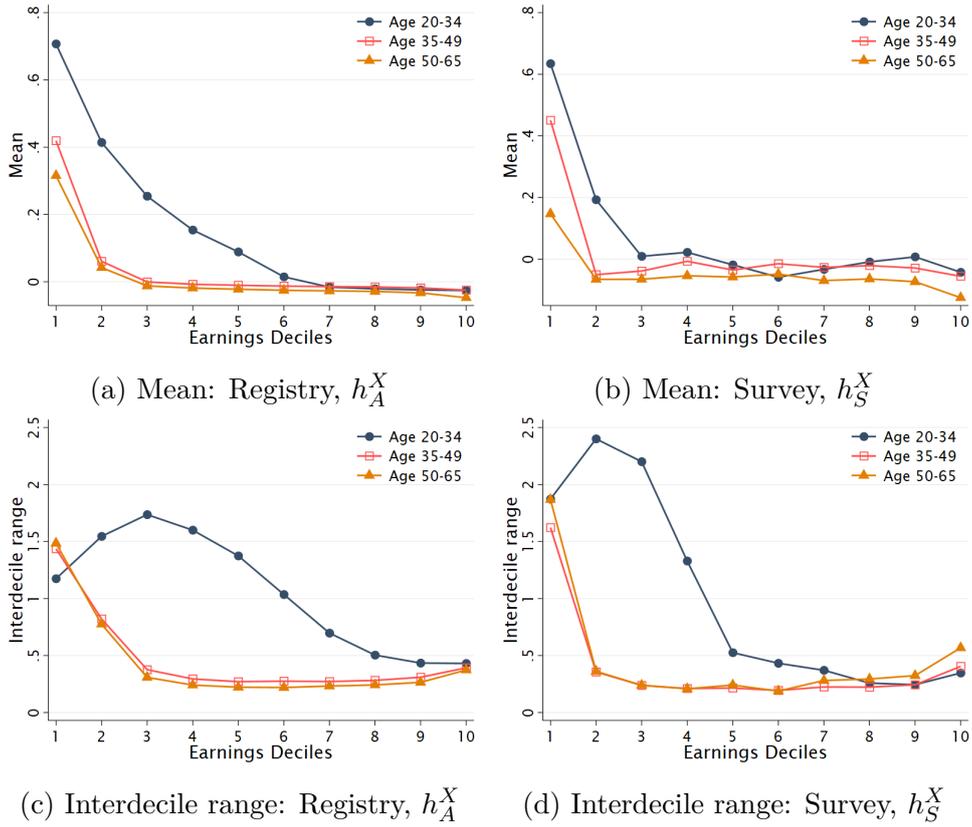

(a) Standard Deviation   (b) Skewness   (c) Kurtosis
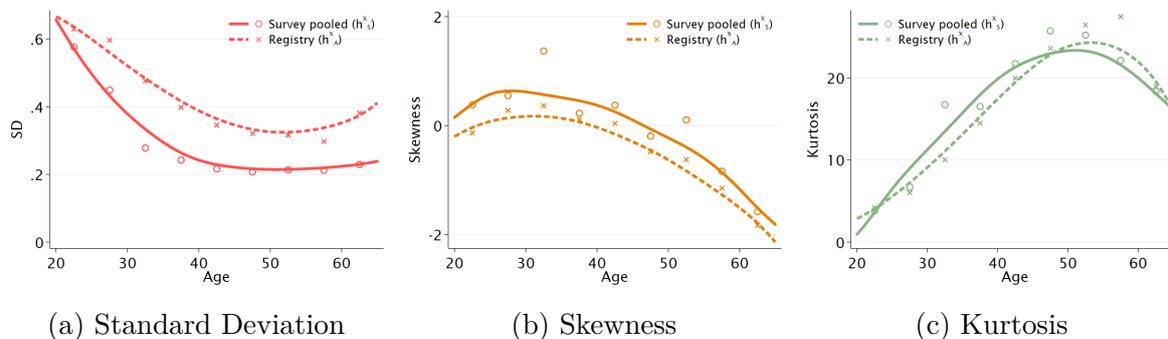
Note: The figure shows the average value of the 1st to 4th moments of the pooled earnings distribution by age in the survey, h_S^X, and in the administrative data, h_A^X. "o" and "x" represent the empirical mean across 5-years age bins for the survey and the administrative data for 2020, respectively.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

Figure C.8: Life cycle patterns in pooled risks moments

C.9 Comparing Earnings Growth across Earnings levels

In this section, we compare how the moments of the distribution of earnings growth in the survey and the administrative data evolve across the life cycle and earnings distribution. Specifically, we divide the data into cells consisting of three broad age groups and deciles of the distribution of earnings growth within these broad age groups. Again, for the survey, we consider the distribution of pooled subjective earnings growth expectations. The result is shown in Figure C.9. Panel (a) and (b) shows how the mean earnings growth evolves across the distribution of earnings levels. We find that earnings growth is generally highest at the lower end of the earnings distribution and that positive earnings growth extends higher up in the distribution for younger workers than for workers aged 35 or more. Importantly, the broad features are similar between the distributions constructed from the survey and the administrative data. Panel (c) and (d) show how $p_{90} - p_{10}$ develop across earnings deciles by age groups. Young workers have the highest level of uncertainty and this is practically the case throughout the earnings growth distribution, but it is most pronounced in the lower half of the distribution of earnings. However, the most important insight is that the patterns are remarkably similar across the survey and the registry data.

Overall, we find that the cross-section distribution of pooled earnings risk and the distribution of realized earnings growth in the administrative data are broadly similar, even when we compare the two distributions by detailed groups of age and earnings levels.

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

(c) Interdecile range: Registry, h_A^X (d) Interdecile range: Survey, h_S^X

Note: The figure shows the first two moments of the pooled earnings distribution in the survey, h_S^X, by earnings deciles and coarse age groups. It also plots the corresponding measures calculated from the administrative data, h_A^X.

Figure C.9: Mean and interdecile range across age and earnings level

21

C.10   Standard Moments of Subjective Earnings Risks

In this subsection, we show the original moments (standard deviations, skewness, and kurtosis) subjective earnings risk. First, the standard deviation is consistent with the $p_{90} - p_{10}$ which shows higher degree of spread for younger workers. Second, the skewness is also consistent with Kelley skewness in the main text and show a consistent negative skewed pattern. Lastly, kurtosis shows a slightly different pattern when compared to the Crow Siddiqui kurtosis. Especially, older workers exhibit less kurtosis according to the standard measure than according to the Crow Siddiqui measure of kurtosis.

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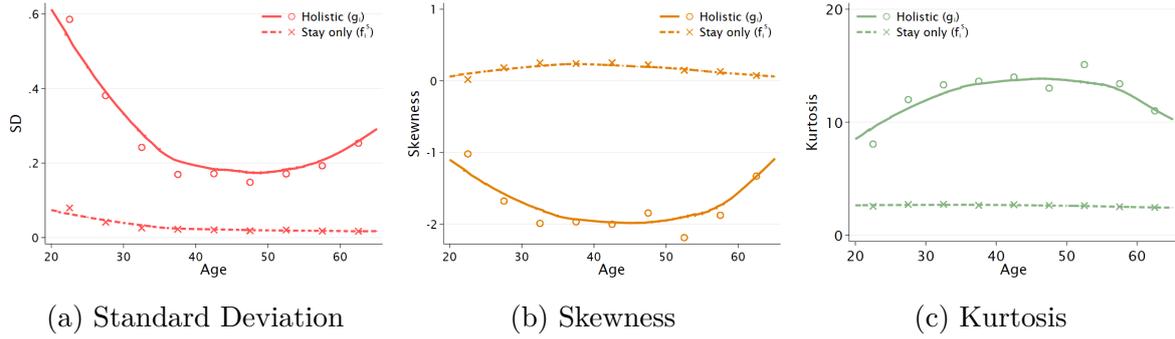
Note: The figure shows the average value of the 2nd to 4th standard moments over the life cycle of holistic earnings risk, g_i, and risk conditional on staying, f_i^S. "o" and "x" represent the empirical mean across 5-years age bins. The lines are local linear polynomials, calculated using a bandwidth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

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

D   Method for Simulating Holistic Earnings Risks

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

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

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

1. Simulating job transitions, $\mathbb{1}[Q_i = 1]$, $\mathbb{1}[L_i = 1]$: We simulate 20,000 job transition events for each individual based on the stated job transition probabilities (stay, laid off, and quit) in 2021.

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

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

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

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

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

   2. Then we simulate the timing of reemployment. We do this using the stated probabilities of being reemployed after 1, 3, and 12 months after the job separation. To do this, we linearly interpolate the probability of reemployment

over months and then construct the monthly reemployment distribution. Using this monthly distribution of reemployment, we simulate the duration of the intermediate job search period after the separation. We assume that the reemployment happens at the beginning of the month.

3. Simulating conditional earnings distributions, $\hat{f}_i^S$, $\hat{f}_i^Q$, $\hat{f}_i^L$

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

4. Weight together the components from steps 1-3

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

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

---

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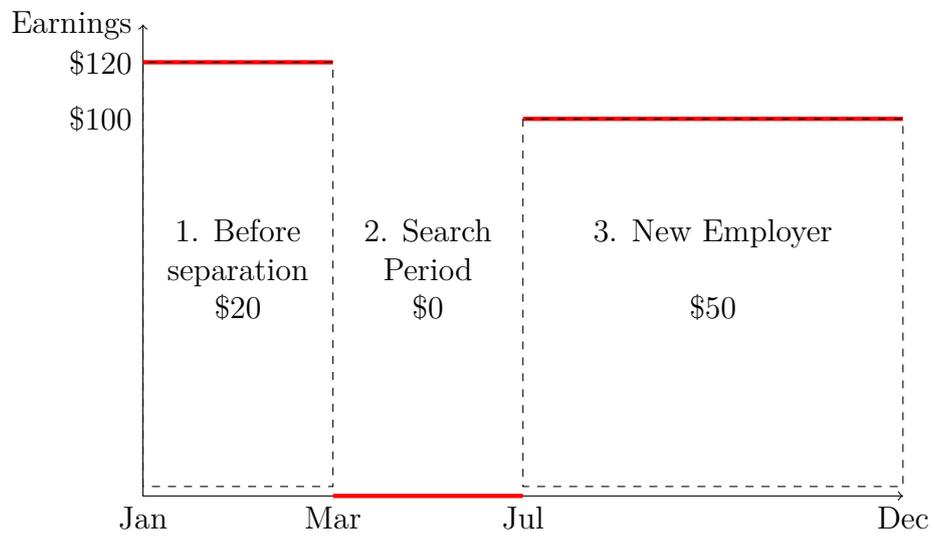
Figure D.1: Holistic earnings risk simulation

E Moments of the Pooled Distribution

## E.1 Age Group and Current Earnings Variations

In this section, we report figures corresponding to Figure 8 in the paper for all the groups defined in section 4.1. We note again that we divide the sample into two earnings groups, above-median (high) and below-median (low), and three broad age groups (20-34, 35-49, 50-65). Figure E.1 shows the results. The number on the label shows their age group and High/Low means their earnings level in the age group. Across all age groups and earnings levels, we find a pattern consistent with the pattern presented in Figure 8 in the paper. We confirm that the administrative based measure, h_X^A, and survey based measure, h_X^S, are well aligned. On the other hand, the subjective measure, g_i, is heterogeneous while being centered around 0.

(a) 20-34:*High* (b) 35-49:*Low* (c) 35-49:*High* (d) 50-65:*Low*
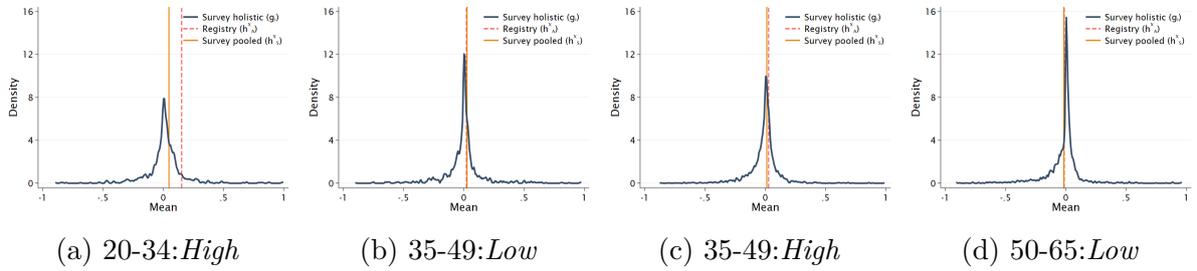
Note: Estimates of the mean for h_S^X, h_A^X, and the distribution of g_i for four subgroups in the data.

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

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

(a) 20-34:*High* (b) 35-49:*Low* (c) 35-49:*High* (d) 50-65:*Low*
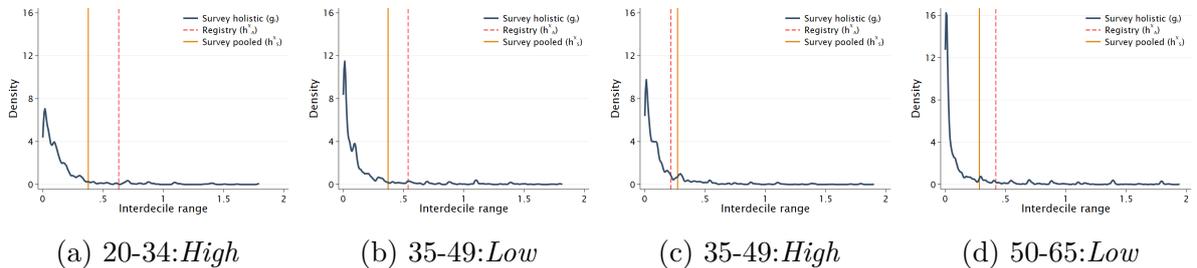
Note: Estimates of the interdecile range for h_S^X, h_A^X, and the distribution of g_i for four subgroups in the data.

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

E.2 Derivations

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

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

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

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

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

The skewness (the centralized third moment) is given by

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

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

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

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

Kurtosis (the centralized fourth moment) is given by

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

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

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

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

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

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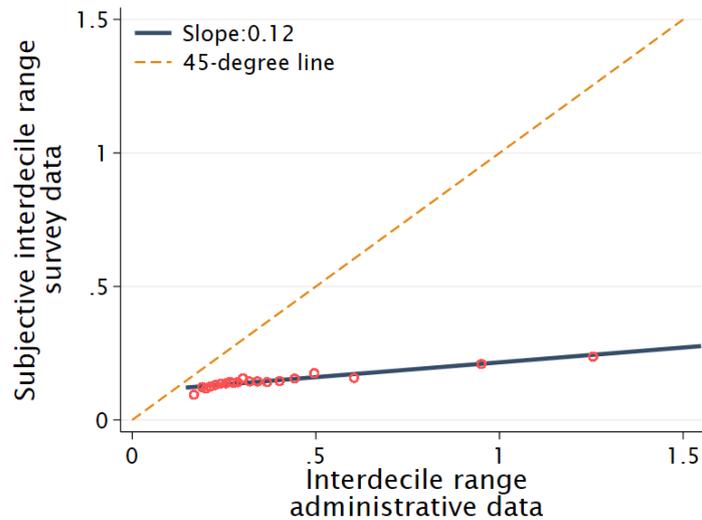

(a) Skewness   (b) Kurtosis
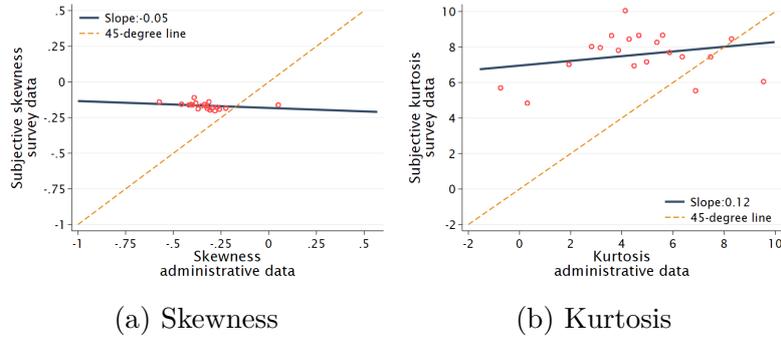
Note: The figure compares average skewness and kurtosis of subjective holistic income expectations, g_i, to those calculated from administrative data. Both are calculated within 300 cells divided by age group and within age group earnings percentiles, h_A^X, i.e, the same partition applied in the construction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

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

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

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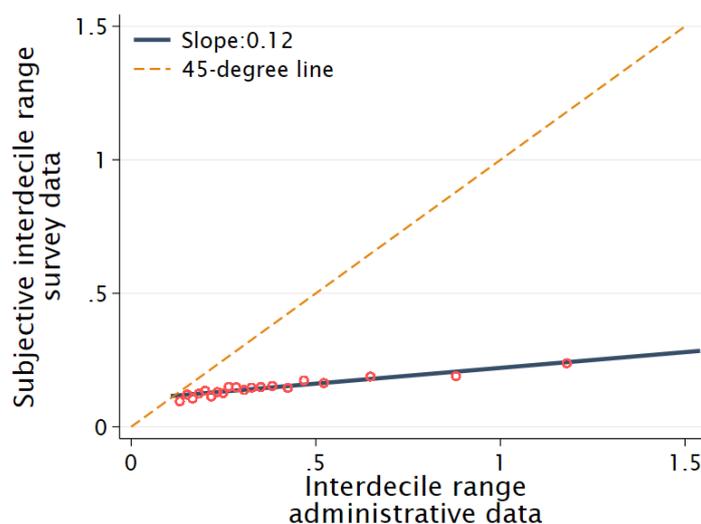

Note: The figure compares average interdecile range, $p_{90} - p_{10}$, of subjective holistic earnings expectations, g_i, to interdecile range calculated from administrative data. Both are calculated within 1,800 cells divided by age, gender, education, and earnings deciles, h_A^X. The panel shows a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

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

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

The literature examining heterogeneous income processes (HIP) assumes that individual-level heterogeneity in earnings growth is present. To address this issue we calculate average individual level earnings growth over the last 5 years (2014-2019). We then divide the sample into cells by age, gender, education, current earnings quintiles, and last 5 years' average earnings growth quintiles, resulting in 4,500 cells over which we conduct an exercise similar to the one presented in Figure 9 in the paper. Again, the results are practically identical to the results presented in Figure E.5.

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

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

F   Search Model: Details and Calibration

F.1   Value Functions

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

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

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

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

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

where

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

The flow value is just the match output (the firm's profit is $zg(y) - w$, but the worker earns w). Next period, the worker gains 1 unit of experience y, but the worker and firm separate with probability d, which is an optimally chosen policy (discussed further in the next section). If they separate, the worker goes to unemployment and gets value $U_{t+1}(y+1, \psi')$. Otherwise, with probability λ_e, the worker gets the opportunity to search

on the job, with value encompassed in $S_{t+1}(z, y+1, \psi')$. In this case, the worker (like the unemployed worker above) chooses the optimal submarket x to search in which maximizes the probability they find a job there times the additional value they get from the new job relative to their old one. If the worker's search is unsuccessful or the worker does not get the opportunity to search, they get $\mathbb{E}_{z'} V_{t+1}(z', y+1, \psi')$ and stay with their existing job. In this case, the match quality may be reset to z_0. This happens with probability η and the worker becomes a match of unknown quality with value $V_{t+1}(z_0, y+1, \psi')$. Otherwise, with probability $1 - \eta$, the match continues as is, $V_{t+1}(z, y+1, \psi')$.

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

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

$$(11)$$

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

F.2 Equilibrium

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

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

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

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

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

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

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

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

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

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

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

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

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

Everything is now in place to define the equilibrium.

Definition: A *Block Recursive Equilibrium* consists of:

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

- Policy functions that determine which submarkets the employed search in and the match quality thresholds for destruction of the match

- A market tightness $\theta_t(x, y, \psi)$ (ratio of vacancies to unemployment) for each submarket

such that

- The value, policy, and market tightness functions do not depend on the aggregate state ψ

- The market tightnesses satisfy (12)

- The policy functions solve (9), (10), and (11)

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

F.3 Calibration Details

F.3.1 The Registry Data

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

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

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

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

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

considering workers who are employed. The proportion of the sample who reported more than 30,000 DKK in self-employment income is around 4% of the population in 2020.

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

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

F.3.2 Calibration Strategy and Targeted Moments

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

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

---

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

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

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

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

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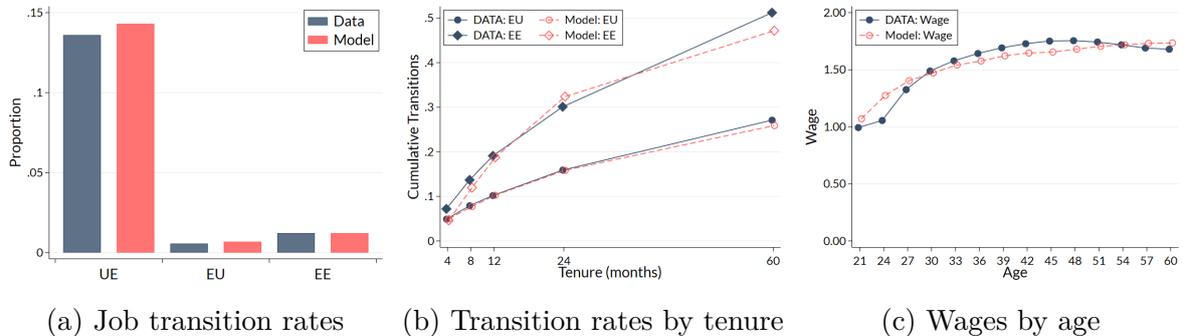

(a) Job transition rates (b) Transition rates by tenure (c) Wages by age

Figure F.1: Targeted moments

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

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

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

Table F.1: Calibrated parameters

|	Description	Value
β	discount factor (externally set)	0.9967
b	home production	2.946
k	vacancy cost	25.579
λ_u	job search probability: unemployed	1
λ_e	job search probability: employed	0.719
δ	exogenous destruction probability	0.0007
σ	scale of match quality distribution	16.311
ϕ	shape of match quality distribution	3.464
α	probability of match quality discovery	0.124
η	probability of match quality resetting	0.025
ρ_1	human capital accumulation: level	5.434
ρ_2	human capital accumulation: curvature	0.095

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

F.3.3 Untargeted Moments

Next, we present some of the untargeted moments produced by the model. We focus on the ones that are closest to moments we are most interested from the survey and the registry: the life-cycle patterns of job transition rates and the distribution of annual earnings growth. As explained in the main part of the paper, we equate EEs in the model and register with quits in the survey and EUs in the model and register with layoffs in the survey.

Figure F.2 compares the life-cycle patterns of both transition rates in the survey, registry, and model. Even though these were not targeted, the model does a remarkable job at matching the registry. 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 – this says that on average and within age groups, people are correct about the chances of undergoing either one of the transition types.

Figure F.3 compares the patterns in time-to-reemployment after a layoff (in the survey) or the length of the U spell in an EUE transition (in the registry and model). Each line shows the probability of being reemployed within 1, 3, 12, or 24 months as a function of age. Again, the model matches the registry very well, despite not being targeted. The
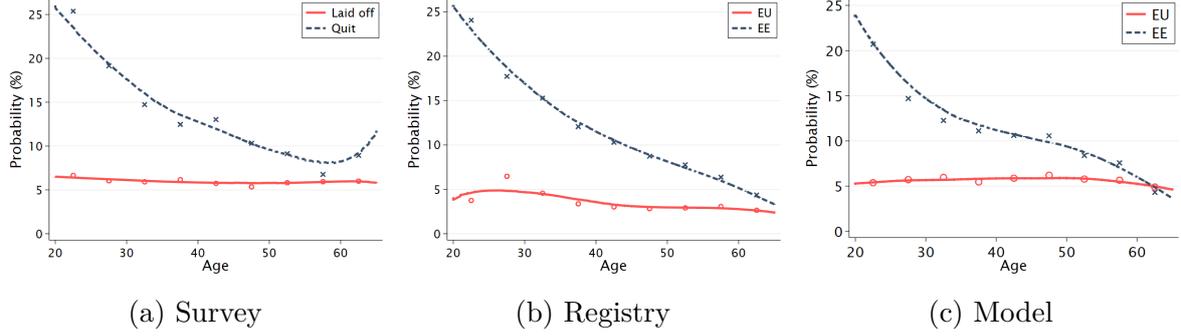
(a) Survey   (b) Registry   (c) Model

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

Figure F.2: Job transition probabilities

pattern in the survey is also quite close, except for those above age 55 or so.
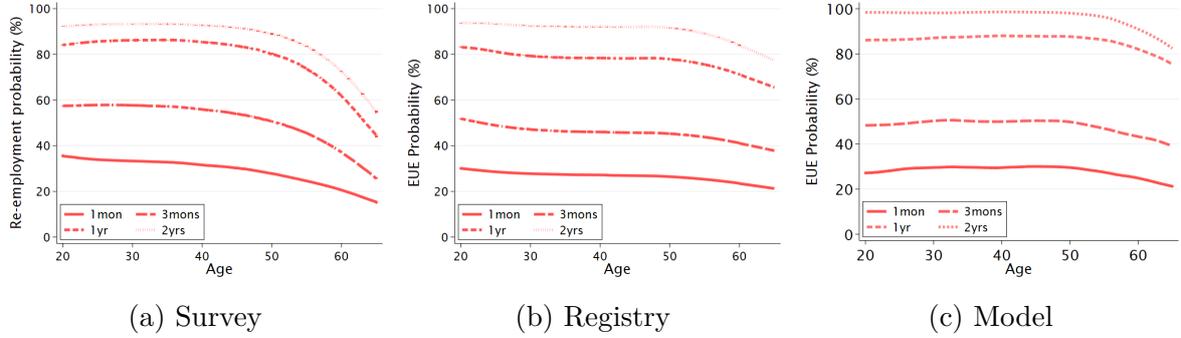
(a) Survey   (b) Registry   (c) Model

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

Figure F.3: Reemployment probabilities

Lastly, Figure F.4 compares the densities of annual earnings growth in the survey, registry, and model. Again, the distribution generated by the model is similar to that of the registry.

F.4   Description of Belief Simulations in the Model

This section describes the details on how we generate the beliefs of agents in the model with a structure that is the same as the survey beliefs we collected.

We start by drawing a sample of 100,000 workers from the stationary distribution. For each worker i in this sample, we will create the model counterparts of $p_i^S, p_i^L, p_i^Q, n_i^L, n_i^Q, f_i^S, f_i^L$, and f_i^Q. However, note that since we equate EEs in the model to quits in the survey, $n_i^Q = 0$ in the model for all i.

1. We interpret each worker's initial state (z, y, t) to represent their "current job" and

| (a) Survey | (b) Registry | (c) Model |
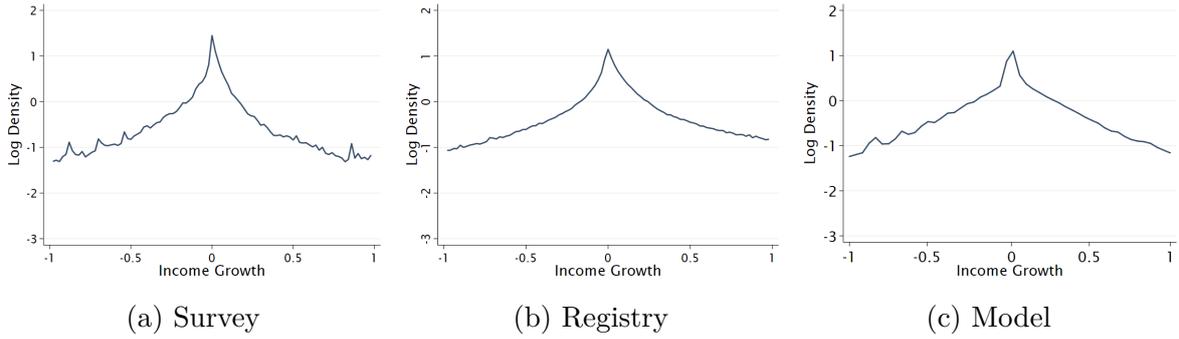
Note: Population weights are used in Panel (a).

Figure F.4: Distributions an annual log earnings growth

its associated earnings, like we do when we simulate out of the survey responses. From there, we draw further sets of simulations for each worker to recover each component of their "survey responses."

2. **Branching probabilities: probability of stay (p_i^S), EU (p_i^L), and EE (p_i^Q):** Given each workers' inital state, we simulate a series of 12-month paths many times and then count the proportion of scenarios in which workers stay with the same employer (p_i^S), make an EU transition (p_i^L), and make an EE transition (p_i^Q). In the simulation, EUs occur with either exogenous job destruction (through δ) or through a change in match quality (it becomes known and is below the threshold for keeping the match, or it is reset and the match is not worth keeping. EEs can occur if the worker successfully searches on-the-job, which is more likely to happen for matches with lower quality.

3. **Stay branch conditional earnings f_i^S:** We simulate a set of scenarios for each worker in which they stay at their current employer for 12 more months. We do not allow them to get exogenous job destruction shocks or search on-the-job ($\delta = 0$ and $\lambda_1 = 0$). The only source of risk is changes in match quality z: it can still go from unknown to known and known to unknown at the same rates and with the same distribution as the calibrated model. In each simulation, we calculate their average monthly earnings over the 12 months, or if they leave earlier because of a match quality shock, the average monthly earnings for the time they are still there. Taking the log difference with their monthly earnings in their original state and aggregating over the simulations gives the distribution, f_i^S.

4. **Layoff branch conditional earnings, f_i^L, and time out of work after a layoff, n_i^L:** These are done together in the same simulation, with a length of 3

40

years so workers have time to find a new job and spend some time there. We separate each worker from their existing job ($\delta = 1$ in the first period) and simulate multiple paths forward as they remain unemployed and eventually find a new job. We count up how many months it took them to find their new job. The worker's n_i^L is the average of this over the simulations. Workers will have different job-finding rates depending on their experience and age. Then we use their earnings at their new employer. Since all of these jobs are new jobs, they all start off with unknown match quality, and from there it evolves in the same way as it does in the calibrated model. Like on the stay branch, we calculate the average monthly earnings in this new job for up to 12 months. Taking the log difference with their monthly earnings in their original state and aggregating over the simulations gives the distribution, f_i^L.

5. **Quit branch conditional earnings** f_i^Q: We assign workers to a new job immediately and simulate up to 12 months with the employer. There are two key differences compared to the layoff branch. First, we do not model time out of work, as mentioned above. Second, we only simulate this branch for workers whose on-the-job search policy has them applying to submarkets with positive probability of finding a job, i.e., they find it optimal to search on-the-job in the first place (workers with high enough match quality do not). This matches the structure of the survey: if someone reports a 0% chance of quitting, they are not asked follow-up questions on that branch. Like on the layoff branch above, all of these new jobs have unknown quality initially and from there it evolves in the same was an in the calibrated model. Like on the stay and laid-off branches, we calculate the average monthly earnings in this new job for up to 12 months and repeat this process many times for each worker. Taking the log difference with their monthly earnings in their original state and aggregating over the simulations gives the distribution, f_i^Q.

References

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

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

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

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