Subjective Earnings and Employment Dynamics

Manuel Arellano, Orazio Attanasio, Margherita Borella,

Mariacristina De Nardi, and Gonzalo Paz-Pardo*

August 27, 2025

Abstract

We develop a new approach to estimating earnings and job and employment dynamics using subjective expectations data from the Federal Reserve Bank of New York's Survey of Consumer Expectations. These data provide beliefs about future earnings offers and acceptance probabilities, offering direct information on counterfactual outcomes and enabling identification under weaker assumptions. Our framework avoids biases from selection and unobserved heterogeneity that affect models using realized outcomes. First-step fixed-effects regressions identify risk, persistence, and transition effects; second-step GMM recovers the covariance structure of unobserved heterogeneities such as ability, mobility, and match quality. We find lower risk and persistence than in prior work, but greater heterogeneity in ability and match quality. Simulations show that reduced-form estimates overstate persistence and volatility on individual-level productivity due to job transitions and sorting. After accounting for heterogeneity, volatility declines and becomes flat across the earnings distribution. These results underscore the value of expectations data for studying labor market outcomes.

JEL Codes: C23, C81, D15.
Keywords: Subjective expectations, earnings dynamics models.

*Arellano: CEMFI, Madrid; Attanasio: Yale University and NBER; Borella: Università di Torino; De Nardi: University of Minnesota, Federal Reserve Bank of Minneapolis, CEPR and NBER; Paz-Pardo: European Central Bank. The opinions and conclusions expressed are solely those of the authors and do not represent the opinions or policy of the European Central Bank, the Eurosystem, any agency of the Federal Government, NBER, CEPR, the Federal Reserve Bank of Minneapolis, or the Federal Reserve System.

1 Introduction

This paper revisits long-standing questions in income dynamics: how to measure risk, heterogeneity, and persistence in earnings and employment, and how job transitions contribute to these dynamics.[1] Our key innovation is the use of subjective expectations data from the Federal Reserve Bank of New York's Survey of Consumer Expectations (SCE) to address these questions. These data—which include information not only about expected future earnings but also beliefs about earnings potential offers and the probability of accepting them—open new identification strategies and estimation methods that overcome long-recognized limitations in standard approaches.

The standard approach using observed data suffers from identification and estimation difficulties. First, estimates of earnings risk and persistence may be biased by unobserved heterogeneity: differences in ability, job match quality, or preferences that are not separately identified from stochastic variation. Second, selection into employment and jobs creates identification challenges, because earnings are observed only when individuals are employed. If selection depends on unobserved shocks, then standard estimators confound risk with endogenous sorting. Third, realized outcomes—particularly discrete ones—often need to be observed over long panels in order to separately identify latent factors like ability and match quality from transitory noise. This requirement poses practical limitations, particularly when the available data span only a few periods per individual.

We show that a specific kind of subjective expectations data can help address all three problems. In particular, we use information on the probability distribution of potential offers individual might receive to inform the latent earnings distribution, and subjective probabilities of accepting offers to proxy the selection mechanism. Because individuals report their beliefs about earnings and choices under counterfactual scenarios, these data offer direct information on both the latent outcome process and the decision rules that inform selection choices. This information allows us to achieve identification even with short panels, leveraging rich cross-sectional variation in expectations rather than requiring long individual earnings histories. These data allow us to identify and

[1] For surveys of the income dynamics literature, see Meghir and Pistaferri (2011), Arellano (2014) or Altonji et al. (2023)

estimate a structural model of earnings, employment, and job transitions—closely related to Altonji et al. (2013)—under weaker assumptions than would be required when using realized outcomes alone.

Our model builds on the "standard selection framework," which combines a latent outcome process (earnings) and a selection mechanism (employment and job transitions) that depends on latent variables. We associate the subjective earnings offer distribution with the latent outcome process and the acceptance probabilities with the conditional selection rule. This structure allows us to estimate the model using methods that are both transparent and tractable.

In our framework, the residuals from subjective expectations equations reflect measurement error—such as reporting noise—and are orthogonal to the underlying shocks that drive behavior. When subjective probabilities about future outcomes are elicited, those outcomes have not yet materialized. As a result, the subjective probabilities reflect the stochastic properties of future shocks (as in Arellano et al. (2024)), while the realized shocks are absent from the subjective variables we model. Because individuals report their own beliefs about future outcomes—including counterfactual scenarios—the model can condition directly on their information set. This setup avoids selection bias: the residuals do not correlate with unobserved determinants of choices or outcomes.

By contrast, in observational data, selection bias arises when unobserved shocks influence both the outcome and the decision to be observed. For example, a shock to productivity may affect both earnings and the probability of employment. If these shocks are omitted from the regression, the residual becomes correlated with the selection mechanism, leading to biased estimates. This distinction is central to our identification strategy: subjective expectations allow us to observe the information that typically drives selection, avoiding the omitted variable problem that plagues purely observational analyses.

We estimate the model in two simple steps. First, we use fixed-effects regressions to estimate risk, persistence, and the effect of offered wages on transitions, while isolating reduced-form unobserved heterogeneity. This avoids the nonlinear estimation challenges typically encountered in models estimated on realized outcomes. Second, we impose a covariance structure on the residuals from the first step and estimate the structural

components of heterogeneity—ability, mobility, and match quality—using GMM. The second step is likewise computationally straightforward.

Our estimates yield several findings that highlight the value of incorporating subjective expectations into earnings dynamics. First, relative to Altonji et al. (2013), who estimate their model using PSID data, we find substantially less risk and lower persistence in productivity and match-specific shocks. In contrast, we uncover more heterogeneity, suggesting that individual and job-match differences play a larger role than previously estimated.

Second, we find strong positive effects of offered earnings on employment transitions. Higher offered earnings significantly increase both the probability of transitioning out of unemployment and the likelihood of changing jobs among the employed. Third, we estimate that expected earnings in the current job have a positive but smaller and less precisely estimated effect on the probability of remaining in that job.

Finally, we relate our findings to the existing literature. Our model and estimation allow us to understand the determinants of the high earnings persistence that we estimate in both observed and subjective expectations data. Empirically, we find that the high earnings persistence that we observe comes from lower risk and persistence of shocks than standard reduced-form estimates suggest, but more heterogeneity. We show that standard high persistence and volatility estimates arise from unaccounted employer transitions and heterogeneity, rather than persistent productivity shocks. Similarly, we find that much of the nonlinear relationship between earnings volatility and income disappears after accounting for individual- and employer-match effects. These results underscore the value of the combination of expectations data in combination with a rich model of employment dynamics for analyzing labor market outcomes.

We also relate our findings to the literature on risk and nonlinear earnings dynamics. In our simulated data, the standard deviation of reduced-form earnings volatility averages 7% and varies systematically with earnings, in line with previous studies on risky and non-linear earnings dynamics. However, once we remove the influence of person- and match-specific heterogeneity, volatility drops to 4.5% and becomes approximately constant across the earnings distribution. This indicates that much of the observed

nonlinearity in earnings volatility reflects unobserved heterogeneity and job transitions, rather than individual productivity risk.

These results suggest that subjective expectations data provide new insights into labor market dynamics. Our paper is not the first to use such data to estimate a well-specified economic model. Faberman et al. (2022), in a seminal contribution, use the same dataset to estimate a search model and assess how search effort translates into outcomes for both the employed and unemployed. Our contribution focuses on earnings dynamics and shows how subjective expectations data and the *experimentation within the survey* aid identification under weaker assumptions and enable transparent estimation of models. Such models are difficult to identify with realized outcomes alone.

Our contributions are fourfold. First, we develop econometric techniques linking subjective data to structural models, allowing for straightforward parameter estimation through linear regressions. By relying on simple linear fixed-effects methods, we are able to estimate a nonlinear dynamic model with unobserved heterogeneity, a task that is challenging when only using realized outcome data. Second, we demonstrate that subjective expectations data enable identification of models with employment and job risk under weaker assumptions that are not driven by functional form. Third, we show that these data allow us to estimate such models from short panels, avoiding the need to track long individual histories or impose restrictive functional form assumptions on unobserved heterogeneity. Fourth, we show that perceived income uncertainty is significantly smaller than traditionally estimated risks, while heterogeneity plays a larger role in shaping observed outcomes.

While our framework leverages the tractability of linear fixed-effects methods, we highlight that the SCE data contains additional information that could be exploited using nonlinear estimation techniques. This suggests that subjective expectations data offer opportunities to further enhance model identification and estimation, unlocking even richer insights into labor market dynamics.

Table 1: SCE data, sample summary statistics.

	Women			Men		
	N	Mean	Std.Dev.	N	Mean	Std.Dev.
Age	4796	42.57	10.33	4552	43.17	10.10
White	4796	0.78	0.41	4552	0.85	0.36
Couple	4483	0.59	0.49	4240	0.74	0.44
HS	4796	0.09	0.28	4552	0.09	0.29
Some college	4796	0.33	0.47	4552	0.27	0.45
College degree	4796	0.33	0.47	4552	0.38	0.48
Post-graduate education	4796	0.25	0.43	4552	0.26	0.44
Unemployed	4796	0.06	0.24	4552	0.04	0.19
Workers						
Prob. losing job next 4 mo.	4421	12.77	18.36	4310	13.10	18.16
Current earnings	4243	57.38	113.46	4171	85.30	75.09
Expected offer in 4 mo.	3003	59.20	46.83	3062	89.53	79.34
Prob. receiving offer next 4 mo.	4133	0.22	0.28	3997	0.22	0.27
Prob. employed in 4 mo.	4190	0.97	0.11	4049	0.97	0.11
Prob. change job in 4 mo.	4472	0.09	0.18	4234	0.08	0.17
Prob. unemployed in 4 mo.	4190	0.03	0.11	4049	0.03	0.11
Unemployed/laid off						
Previous earnings	291	30.80	48.87	192	51.77	53.02
Expected offer in 4 mo.	257	52.44	219.32	173	49.22	45.17
Prob. receiving offer next 4 mo.	279	0.53	0.34	184	0.56	0.33
Prob. employed in 4 mo.	276	0.52	0.34	181	0.53	0.34

2 Data: The Survey of Consumer Expectations

We use data from the Survey of Consumer Expectations, a monthly internet-based survey administered by the Federal Reserve Bank of New York. The SCE began in June 2013 and consists of a rotating panel in which respondents participate for up to twelve months. Each month, approximately 1,300 heads of U.S. households complete the survey.

Each respondent answers a monthly core module and one or more supplementary modules. The core module collects expectations on inflation, unemployment, and household-level outcomes. Three supplementary modules rotate on a four-month schedule: household finances, expenditures, and labor market activity.[2]

We focus on the labor market module, fielded every March, July, and November.

[2]For more details about the survey, its design, sampling methods and questions, see Armantier et al. (2017).

Respondents in this module report their labor market status and their expectations about job offers and earnings over the next four months. Because the module is administered every four months and the SCE rotates respondents annually, we observe each respondent in this module at most three times before they leave the sample. While this limits the length of individual panels, the richness of the subjective questions we use offers a powerful substitute for long earnings histories.

As discussed in the introduction, long individual panels are typically required to nonparametrically identify latent components—such as ability, match quality, and productivity shocks—from discrete outcomes. In our setting, these requirements are relaxed. The subjective expectations data in the SCE provide direct information about both the latent outcome process and the decision rule. Beliefs about future wage offers inform the perceived distribution of counterfactual earnings outcomes, while reported probabilities of accepting hypothetical offers reveal how individuals would respond to those outcomes. These expectations allow us to identify key features of the model—risk, heterogeneity, and selection behavior—using rich cross-sectional variation rather than extended time series data. This enables the credible estimation of dynamic earnings and employment transition models even from a short panel.

Our sample spans from the inception of the survey in June 2013 through March 2020. We use both observed and subjective variables, including annual labor market income, tenure, job type, part-time or full-time status, sector, employer-provided benefits, and standard demographics (age, state, marital status, education, etc.). Table 1 provides some demographics and summary statistics from our sample.[3]

2.1 Key questions

Our analysis focuses on three survey questions from the labor market module of the SCE. These questions elicit respondents' expectations about potential job offers and

[3]Several papers have studied the features of workers' self-reported expectations in the SCE. Koşar and Van der Klaauw (2025) show that workers' expectations about earnings growth and employment risk tend to align well with realized outcomes, both over the life cycle and over the business cycle. Mueller and Spinnewijn (2023) show that unemployed workers tend to have a good estimate of their job finding probabilities at the beginning of an unemployment spell, but fail to update downwards the probabilities of finding a job as their spell becomes longer. Wang (2023) shows that an estimated model based on perceived income risks from the SCE better matches liquid wealth holdings.

their willingness to accept them. They include the expected annual salary from the best offer received in the next four months, the probability distribution of the best offer in the next four months, and the probability of accepting hypothetical offers. All three are asked of both employed and unemployed individuals.

Question on expected best offer. *Think again about the job offers that you may receive within the coming four months. Roughly speaking, what do you think the annual salary for the best offer will be for the first year? Note: the best offer is the offer you would be most likely to accept.*

This question elicits a point estimate for the expected annual salary from the best offer that respondents anticipate receiving in the next four months. We denote this value as $\overline{y}_{it}^{o_{t+1}}$.

Probability distribution for job offers. *Think again about the job offers that you may receive within the coming four months. What do you think is the percent chance that the job with the best offer will have an annual salary for the first year of...*

The best offer is the offer you would be most likely to accept.

- *Less than [80% * $\overline{y}_{it}^{o_{t+1}}$] dollars: ... %*
- *Between [80% * $\overline{y}_{it}^{o_{t+1}}$] and [90% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... %*
- *Between [90% * $\overline{y}_{it}^{o_{t+1}}$] and [100% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... %*
- *Between [100% * $\overline{y}_{it}^{o_{t+1}}$] and [110% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... %*
- *Between [110% * $\overline{y}_{it}^{o_{t+1}}$] and [120% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... %*
- *More than [120% * $\overline{y}_{it}^{o_{t+1}}$] dollars: ... %*

Respondents allocate probabilities across six bins defined relative to $\overline{y}_{it}^{o_{t+1}}$: less than 80% of $\overline{y}_{it}^{o_{t+1}}$, 80–90%, 90–100%, 100–110%, 110–120%, and more than 120%. The survey presents these values in dollar terms, and the system warns respondents if probabilities do not add up to 100%.

Conditional probabilities of accepting job offers. *Think again about the job offers that you may receive within the coming four months. What do you think is the percent chance that you will accept the job if it offers an annual salary for the first year of...*

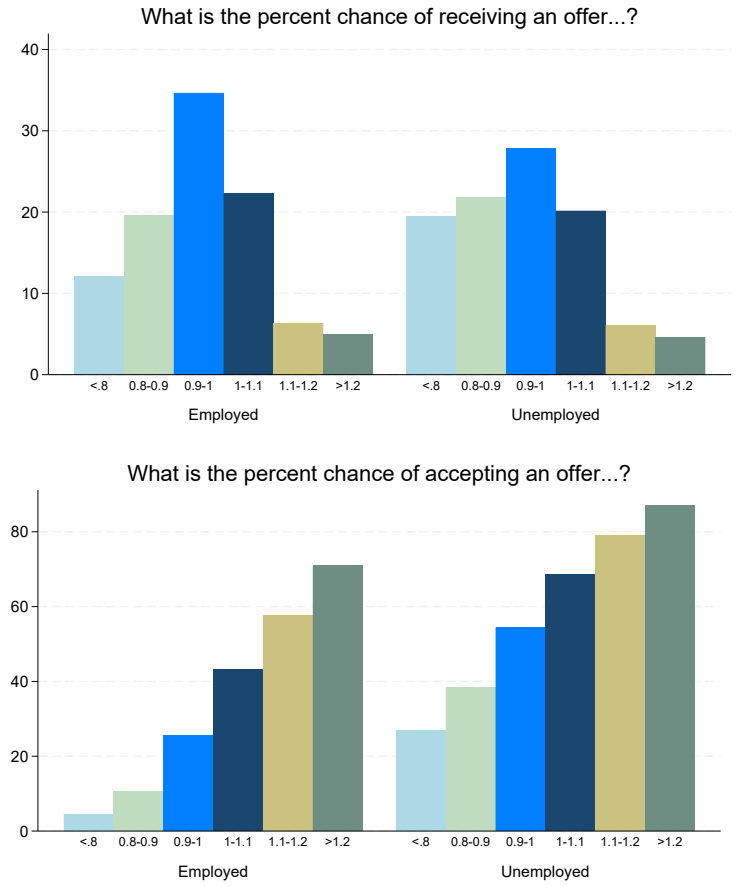
- *Less than [80% * $\overline{y}_{it}^{o_{t+1}}$] dollars: ... percent chance*
- *Between [80% * $\overline{y}_{it}^{o_{t+1}}$] and [90% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... percent chance*
- *Between [90% * $\overline{y}_{it}^{o_{t+1}}$] and [100% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... percent chance*
- *Between [100% * $\overline{y}_{it}^{o_{t+1}}$] and [110% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... percent chance*
- *Between [110% * $\overline{y}_{it}^{o_{t+1}}$] and [120% * $\overline{y}_{it}^{o_{t+1}}$]dollars: ... percent chance*
- *More than [120% * $\overline{y}_{it}^{o_{t+1}}$] dollars: ... percent chance*

Figure 1 shows the average cross-sectional probability distribution for job offers (top panel) and the conditional probability of accepting them (bottom panel). In both cases, the x-axis represents the six bins defined relative to $y_{it}^{o_{t+1}}$, with the leftmost bar corresponding to offers below 80% of the reference value, the next to 80–90%, and so on. Because $y_{it}^{o_{t+1}}$ varies across individuals and over time, the dollar amounts represented by each bin also vary across respondents.

Comparing the distribution of expected offers (top panel) to the reference value $y_{it}^{o_{t+1}}$ provides insight into how respondents interpret the "best offer" question. The distributions are generally consistent with the best offer lying near the mean or median of the expected offer distribution. Unemployed respondents report more skewed distributions, with lower-salary offers being relatively more likely. The bottom panel shows that, as expected, better offers are more likely to be accepted, and that unemployed workers are more likely to accept a given offer than employed workers.

Figure 2 plots the cross-sectional distribution of responses across bins. Each box shows the median (central line), interquartile range (box edges), and the 10th and 90th percentiles (whiskers) for each offer bin. The left panel shows wide variation in the perceived probabilities of receiving offers: for example, the probability of receiving a high-salary offer (in the top two bins) is close to zero for the median respondent but is as high as 10-15% for others. The right panel reveals similar heterogeneity in acceptance

Figure 1: Expected offers (left) and probability of accepting the offer (right).



Note: Probability distribution (cross-sectional averages) of expected job offers and the likelihood of accepting these offers by employment status. y-axis: probability; x-axis: wage offer bins, where e.g. 0.9-1 represents between 90% and 100% of the reference value (the expected best offer).

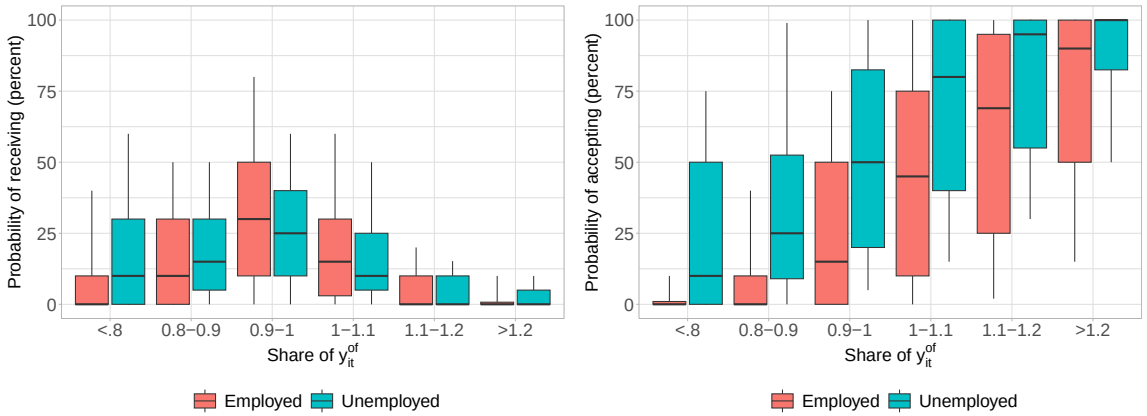
probabilities. In the 100–110% bin, the median reported probability of acceptance is around 50%, but ranges from 15% at the 25th percentile to 75% at the 75th percentile.

2.2 Variation in subjective responses that informs identification

The heterogeneity in responses documented above—across bins and across individuals—is central to identifying the structural components of our model. Figure 3 illustrates four reduced-form patterns in the data. Each one connects a specific subjective expectation to either current earnings or hypothetical offers and helps identify a key model object.

Before describing the panels in the figure, it is important to highlight a distinctive feature of our data. The survey does not only ask respondents what they expect to happen. It also asks how they would behave under counterfactual scenarios. For example, employed and unemployed individuals report how likely they would be to accept job offers

Figure 2: Cross-sectional distribution of the probability of receiving an offer (left) and accepting it (right), by offer bin. Bins are expressed as a share of the reference value $y_{it}^{o_{t+1}}$; for example, 0.9–1 represents between 90% and 100% of the expected best offer. Each box shows the median (central line), interquartile range (box edges), and the 10th and 90th percentiles (whiskers).



at various hypothetical wage levels, each of which they have typically assigned a positive probability of occurrence. These questions generate variation similar to that produced by experimental designs. Hence, all of these questions allow us to observe decision rules over a range of hypothetical and counterfactual earnings, rather than inferring them from realized transitions alone.

The top left panel of Figure 3 shows how expected future earnings depend on current earnings. Specifically, it plots log expected best offers against log current earnings, controlling for individual-job fixed effects. The positive relationship shows that respondents who currently earn more also expect to receive better offers in the near future. This pattern reflects *perceived persistence* in earnings and helps identify the degree to which current earnings predict future offers, holding fixed individual and job-specific factors.

The top right panel examines the subjective dispersion of expected offers. It shows how the (residualized) log odds of the subjective cumulative distribution function (CDF) of offers varies with the level of the hypothetical offer. The log odds transform maps probabilities onto a linear scale, making it easier to compare how quickly the CDF rises across earnings levels. The slope of the line in the figure is informative: a steeper slope corresponds to more concentrated beliefs, while a flatter slope reflects greater uncertainty. Thus, the inverse of the slope provides a reduced-form measure of *subjective earnings risk*.

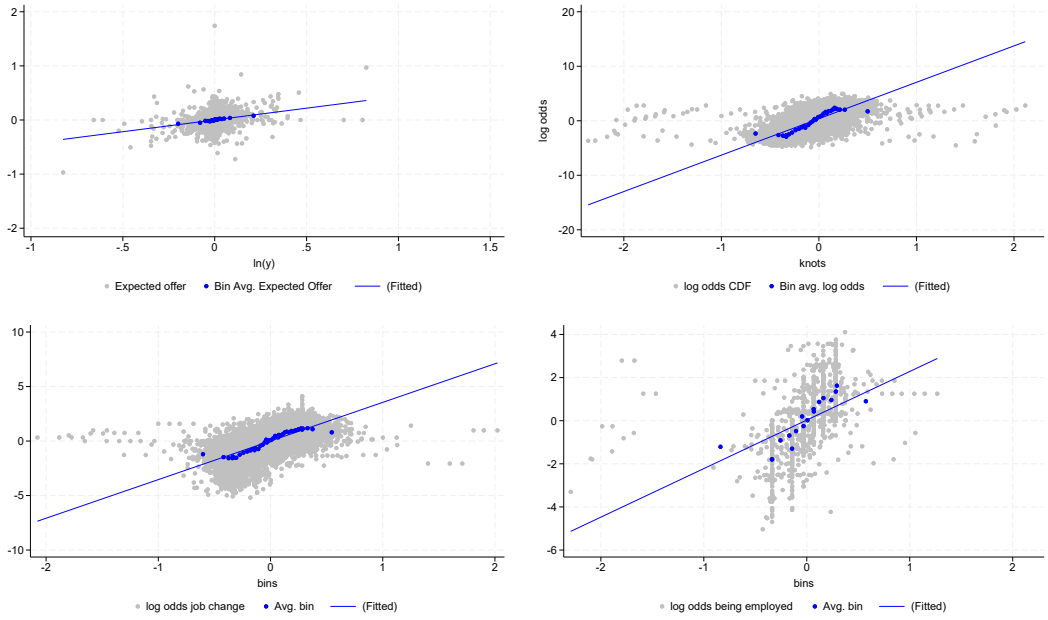
The bottom left panel turns to job-to-job transitions. It plots the (residualized) log

odds of changing jobs as a function of the hypothetical offer. Respondents are more likely to report a willingness to switch jobs when the offer is higher. This variation captures the sensitivity of job mobility decisions to prospective wage gains and helps identify *offer-dependent job transitions* in the model.

The bottom right panel focuses on transitions from unemployment to employment. It shows how the (residualized) log odds of accepting a job varies with the level of the hypothetical offer. As with the employed, unemployed respondents are more likely to accept better offers. This pattern provides information about the reservation wage behavior of the unemployed and contributes to identifying *offer-dependent selection into employment*.

Together, these four reduced-form patterns show that subjective expectations vary systematically with both realized and hypothetical conditions. They also demonstrate that individuals report meaningful, monotonic behavioral responses to counterfactual wage offers. These features provide the variation necessary to identify the persistence, risk, and selection components of our structural model.

Figure 3: Key elements for model identification: scatter plots and fitted lines. Top left: persistence (log expected offers on log earnings). Top right: risk (log odds of the CDF of expected offers on the log of each point/knot). Bottom left: log odds of quitting if employed, by expected offer bin. Bottom right: log odds of becoming employed if unemployed, by expected offer bin.



3 A model of earnings and employment transitions

The model we estimate with subjective expectations data is a version of that developed by Altonji, Smith, and Vidangos (2013), or ASV. It combines two components: an earnings determination block and an employment and job transitions block. The employment block is central to shaping earnings dynamics and selection into both employment and employer transitions.

We depart from the approach taken in ASV along some important dimensions. Our framework and estimation method allow us to disentangle the role of current shocks and heterogeneity in shaping employment transitions without making strong functional form assumptions and using much simpler estimation methods. These shocks, though unobserved by the econometrician, are observed by individuals and thus influence selection into employment and job transitions. This selection, in turn, plays a central role in driving observed earnings dynamics. Furthermore, the simplicity of our approach, leveraging subjective expectations, allows us to estimate employment and job transitions jointly using a multinomial specification in a way that is simple, transparent, and reliable. It is worth noting that ASV regard the multinomial modeling of selection as more attractive, but they do not focus on it due to numerical instability.[4]

Latent earnings dynamics. Starting with the earnings determination block, we assume that log earnings of individual i at time $t+1$, $y_{i,t+1}$, are given by:

$$y_{i,t+1} = y^*_{i,t+1} \times e_{i,t+1} \tag{1}$$

$$y^*_{i,t+1} = x'_{i,t+1}\gamma + \mu_i + \omega_{i,t+1} + \upsilon_{ij(t+1)} \tag{2}$$

$$\omega_{i,t+1} = \rho\omega_{i,t} + \varepsilon^{\omega}_{i,t+1} \tag{3}$$

$$\upsilon_{ij(t+1)} = \begin{cases} \upsilon^0_{ij(t+1)} = \upsilon_{ij(t)} & \text{if } s_{i,t+1} = 0 \\ \upsilon^1_{ij(t+1)} = \phi\upsilon_{ij(t)} + \varepsilon^{\upsilon}_{ij(t+1)} & \text{if } s_{i,t+1} = 1. \end{cases} \tag{4}$$

Log earnings coincide with latent earnings $y^*_{i,t+1}$ for employed individuals ($e_{i,t+1} = 1$). The latter are a function of some observable characteristics, $x_{i,t+1}$, an individual-specific

[4] ASV's main estimated model is a sequential one in which an employed person first decides whether to remain employed and, conditional on that decision, whether to change jobs.

fixed effect, μ_i, and a persistent shock, $\varepsilon^{\omega}_{i,t+1}$ (both unobserved to the econometrician), and a match-specific effect, $\upsilon_{ij(t+1)}$. The observed characteristics $x_{i,t+1}$ include an age profile and an employer tenure polynomial. Tenure is zero for unemployed workers. The time invariant individual-specific effect μ_i captures permanent ability effects. The autoregressive component $\omega_{i,t+1}$ specifies the transmission of productivity shocks, $\varepsilon^{\omega}_{i,t+1}$, with persistence coefficient ρ.

The term $\upsilon_{ij(t+1)}$ represents an employer-specific effect for individual i at $t+1$ (where $j(t+1)$ denotes such individual-employer match). It depends, as specified in equation (4), on the match-specific effect in period t and an employer-specific shock, $\varepsilon^{\upsilon}_{ij(t+1)}$, which becomes operative when the employer-switching indicator $s_{i,t+1}$ equals 1.[5] When there is no employer change, the quality of the match is fixed, otherwise, its persistence is ϕ.

Equations (1) to (4) describe the evolution of latent earnings, given individual employment status and job changes. These transitions are influenced by the innovations $\varepsilon^{\omega}_{i,t+1}$ and $\varepsilon^{\upsilon}_{ij(t+1)}$, as well as by other exogenous employment and mobility shocks. The match shock, $\varepsilon^{\upsilon}_{ij(t+1)}$, affects actual earnings only when $e_{i,t+1} = s_{i,t+1} = 1$.

For notational convenience later on, it is useful to define latent earnings next period in the case of staying with the same employer, $y^{0*}_{i,t+1}$, or switching, $y^{1*}_{i,t+1}$.

$$y^{*}_{i,t+1} = \begin{cases} y^{0*}_{i,t+1} & \text{if } s_{i,t+1} = 0 \\ y^{1*}_{i,t+1} & \text{if } s_{i,t+1} = 1. \end{cases} \tag{5}$$

Employment transitions. Turning to the determinants of employer changes and employment status, each period individuals face a discrete set of choices: remain with their current employment, switch to a new employer, or transition into or out of employment.

For *currently employed* workers ($e_{i,t} = 1$), the multinomial model proceeds by specifying the value of the $t+1$ alternatives "employed in the current job" $ee^{0*}_{i,t+1}$ and "employed

[5] ASV models separately hours and wages, which we combine into earnings, due to data limitations. They also include additional terms in equation (2): a constant if unemployed and another constant if previously unemployed to capture human capital depreciation. While we omit these for simplicity here, our methodology allows for heterogeneous drifts in the productivity process for the employed and unemployed. The shocks $\varepsilon^{\omega}_{i,t+1}$ and $\varepsilon^{\upsilon}_{ij(t+1)}$ are *iid* and independent of each other.

in a new job" $ee^{1*}_{i,t+1}$ relative to "non-employment," whose value is normalized to zero:

$$ee^{0*}_{i,t+1} = x^{0\prime}_{i,t+1}\gamma^0 + \delta^0_y y^{0*}_{i,t+1} + b^0_\mu \mu_i + b^0_\eta \eta_i + \varepsilon^0_{i,t+1} \quad (6)$$

$$ee^{1*}_{i,t+1} = x^{1\prime}_{i,t+1}\gamma^1 + \delta^1_y y^{1*}_{i,t+1} + b^1_\mu \mu_i + b^1_\eta \eta_i + \varepsilon^1_{i,t+1} \quad (7)$$

In equations (6) and (7) the term η_i is a "mobility" fixed effect, representing individual propensity to switch employers and the latent earnings $y^{0*}_{i,t+1}$ and $y^{1*}_{i,t+1}$ are defined in equation 5.[6]

Assuming that $\varepsilon^0_{i,t+1}$ and $\varepsilon^1_{i,t+1}$ are independent with an extreme value distribution (conditionally on $\varepsilon_{i,t+1} = \left(\varepsilon^{\omega}_{i,t+1}, \varepsilon^{\upsilon}_{ij(t+1)}\right)$ and $\Omega_{it} = \left(y^*_{i,t}, x_{i,s}, \mu_i, \upsilon_{ij(t)}\right)$) gives rise to the multinomial logit model. We define the probabilities of staying in the same job $p^0_{i,t}$ and of changing jobs $p^1_{i,t}$ as follows:

$$p^0_{i,t} = \Pr\left(s_{i,t+1} = 0, e_{i,t+1} = 1 \mid \varepsilon_{i,t+1}, \Omega_{it}, e_{i,t} = 1\right), \quad (8)$$

$$p^1_{i,t} = \Pr\left(s_{i,t+1} = 1, e_{i,t+1} = 1 \mid \varepsilon_{i,t+1}, \Omega_{it}, e_{i,t} = 1\right). \quad (9)$$

Based on these definitions and equations (6) and (7), we obtain the following log odds ratios:

$$\ln\left(\frac{p^0_{i,t}}{1 - p^1_{i,t} - p^0_{i,t}}\right) = x^{0\prime}_{i,t+1}\gamma^0 + \delta^0_y y^{0*}_{i,t+1} + b^0_\mu \mu_i + b^0_\eta \eta_i. \quad (10)$$

$$\ln\left(\frac{p^1_{i,t}}{1 - p^1_{i,t} - p^0_{i,t}}\right) = x^{1\prime}_{i,t+1}\gamma^q + \delta^1_y y^{1*}_{i,t+1} + b^1_\mu \mu_i + b^1_\eta \eta_i \quad (11)$$

For the *currently unemployed* workers ($e_{i,t} = 0$), the value of "employment in a new job" relative to non-employment is given by:

$$ue^*_{i,t+1} = x^{u\prime}_{i,t+1}\gamma^u + \delta^u_y y^{1*}_{i,t+1} + b^u_\mu \mu_i + b^u_\eta \eta_i + \varepsilon^u_{i,t+1}. \quad (12)$$

[6]Equations (6), (7), and, later, equation 12 do not allow the match component to have an independent effect over and above that through latent earnings. This restriction is because our data do not allow us to identify such an effect.

With the logistic assumption for $\varepsilon^u_{i,t+1}$ (again conditionally on $\varepsilon_{i,t+1}$ and Ω_{it}) we get:

$$\ln\left(\frac{p^{ue}_{i,t}}{1-p^{ue}_{i,t}}\right) = x^{u\prime}_{i,t+1}\gamma^u + \delta^u_y y^{1*}_{i,t+1} + b^u_\mu \mu_i + b^u_\eta \eta_i. \tag{13}$$

Unlike ASV, we do not constrain δ^u_y to be zero, thus allowing for endogenous self-selection into employment. This coefficient captures the reaction of unemployment individuals to their labor market opportunities and, thus, reflects an important economic mechanism.

To sum up, endogenous self-selection in our model of earnings and employment dynamics arises as a result of dependence of the transition probabilities $p^1_{i,t}$, $p^0_{i,t}$ and $p^{ue}_{i,t}$ on the time $t+1$ innovations to latent earnings, as specified in equations (2) and (3) conditionally on the information set up to period t. As we discuss later, our approach is possible because the subjective expectations data provide direct information on y^*_{t+1}.

4 Using subjective expectations to identify the model

The model we discussed in Section 3 is made of two main blocks: the first characterizes the evolution of latent earnings while the second deals with job transitions (to and from unemployment and/or to different employers). Latent earnings are determined by an individual fixed effect, by a stochastic process that we label 'productivity' and by employer-match quality. As for transitions, the currently employed face three potential outcomes next period: remain employed with the same employer, remain employed with another employer, and become unemployed. The currently unemployed, instead, face two potential outcomes: remain unemployed, or become employed.

Our identification and estimation approach is based on the idea of matching expectations in the data to those in our model. Therefore, in this section, we first compute the conditional earning expectations and the distribution of best offers, and then move to the probability of employment and of employer changes. This allows us to relate these quantities to those observed in our data, which in this section appear on the left-hand-side of some of the equations and which we bold for ease of exposition.

4.1 Latent earnings and wage offers

For future earnings, in the data, we have information about the entire probability distribution. However, for expositional purposes, we start with a discussion of the means of its distribution. We then extend this discussion to the entire distributions.

We start by expressing latent earnings as an autoregressive process by rewriting the AR(1) process for $\omega_{i,t+1}$ in equations (2) and (3) as:

$$y^*_{i,t+1} - x'_{i,t+1}\gamma - \mu_i - \upsilon_{i,j(t+1)} = \rho\left(y^*_{i,t} - x'_{i,t}\gamma - \mu_i - \upsilon_{i,j(t)}\right) + \varepsilon^{\omega}_{it} \tag{14}$$

Re-arranging terms, we get:

$$y^*_{i,t+1} = \rho y^*_{i,t} + \left(x_{i,t+1} - \rho x_{i,t}\right)'\gamma + (1-\rho)\,\mu_i + \upsilon_{i,j(t+1)} - \rho\upsilon_{i,j(t)} + \varepsilon^{\omega}_{i,t+1}, \tag{15}$$

subject to equation 4 for the law of motion for $\upsilon_{i,j(t+1)}$.

Potential offers. As discussed in Section 2, the SCE contains information on both the prediction and the probability distribution of the yearly salary one might be offered during the next four months, for both the employed and the unemployed. To map these data to the model's quantities, it is convenient to distinguish between earnings from changing employer and earnings with the same employer for the currently employed, $\left(y^{1*}_{i,t+1}, y^{0*}_{i,t+1}\right)$, which are given by the following expressions:

$$y^{1*}_{i,t+1} = \rho y^*_{i,t} + \left(x_{i,t+1} - \rho x_{i,t}\right)'\gamma + (1-\rho)\,\mu_i + (\phi-\rho)\,\upsilon_{i,j(t)} + \varepsilon^{\upsilon}_{i,j(t+1)} + \varepsilon^{\omega}_{i,t+1} \tag{16}$$

$$y^{0*}_{i,t+1} = \rho y^*_{i,t} + \left(x_{i,t+1} - \rho x_{i,t}\right)'\gamma + (1-\rho)\,\mu_i + (1-\rho)\,\upsilon_{i,j(t)} + \varepsilon^{\omega}_{i,t+1} \tag{17}$$

Our approach exploits the observability in the SCE of the expected potential earnings in $t+1$ with a new employer, which in logs we denote with $\overline{\mathbf{y}}^{\mathbf{o_{t+1}}}_{\mathbf{it}}$, as well as several points of its probability distribution. For expositional purposes, we start by equating $\overline{\mathbf{y}}^{\mathbf{o_{t+1}}}_{\mathbf{it}}$ to the expected value of next period's log earnings in the event of switching employer, subject to additive elicitation error that is mean independent of the respondent's information set.

This implies that, for job changers, $\upsilon_{i,j(t+1)} = \phi\upsilon_{i,j(t)} + \epsilon^{\upsilon}_{i,j(t+1)}$, hence:

$$\overline{\mathbf{y}}_{\mathbf{it}}^{\mathbf{o_{t+1}}} \equiv \quad E\left(y_{i,t+1}^{1*} \mid \Omega_{it}\right) = \rho y_{it}^{*} + \left(x_{i,t+1} - \rho x_{it}\right)'\gamma + (1-\rho)\,\mu_i + (\phi - \rho)\upsilon_{i,j(t)}. \quad (18)$$

In addition to $\overline{\mathbf{y}}_{\mathbf{it}}^{\mathbf{o_{t+1}}}$, the SCE also contains information on its probability distribution, which we can match to the corresponding model quantities, as we discuss below. Importantly, the parameters of interest can be estimated using only the information on the probability distribution, avoiding the assumption that $\overline{\mathbf{y}}_{\mathbf{it}}^{\mathbf{o_{t+1}}}$ is interpreted as the *mean* of the potential offer with a new employer.

It is also worthwhile noticing that we observe the distribution of latent earnings for all individuals, regardless of whether they work or switch employers in the next period. By contrast, the common practice of using realized earnings in $t+1$ restricts the sample to those who actually work. Consequently, our specification avoids the selection bias that typically arises when estimation relies on realized earnings alone. This implies that we can obtain consistent estimates of the parameters in equation (18) using only data on the employed because in our case $e_{i,t} = 1$ represents an exogenous selection relative to shocks in $t+1$.

This last consideration is particularly noticeable. Estimating equation 18 requires observing $y_{i,t}^{*}$, which for employed individuals coincides with current earnings. The model also has implications for the unemployed, but their latent earnings are unobserved. Furthermore, the SCE contains only a small sample of unemployed respondents. We therefore drop the unemployed equations, while reporting them for completeness in Appendix A.1. This exclusion does not introduce selection bias.

Future earnings probability distribution for the employed. As we discussed in Section 2, in addition to point expectations, the survey reports the probability distribution of the "annual salary of the best offer likely to be received within the next four months" in 10-percentage-point bins, ranging from less than 80% to more than 120% of the expected offer. Assuming classical measurement error in the observed log odds

transformation of the CDF,[7] u^p_{kit}, we have:

$$\text{logit}(\overline{\mathbf{p}}^{\mathbf{o_{t+1}}}_{\mathbf{kit}}) = \text{logit}(\Pr\left(y^{1*}_{i,t+1} \leq r_{kit} \mid \Omega_{it}\right)) + u^p_{kit} \tag{19}$$

where $\mathbf{r_{kit}} = \tau_k \times \overline{y}^{o_{t+1}}_{it}$ and $\tau_k \in \{0.8, 0.9, 1.0, 1.1, 1.2\}$. So, in view of (16) we have:

$$\begin{aligned} & \Pr\left(y^{1*}_{i,t+1} \leq r_{kit} \mid \Omega_{it}\right) \\ = \; & \Pr\left(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{i,j(t+1)} \leq r_{kit} - \rho y^*_{i,t} - \left(x_{i,t+1} - \rho x_{i,t}\right)'\gamma - (1-\rho)\,\mu_i - (\phi - \rho)\,\upsilon_{i,j(t)} \mid \Omega_{it}\right) \end{aligned}$$

Assuming that $\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{i,j(t+1)}$ has a logistic distribution, our model for the log odds transformation $\overline{\boldsymbol{\ell}}^{\mathbf{o_{t+1}}}_{\mathbf{kit}} = \ln\left(\frac{\overline{\mathbf{p}}^{\mathbf{o_{t+1}}}_{\mathbf{kit}}}{\mathbf{1}-\overline{\mathbf{p}}^{\mathbf{o_{t+1}}}_{\mathbf{kit}}}\right)$ is linear in the variables:

$$\begin{aligned} \overline{\boldsymbol{\ell}}^{\mathbf{o_{t+1}}}_{\mathbf{kit}} = \; & (1/\sigma_e)\, r_{kit} + (-\rho/\sigma_e)\, y^*_{i,t} + \left(x_{i,t+1} - \rho x_{i,t}\right)'(-\gamma/\sigma_e) \\ & -\mu_i\,(1-\rho)\,/\sigma_e - (1/\sigma_e)\,(\phi - \rho)\,\upsilon_{i,j(t)} + u^p_{kit} \end{aligned} \tag{20}$$

where σ_e denotes the standard deviation of $\left(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{i,j(t+1)}\right)$.[8]

As noted above, the parameters of interest can be identified using only the probability distribution of best offers (equation 20), so the equation for expected future latent earnings is not needed. Accordingly, interpreting $\overline{\mathbf{y}}^{\mathbf{ot+1}}_{\mathbf{it}}$ as the *mean* of the relevant variable is unnecessary; instead, it simply anchors the probability distribution.

As with expected latent earnings, it is possible to derive expressions for their probability distribution also for the unemployed. However, we relegate them in Appendix A.1, like we did for the discussion for the means.

4.2 Employment transitions

In the previous section, we describe how, based on one's expected best offer, the SCE provides potential offer bins, with respondents reporting the probability of receiving

[7] Respondents report the probabilities for each bin, which we then use to construct the observed CDF and log cumulative odds $\text{logit}(CDF) = \ln[CDF/(1-CDF)]$. It is natural to assume additive measurement error in the log cumulative odds, as these values can span the entire real line.

[8] While not pursued in our current implementation of estimators, a sufficiently flexible specification of the coefficients in equation (20) as functions of r_{kit} can capture a broad class of probability distributions for $\left(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{i,j(t+1)}\right)$–including those exhibiting skewness–regardless of the logistic transformation that we use (see Arellano et al. (2024)).

an offer in each of them. Furthermore, the SCE also asks respondents to state their probability (in percent) of accepting the offer in each bin. This information allows us to estimate the transition probabilities between labor market states.

Given that in the model transitions are determined by comparing the values of various labor market statuses (as in equations (6 and 7), where the value of unemployment is normalized to zero, we can map the observed probabilities of transitions observed in the data to their counterparts in the model as follows:[9]

$$\begin{aligned}
\mathbf{p^{ue}_{kit}} &= \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid \tau_k \cdot \overline{y}^{o_{t+1}}_{i,t} < y^{1*}_{i,t+1} \leq \tau_{k+1} \cdot \overline{y}^{o_{t+1}}_{i,t}, 0, \Omega_{it}, e_{i,t}=0\right), \quad (21)\\
\mathbf{p^{e1}_{kit}} &= \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid \tau_k \cdot \overline{y}^{o_{t+1}}_{i,t} < y^{1*}_{i,t+1} \leq \tau_{k+1} \cdot \overline{y}^{o_{t+1}}_{i,t}, y^{0*}_{i,t+1}, \Omega_{it}, e_{i,t}=1\right),\\
\mathbf{p^{e0}_{kit}} &= \Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid \tau_k \cdot \overline{y}^{o_{t+1}}_{i,t} < y^{1*}_{i,t+1} \leq \tau_{k+1} \cdot \overline{y}^{o_{t+1}}_{i,t}, y^{0*}_{i,t+1}, \Omega_{it}, e_{i,t}=1\right).\\
&\quad \tau \in \{-\infty, 0.8, 0.9, 1.0, 1.1, 1.2, +\infty\}, \quad k=1,\ldots,6.
\end{aligned}$$

The term $\mathbf{p^{ue}_{kit}}$ refers to the probability of becoming employed for an unemployed person (with $(\mathbf{1}-\mathbf{p^{ue}_{kit}})$ being the probability of remaining unemployed). The terms $\mathbf{p^{e1}_{kit}}$ and $\mathbf{p^{e0}_{kit}}$ are the probabilities for the currently employed of switching employer and staying with the same employer, respectively. The probability to become unemployed is thus $(1-\mathbf{p^{e1}_{kit}}-\mathbf{p^{e0}_{kit}})$.

To implement this approach, we approximate the quantities in equation (21) by attributing the probability mass in each interval to the mid point of the grid bin for the interior points, and to 0.75 or 1.25 for the the grid boundaries. For example, for the unemployed, we set it as follows:

$$\begin{aligned}
\mathbf{p^{ue}_{kit}} &\approx \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y^{1*}_{i,t+1} = \tilde{\tau}_k \cdot \overline{y}^{o_{t+1}}_{it}, \Omega_{it}, e_{i,t}=0\right) \quad (22)\\
&\quad \tilde{\tau}_k \in \{0.75, 0.85, 0.95, 1.05, 1.15, 1.25\}, \quad k=1,\ldots,6,
\end{aligned}$$

where $\tilde{\tau}_k$ is the point used for each bin.

The next step is to take the log-odds ratios and use our model to obtain the following

[9]Subject to classical additive measurement error in the logit of the observed probabilities, as before.

linear specifications, which are the empirical equivalent of equation (13):

$$\overline{\ell}^{\mathbf{ue}}_{\mathbf{kit}} = \ln\left(\frac{\mathbf{p^{ue}_{kit}}}{1-\mathbf{p^{ue}_{kit}}}\right) = x^{u\prime}_{i,t+1}\gamma^u + \delta^u_y\left(\tilde{\tau}_k \cdot \overline{y}^{o_{t+1}}_{it}\right) + b^u_\mu \mu_i + b^u_\eta \eta_i + \xi^{ue}_{kit}, \quad (23)$$

$$\overline{\ell}^{\mathbf{e1}}_{\mathbf{kit}} = \ln\left(\frac{\mathbf{p^{e1}_{kit}}}{1-\mathbf{p^{e1}_{kit}}-\mathbf{p^{e0}_{kit}}}\right) = x^{1\prime}_{i,t+1}\gamma^1 + \delta^1_y\left(\tilde{\tau}_k \cdot \overline{y}^{o_{t+1}}_{it}\right) + b^1_\mu \mu_i + b^1_\eta \eta_i + \xi^{e1}_{kit}, \quad (24)$$

$$\overline{\ell}^{\mathbf{e0}}_{\mathbf{kit}} = \ln\left(\frac{\mathbf{p^{e0}_{kit}}}{1-\mathbf{p^{e1}_{kit}}-\mathbf{p^{e0}_{kit}}}\right) = x^{0\prime}_{i,t+1}\gamma^0 + \delta^0_y \widehat{y}^{0*}_{i,t+1} + b^0_\mu \mu_i + b^0_\eta \eta_i + \xi^{e0}_{kit}. \quad (25)$$

Here, four points are worth noticing. First, while both $y^{1*}_{i,t+1}$ and $y^{0*}_{i,t+1}$ (latent earnings with a new employer and with the current employer) enter the probabilities that the employed switch employers or staying, due to the properties of the logistic and our functional form assumptions, only one of the two enters the log-odds ratios.

Second, and very importantly, because in the SCE people report the probability of becoming employed (if unemployed) or of changing employer (if employed) conditionally on various hypothetical offered earnings, the term $\tilde{\tau}_k \cdot \overline{y}^{o_{t+1}}_{it}$ in equations (23) and (24) is observed. While Altonji et al. (2013) attempt including a similar, model-predicted term, they encountered poor identification issues.

Third, because the SCE does not ask about the probability of staying with the same employer, *conditional on receiving a certain job offer* $\mathbf{p^{e0}_{kit}}$, we cannot decompose the probability of "not accepting an offer" $1 - \mathbf{p^{e1}_{kit}}$ between that of staying on the same job and that of moving into unemployment. However, because we observe the unconditional probabilities of changing jobs, staying with the same employer and moving into unemployment, we can use them to reconstruct $\mathbf{p^{e0}_{kit}}$. To do so, we assume that individuals who receive relatively low job offers and reject them are more likely to transition into unemployment, while individuals who receive relatively high offers and reject them are more likely to stay with the same employer. We provide further details in Appendix B.1.

Fourth, because the SCE does not ask future income with the same employer upon the receipt of a potential outside offer, in equation (25), we replace $y^{0*}_{i,t+1}$ with $\widehat{y}^{0*}_{i,t+1} = E\left(y^{0*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1} = e_{i,t} = 1, s_{i,t+1} = 0\right)$, that is, its conditional expectation, for which we can construct an empirical measure from the SCE data. This substitution is equivalent to assuming that endogenous self-selection is driven by $\varepsilon^{\upsilon}_{i,j(t+1)}$ but not $\varepsilon^{\omega}_{i,t+1}$. In other words, that the reported offer reflects the future employer specific shock in the match,

but not the one in human capital.[10] As a result, the model conveniently simplifies to a system of linear equations.

We can use two survey questions to construct $\widehat{\mathbf{y}}^{\mathbf{0}*}_{\mathbf{i,t+1}}$. The first explicitly asks for a point estimate of earnings, conditional on maintaining the same employer and working the same number of hours, 12 months ahead. The second one pertains to unconditional subjective expected earnings in four months' time. The problem with the first question is that it conditions on the same number of hours and has a slight time horizon mismatch. This lead us to use the second question and additional survey information to construct the conditional subjective expected earnings that we need. We discuss the procedure we follow in Appendix A.2.

5 Two-step model estimation

Section 4 describes how subjective expectations provide the key inputs for identifying our earnings and employment transitions. This section focuses on estimating those equations and the error-components structure they imply. Our estimation proceeds in two steps.

In **Step 1**, we estimate ρ, σ_e and γ using the equations on latent earnings while allowing for fixed effects of ability, mobility and employer-match, as well as δ_y^u, δ_y^1 and δ_y^0 using the equations that determine job mobility. In **Step 2**, we estimate the remaining model parameters (ϕ, b_μ^u, b_μ^0, b_μ^1, b_η^u, b_η^0, b_η^1) as well as σ_e, by enforcing error-component restrictions on the covariance matrix of the Step 1 residuals using a minimum-distance approach. We keep the parameter ρ fixed at its within-group estimated value.[11]

Step 1. We organize Step 1 estimation into two main blocks of equations. The first block uses SCE data on future latent earnings and includes equations (18) and (20). It includes both the reported point earnings expectations (equated to the model-implied mean future earnings for employer switchers) and their distribution. The second block includes information on the probability of changing employer conditional on being in a certain offer bin (and equates the observed log-odds ratios in each latent offer bin for

[10] Appendix A.3 examines more complex versions of the multinomial model that incorporate endogenous selection in $\varepsilon^{\omega}_{i,t+1}$ and discusses the challenges associated with their estimation.

[11] We re-estimate the standard deviation σ_e in Step 2, but results are unaffected if we just keep the value obtained in the first step.

the employed). There are six bins, hence six estimating equations. In each of these equations we also consider the possible presence of elicitation error, which is assumed to be uncorrelated with the variables considered in these equations.

For each equation, we estimate ρ, and then obtain a single estimate for it by minimum distance, with bootstrap standard errors.

Here, it is important to note that our methodology identifies the model's structural parameters using only information about the probability distribution. This approach does not require assuming that the point expectation of the "best offer" coincides with the mean (rather than another moment, such as the median) implied by our model.

The second block of estimating equations uses SCE data on job transition probabilities. Equations (23), (24), and (25) correspond to transitions from unemployment to employment, and to transitions for employed switchers and stayers, respectively, which again include classical elicitation errors.

Step 2. In a second step, we use GMM to estimate fixed-effect heterogeneity σ_μ, the persistence of job-match quality ϕ, σ_ν and σ_e. Having estimated the equations in the first step, we can obtain estimates of their residuals, which, in addition to the fixed effects contain elicitation errors in the subjective expectations questions. In particular, the residuals we use are given by the following set of equations:

$$\overline{ry}^{of}_{it} = (1-\rho)\mu_i + \upsilon_{ij(t)}(\phi-\rho) + \xi^{m-offer}_{it}, \tag{26}$$

$$\overline{rl}^{of}_{jit} = -\mu_i(1-\rho)/\sigma_e - \upsilon_{ij(t)}(\phi-\rho)/\sigma_e + \xi^{\ell-offer}_{jit}, \tag{27}$$

$$\overline{rp}^{ue}_{jit} = b^u_\mu \mu_i + b^u_\eta \eta_i + \xi^{ue}_{jit}, \tag{28}$$

$$\overline{rp}^{0}_{jit} = b^0_\mu \mu_i + b^0_\eta \eta_i + \xi^{e0}_{jit}, \tag{29}$$

$$\overline{rp}^{1}_{jit} = b^1_\mu \mu_i + b^1_\eta \eta_i + \xi^{e1}_{jit}, \tag{30}$$

where $\xi^{m-offer}_{it}$, $\xi^{\ell-offer}_{jit}$, ξ^{ue}_{jit}, ξ^{e0}_{jit}, and ξ^{e1}_{jit} are elicitation errors. Appendix B.2 provides detailed information on the implementation of the minimization problem, the weighting scheme, and other technical considerations.

The estimation uses GMM on both the variances of the residuals and their cross-equation covariances, including contemporaneous components and those computed over

the three observation periods for each worker. When analyzing autocovariances, we account for whether individuals have transitioned across employers or not.

The system of equations (26)–(30) constitutes a linear factor model. Under sufficient independence assumptions regarding the elicitation errors, the probability distributions of the ability, mobility, and match factors are nonparametrically identified. This contrasts sharply with discrete choice models based on short panels of realized outcomes, where fixed-effects identification is typically not feasible without strong additional assumptions. A potential third step in our procedure—currently not implemented—would be to estimate the components of heterogeneity nonparametrically via deconvolution.

Given the complexity of the model and of the estimation approach, we summarize all the notation used so far in Tables 7 and 8.

6 Results

In the previous sections, we discussed the available data, the theoretical model, the mapping of some equations to the data, and the identification of the structural parameters. We now present our empirical analysis, starting with the criteria that inform the sample selection and then move on to our parameter estimates.

We focus on individuals aged 25 to 62 and exclude the self-employed. Given that our panel dimension is short, we only include employer tenure in our explanatory variables x_{it}.

In our model, all individuals face a distribution of potential offers each period. In the data, 20% of respondents report a zero probability of receiving an offer in the next period and therefore provide no information about their distribution of expected potential earnings. Consequently, we only use data for those who report a positive probability of receiving an offer. By an argument analogous to the one we used for the unemployed, this restriction does not introduce selection bias.

Estimated parameters. Table 2 reports the parameters affecting the evolution of latent earnings (equations (1) to (4)) obtained from the empirical equations‘(18 and 20. Starting with ρ, which determines the degree of persistence of the productivity process

Table 2: Persistence and risk parameters

	Coefficient	Men	Women
Persistence in productivity	ρ	0.555***	0.433***
		(0.095)	(0.068)
Pers. job-specific component	ϕ	0.156	0.194
		(0.205)	(0.346)
SD individual FE	σ_μ	0.549***	0.493***
		(0.076)	(0.101)
SD $(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{ij,t+1})$	σ_e	0.117***	0.144***
		(0.007)	(0.012)
SD job-specific component	σ_υ	0.542	0.789
		(0.510)	(0.850)

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

ω_t, we estimate a value 0.555 (se 0.095) for men and 0.433 (se 0.068) for women.[12] These are considerably lower than those commonly estimated or assumed in the labor literature, which typically range between 0.7 and 1.0 (for instance, Hall and Mishkin (1982), MaCurdy (1982), Abowd and Card (1989), and Meghir and Pistaferri (2004)). The magnitude of the estimated persistence parameters often depends on how individual unobserved heterogeneity is modeled; estimates tend to be lower when more flexible forms of heterogeneity are allowed (e.g., Browning et al. (2010), Hospido (2015), and Alvarez and Arellano (2022)). Altonji et al. (2013), who estimate the same model, including random affects, on PSID data, obtain $\rho \approx 0.95$. As we will see in Section 7, a lower structural persistence of general productivity shocks does not preclude the estimation of high reduced-form earnings persistence.

Another parameter affecting persistence in observed earnings is ϕ, which captures the extent to which employer-match quality carries over after a job switch. We estimate ϕ at 0.156 (se 0.205) for men and 0.194 (se 0.346) for women. Both are statistically indistinguishable from zero, implying substantial depreciation of employer-specific capital. These estimates contrast sharply with the value of 0.7 reported in Altonji et al. (2013)

Our estimates for the standard deviation of the "ability" fixed effects (σ_μ) (0.55 (se 0.076) for men and 0.49 (se 0.101) for women) are considerably higher than ASV's, which finds 0.081.[13] Importantly, our estimates are obtained without any functional

[12]We report the different equations with which we estimate ρ in Appendix **??**.

[13]We normalize the coefficient of the ability fixed effect μ_i in the latent earning equation to 1 and report the standard deviation of the fixed effects. ASV, instead, normalize the standard deviation of the same variable to 1. The 0.081 we report in the text is the coefficient of the ability fixed effect in the

form assumption on the distribution of fixed effects.

In our model, latent earnings vary due to innovations in ability, ω_t, and employer-match quality, $\nu_{ij,(t)}$. We estimate the standard deviation of their combined innovation, $(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{ij,t+1})$, denoted σ_e, with high precision. We find $\sigma_e = 0.12$ (se 0.007) for men and 0.14 (se 0.012) for women, both well below the 0.29 reported by Altonji et al. (2013).

The last row of the table reports the cross-sectional standard deviation of the employer-match quality fixed effects. Although it is not statistically significant, its point estimate is similar in magnitude to that of σ_μ.

Tables 3 and 4 report the parameter estimates affecting transitions to new jobs, as specified in equations (23), (24) and (25). Table 3 reports the estimated transition parameters. The term δ^u_y measures the effect of potential offers on the probability of leaving unemployment, δ^0_y the effect of current earnings on the probability of staying with the same employer, and δ^1_y the effect of potential offers on the probability of quitting. The coefficient estimates in columns (2) and (4) for men and women, respectively, display the impact of these variables on the log-odds ratio of the corresponding probability. Columns (3) and (5) translate these effects into changes in probability points.

Table 3: Job transition parameters

		Men		Women	
	Coefficient	Estimate	PP change	Estimate	PP change
	(1)	(2)	(3)	(4)	(5)
Effect of exp. offer	δ^u_y	3.361***	0.80	1.549***	0.34
on Pr(working)		(0.177)		0.118)	
Effect of exp. offer	δ^1_y	3.661***	0.60	2.809***	0.43
on Pr(quitting)		(0.032)		(0.032)	
Effect of earnings	δ^0_y	0.294*	0.03	0.458**	0.05
on Pr(staying)		(0.160)		(0.179)	

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The impacts of expected offers on the probability of leaving unemployment and on quitting are precisely estimated and have the expected sign. These effects are larger for men than for women, especially for the probability of leaving unemployment. The impact of expected offers is greater for quitting than for leaving unemployment, which might be due to the fact that it is more costly for the unemployed to wait for the next offer, hence

latent earnings equation. The two normalizations are obviously equivalent.

they might be less sensitive to the salary offered.

The effect of current earnings on the probability of staying is small and imprecisely estimated, especially for men, where it is statistically different from zero only at the 10% level. This imprecision likely reflects the absence of data on how current employers adjust pay in response to outside offers and our need to approximate y_{it}^0. It also stems from the survey design: while it elicits responses to potential job offers and acceptance probabilities, it does not ask how workers would change their probability of staying if their current employer raised or cut their pay.

Table 4: Individual fixed effects and transitions

		Men	Women
	Coefficient		
	(1)	(2)	(3)
Standard deviation of η	σ_η	0.889**	0.581*
		(0.381)	(0.308)
Effect of μ	b_μ^u	-7.064***	-2.593**
on Pr(working)		(2.216)	(1.297)
Effect of η	b_η^u	1.526	-4.054*
on Pr(working)		(5.068)	(2.433)
Effect of μ	b_μ^1	-4.750***	-3.929***
on Pr(quitting)		(0.845)	(1.056)
Effect of μ	b_μ^0	-0.590***	-0.292
on Pr(staying)		(0.222)	(0.399)
Effect of η	b_η^0	-0.604	-0.999
on Pr(staying)		(0.732)	(0.870)

Table 4 reports the effects of fixed abilities and mobility effects on transition probabilities: leaving unemployment (first two rows), changing jobs for the employed (next row), and staying with the same employer (final two rows). We normalize the coefficient on the mobility fixed effect η in the equation for the probability of changing jobs for the employed to 1.

For men, the coefficients on the ability fixed effects are all negative and highly significant across the three probabilities. By contrast, the coefficient on mobility fixed effects are estimated less precisely. However, as expected, the point estimate for leaving unemployment is positive, and the point estimate for staying in the same job is negative.

For women, the ability fixed effects follow a similar pattern but are smaller in magni-

tude, and the coefficient for staying is not significantly different from zero. The mobility fixed effects are, similarly to men, negative and insignificant for staying but, unlike for men, negative and significant for leaving unemployment. The surprising fact that the mobility fixed effect reduces the probability of leaving unemployment suggests that, in the case of women, the ability and mobility fixed effects might not be sufficient to capture the total extent of unobserved heterogeneity, which may also include a (dis)taste for work. We explore this possibility further in Appendix **??**.

7 The role of risk and unobserved heterogeneity

As we mentioned above, the model we have estimated is a rich a complex one. The dynamics of observed earnings depends on the the dynamics of ability, on the persistence of job-specific human capital and on selections into and out of unemployment or job changes. Furthermore, quantifying the persistence and variability of earnings, not to mention risk, using a cross section with relatively short horizon makes the problem harder if not impossible with only observational data. Having estimated the parameters of this model with the help of subjective expectations complemented with actual realizations, we are now in a position to dissect the various components of this model and get a better understanding of its dynamic properties. We do this by simulating the estimated model given the estimates we have. We stress that identification was achieved with a minimal set of assumptions on the primitives of the model, including the distribution of unobserved heterogeneity.

7.1 Decomposing the variance and persistence of earnings

We now turn to the relative contributions of risk and unobserved heterogeneity in explaining fluctuations in workers' earnings.

Let α_i denote unobserved heterogeneity in the form of an individual fixed effect. Earnings volatility conditional on $y_{i,t}$ can be decomposed into a risk component and an

unobserved heterogeneity component:[14]

$$\begin{aligned}\mathrm{Var}(y_{i,t+1} \mid y_{i,t}) = \; & \mathbb{E}_{\alpha_i|y_{i,t}}\left[\mathrm{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right] + \\ & \mathrm{Var}_{\alpha_i|y_{i,t}}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right], \end{aligned} \tag{31}$$

where $\mathrm{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i)$ is the earnings risk faced by an individual with current income $y_{i,t}$ and heterogeneity α_i. The first term is the average risk among individuals with income $y_{i,t}$, while the second captures the contribution of unobserved heterogeneity to $\mathrm{Var}(y_{i,t+1} \mid y_{i,t})$.

One can further decompose the first term on the right hand side of Equation 31 by taking into account an additional component of unobserved heterogeneity which is individual-employer-specific and thus can vary over t $\zeta_{ij(t)}$:

$$\begin{aligned}\mathrm{Var}(y_{i,t+1} \mid y_{i,t}) = \; & \mathbb{E}_{\zeta_{ij(t)},\alpha_i|y_{i,t}}\left[\mathrm{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{ij(t)}) \mid y_{i,t}\right] + \\ & \mathbb{E}_{\alpha_i|y_{i,t}}\left(\mathrm{Var}_{\zeta_{ij(t)}|y_{i,t},\alpha_i}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{ij(t)}) \mid y_{i,t}\right]\right) + \\ & \mathrm{Var}_{\alpha_i|y_{i,t}}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right]. \end{aligned} \tag{32}$$

Here, the first term captures the earnings risk faced by an individual with current income $y_{i,t}$, fixed effect α_i, and employer-specific effect $\zeta_{ij(t)}$; the second term reflects the contribution of unobserved heterogeneity from the employer-specific component; and the third term reflects the contribution of individual ability fixed effects.

Likewise, we can decompose the persistence of earnings $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t})}{\partial y_{i,t}}$ as follows:

$$\begin{aligned}\frac{\partial \mathbb{E}(y_{i,t+1} \mid y_{i,t})}{\partial y_{i,t}} = \; & \mathbb{E}\left[\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\alpha_i)}{\partial y_{i,t}} \mid y_{i,t}\right] + \\ & \mathbb{E}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i)\frac{\partial \ln f(\alpha_i|y_{i,t})}{\partial y_{i,t}} \mid y_{i,t}\right]. \end{aligned} \tag{33}$$

where $\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i)$ denotes a reduced form conditional-mean regression of $y_{i,t+1}$ on $y_{i,t}$ and a fixed effect. One can also expand the first term in this equation to separate

[14]For now, we abstract from unemployment risk and restrict attention to workers employed in both periods t and $t+1$. Note that α_i differs from μ_i, and $\zeta_{ij(t)}$ below from $\upsilon_{ij(t)}$.

employer-specific and individual-specific time invariant components to obtain:

$$\begin{aligned}\frac{\partial \mathbb{E}(y_{i,t+1} \mid y_{i,t})}{\partial y_{i,t}} = & \quad \mathbb{E}\left[\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\zeta_{ij(t)})}{\partial y_{i,t}} \mid y_{i,t}\right] + \\ & \mathbb{E}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t},\zeta_{ij(t)})\frac{\partial \ln f(\zeta_{ij(t)}|y_{i,t},\alpha_i)}{\partial y_{i,t}} \mid y_{i,t}\right] + \\ & \mathbb{E}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t},\alpha_i)\frac{\partial \ln f(\alpha_i|y_{i,t})}{\partial y_{i,t}} \mid y_{i,t}\right]\end{aligned} \tag{34}$$

In this expression, the first component represents the contribution of persistence in earnings after removing fixed effects to overall persistence, the second component represents the contribution of the individual-employer specific component $\zeta_{ij(t)}$ to overall persistence, and the third component represents the contribution of the individual-specific component α_i to overall persistence.

7.2 Implementation and results

We begin by using our model and estimated equations to simulate a panel of individuals. Because our exercise requires estimating fixed-effects regressions on simulated earnings, we can simulate a long panel of 100 periods for 10,000 individuals. This allows us to avoid the Nickell bias (Nickell, 1981).[15]

To reflect that the probability, in the observed data, of receiving an offer is, on average, 25%, we simulate transitions for employed individuals by setting the value corresponding to receiving an offer to a large negative number (i.e., no offer) in 75% of cases. This adjustment prevents overstating employer transitions and their contribution to earnings dynamics.[16]

[15] Simulations indicate that the Nickell bias becomes negligible after 25–30 periods. However, since our decomposition involves regressing earnings on employer-specific fixed effects, we require each employer spell to last at least 25 periods to obtain reliable estimates.

[16] In the SCE, the average self-reported probability of receiving an offer over the following four months for the employed is around 25%. While we do not explicitly incorporate this question in our baseline estimation, we have verified that our results do not change substantially if we were to use it to adjust the self-reported offer probability distributions accordingly.

Variance. We estimate the following three regressions in our simulated data:

$$y_{i,t} = \beta_0^A + \beta_1^A y_{i,t-1} + \beta_2^A y_{i,t-1}^2 + \varepsilon_{i,t}^A \tag{35}$$

$$y_{i,t} = \beta_0^B + \beta_1^B y_{i,t-1} + \beta_2^B y_{i,t-1}^2 + \alpha_i + \varepsilon_{i,t}^B \tag{36}$$

$$y_{i,t} = \beta_0^C + \beta_1^C y_{i,t-1} + \beta_2^C y_{i,t-1}^2 + \alpha_i + \zeta_{ij(t)} + \varepsilon_{i,t}^C \tag{37}$$

We approximate $\text{Var}(y_{i,t+1} \mid y_{i,t})$ by the fitted values of a linear regression of $(\hat{\varepsilon}_{it}^A)^2$ on $(y_{i,t}, y_{i,t}^2)$. Similarly, we approximate $\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i)$ by the fitted values of a linear regression of $(\hat{\varepsilon}_{it}^B)^2$ on $(y_{i,t}, y_{i,t}^2)$, and $\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{ij(t)})$ by the fitted values of a linear regression of $(\hat{\varepsilon}_{it}^C)^2$ on $(y_{i,t}, y_{i,t}^2)$. We then extract the fitted values of these two latter regressions, which we denote $\hat{\kappa}_{it}^B$ and $\hat{\kappa}_{it}^C$ respectively.

We use these to obtain an approximation for $\mathbb{E}_{\alpha_i|y_{i,t}}\left[\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right]$ by obtaining the fitted values of a linear regression of $\hat{\kappa}_{it}^B$ on $(y_{i,t}, y_{i,t}^2)$. Likewise, we approximate $\mathbb{E}_{\alpha_i,\zeta_{ij(t)}|y_{i,t}}\left[\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{ij(t)}) \mid y_{i,t}\right]$ by estimating the fitted values of a linear regression of $\hat{\kappa}_{it}^C$ on $(y_{i,t}, y_{i,t}^2)$.

By obtaining these two objects—and representing them in Figure 4 alongside overall earnings variability—we can compute every term in the decompositions of Equations 31 and 32, both on average and across different points of the $y_{i,t}$ distribution. This figure displays three lines. The red line refers to variance of next period's earning conditional on past earnings and its square as a function of past earnings. The blue line also conditions the variance of next period's earnings on individual fixed-effects. The green line also conditions it on employer-specific match.

The first result from this exercise is that the estimated variance of the earnings process using the SCE data also shows the nonlinearities in the variance of future earnings as a function of current earnings that has been previously documented in the literature using observed earnings (for instance by Guvenen et al. (2021)). The second important point is that our decompositions allows us to determine which aspects of our model generate this non-linearity. The comparison of the red, blue, and green lines, shows that this nonlinearity is due to the effects of switching employers. Furthermore, the employer fixed effect also has a large effect on the level of the variance, which is much lower once one conditions on them. This is an important result showing that most earnings risk

comes from employment status and employer changes.

Figure 4: Risk decomposition. The term ζ represents individual-employer fixed effects, while α represents individual fixed effects

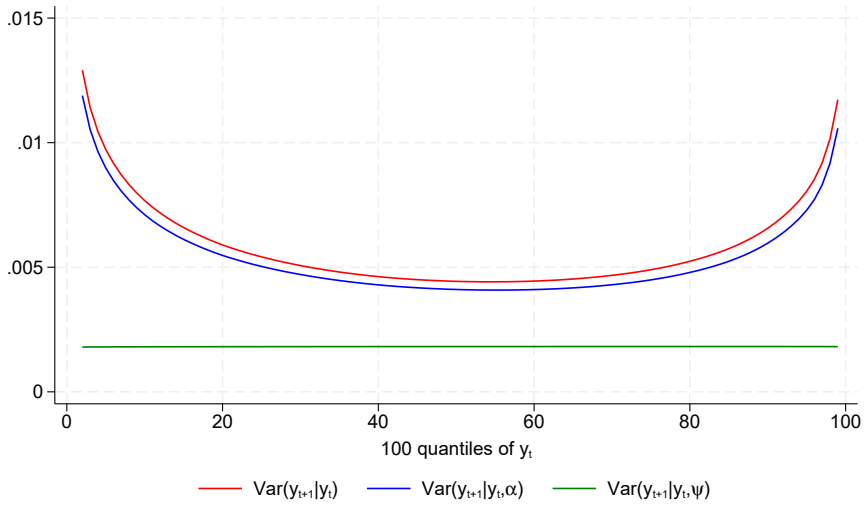


Table 5 shows the contribution of the components in equation (32). It shows that most fluctuations in earnings are related to heterogeneity in the employer-specific components.

Table 5: Risk decomposition, averages of Figure 4 (top) and contributions of individual fixed effects, employer transitions and employer fixed effects, and remainder (bottom).

Risk Measure	Average	Contribution of...	
$Var(y_{t+1}\|y_t)$	0.006	Risk conditional on α and ζ	29.86%
$Var(y_{t+1}\|y_t,\alpha)$	0.006	Heterogeneity in ζ	62.17%
$Var(y_{t+1}\|y_t,\alpha,\zeta)$	0.002	Heterogeneity in α	7.97%

Persistence. To understand the determinants of earnings persistence, we estimate the following three regressions on our simulated data:

$$y_{it+1} = \beta_0^D + \beta_1^D y_{it} + \epsilon_{it+1}^D \tag{38}$$

$$y_{it+1} = \beta_0^E + \beta_1^E y_{it} + \mu_i + \epsilon_{it+1}^E \tag{39}$$

$$y_{it+1} = \beta_0^F + \beta_1^F y_{it} + \mu_i + \upsilon_{ji(t)} + \epsilon_{ijt+1}^F. \tag{40}$$

With β_1^D we obtain an estimate of $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t})}{\partial y_{i,t}}$ in equation (31). The estimated β_1^E, instead, provides us with an approximation of $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\alpha_i)}{\partial y_{i,t}}$ in the same equation. Finally,

with β_1^F we estimate $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\alpha_i,\zeta_{ij(t)})}{\partial y_{i,t}}$ in equation (32). We show the results in Table 6.

We can also perform a similar decomposition as that for the variance, based on equation (35), which we report in the bottom panel of Table 6, which shows that the persistence in employer-specific components also has an important effect in overall earnings persistence. Remarkably, the persistence estimated on (simulated) realized earnings (0.506) is not very different from the estimate of ρ we obtain with subjective expectations data. We also note that, while the data used to generate the estimates in Table 6 are simulated from the model we estimated, "realized" earnings are exempt from measurement error.

Table 6: Slope coefficients (top) and implied contribution of different components to overall persistence (bottom), obtained from simulated data, for different specifications. $T = 100$. The three numbers in the right-hand side panel sum to 0.988.

Parameter	Value	Contribution of...	
β_1^A	0.988	Individual heterogeneity α	0.115
β_1^B	0.873	Individual-employer fixed effects ζ	0.367
β_1^C	0.506	Remaining factors	0.506

8 Conclusions

Expectations play a critical role in economic decision-making. Beyond expected values, the perceived distribution of future outcomes, including uncertainty, shapes individual behavior. While economists were initially reluctant to quantify subjective expectations, recent decades have seen significant advances in their collection and use, spearheaded by scholars such as Manski and collaborators (e.g., Manski (2004)). These efforts have been embraced by central banks and other institutions globally, opening the door to novel analyses of labor market dynamics.

The NY Fed Survey of Consumer Expectations data offer unique richness, capturing perceptions of future earnings distributions under counterfactual scenarios and the probabilities of discrete events. These data allow us to map model parameters to subjective probabilities while accounting for fixed effects and measurement error. By combining subjective data with observed earnings realizations, our framework delivers a nuanced understanding of labor market risks and choices.

This paper shows that subjective expectations data—particularly those eliciting beliefs over future wage offers and acceptance probabilities—can be used to identify and estimate rich models of earnings and employment dynamics under weaker assumptions than traditional approaches. These data provide direct information on counterfactual outcomes, allowing us to overcome key identification challenges posed by selection and unobserved heterogeneity.

We show that subjective beliefs, when appropriately measured, contain sufficient structure to estimate both the latent earnings process and the employment and employer transition dynamics. Our framework links these expectations to structural model components using tractable fixed-effects and GMM estimation methods, avoiding the strong distributional assumptions often required when using realized outcomes alone.

Empirically, we find that the high earnings persistence that we observe comes from lower risk and persistence of shocks than standard reduced-form estimates suggest, but more heterogeneity. We show that standard high persistence and volatility estimates arise from unaccounted employer transitions and heterogeneity, rather than persistent productivity shocks. Similarly, we find that much of the nonlinear relationship between earnings volatility and income disappears after accounting for individual- and employer-match effects. These results underscore the value of the combination of expectations data in combination with a rich model of employment dynamics for analyzing labor market outcomes.

References

Abowd, J. M. and D. Card (1989). On the covariance structure of earnings and hours changes. *Econometrica 57*(2), 411–445.

Altonji, J. G., D. M. Hynsjö, and I. Vidangos (2023). Individual earnings and family income: Dynamics and distribution. *Review of Economic Dynamics 49*, 225–250.

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

Alvarez, J. and M. Arellano (2022). Robust likelihood estimation of dynamic panel data models. *Journal of Econometrics 226*(1), 21–61.

Arellano, M. (2014). Uncertainty, persistence, and heterogeneity: A panel data perspective. *Journal of the European Economic Association 12*(5), 1127–1153.

Arellano, M., O. Attanasio, S. Crossman, and V. Sancibrián (2024, September). Estimating flexible income processes from subjective expectations data: Evidence from india and colombia. Working Paper 32922, National Bureau of Economic Research.

Armantier, O., G. Topa, W. Van der Klaauw, and B. Zafar (2017). An overview of the survey of consumer expectations. *Economic Policy Review* (23-2), 51–72.

Browning, M., M. Ejrnæs, and J. Alvarez (2010). Modelling income processes with lots of heterogeneity. *Review of Economic Studies 77*(4), 1353–1381.

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.

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

Hall, R. E. and F. S. Mishkin (1982). The sensitivity of consumption to transitory income: Estimates from panel data on households. *Econometrica 50*(2), 461–481.

Hospido, L. (2015). Wage dynamics in the presence of unobserved individual and job heterogeneity. *Labour Economics 33*, 81–93.

Koşar, G. and W. Van der Klaauw (2025). Workers' perceptions of earnings growth and employment risk. *Journal of Labor Economics 43*(S1), S83–S121.

MaCurdy, T. E. (1982). The use of time series processes to model the error structure of earnings in a longitudinal data analysis. *Journal of Econometrics 18*(1), 83–114.

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

Meghir, C. and L. Pistaferri (2004). Income variance dynamics and heterogeneity. *Econometrica 72*(1), 1–32.

Meghir, C. and L. Pistaferri (2011). Earnings, consumption and life cycle choices. In *Handbook of labor economics*, Volume 4, pp. 773–854. Elsevier.

Mueller, A. I. and J. Spinnewijn (2023). Expectations data, labor market, and job search. *Handbook of Economic Expectations*, 677–713.

Nickell, S. (1981). Biases in dynamic models with fixed effects. *Econometrica: Journal of the econometric society*, 1417–1426.

Wang, T. (2023). Perceived versus calibrated income risks in heterogeneous-agent consumption models. Technical report, Bank of Canada.

A Model extensions and discussion

A.1 Earnings equations for the unemployed

As we discussed above, we do not need to use the data for the unemployed. However, we can include them in our estimation procedure as follows.

Earnings expectations for the unemployed. We need to modify our approach in two dimensions. First, only equation 16 is relevant, as upon employment, they will draw a new employer-specific match quality. Second, we do not have information on latent current period earnings $y^*_{i,t}$. Since we observe earnings in their last job and their unemployment duration, we can use our model's equations to infer their earnings evolution. For instance, if their unemployment duration is one period, recursive substitutions of equation (15) yield:

$$y^*_{i,t+1} - x'_{i,t+1}\gamma - \mu_i - \upsilon_{i,j(t+1)} = \rho^2 \left(y^*_{i,t-1} - x'_{i,t-1}\gamma - \mu_i - \upsilon_{i,j(t-1)} \right) + \rho\varepsilon^{\omega}_{i,t} + \varepsilon^{\omega}_{i,t+1}$$

and because $\upsilon_{i,j(t+1)} = \phi\upsilon_{i,j(t)} + \epsilon^{\upsilon}_{i,j(t+1)}$ and $\upsilon_{i,j(t-1)} = (1/\phi)\upsilon_{i,j(t)} - (1/\phi)\epsilon^{\upsilon}_{i,j(t)}$ can then re-write equation 18 as follows:

$$\begin{aligned} \overline{\mathbf{y}}^{\mathbf{o_{t+1}}}_{\mathbf{i,t}} &= \rho^2 y^*_{i,t-1} + \left(x_{i,t+1} - \rho^2 x_{i,t-1} \right)' \gamma + \left(1 - \rho^2 \right) \mu_i + \left(\phi - \rho^2/\phi \right) \upsilon_{i,j(t)} + \\ &\quad + (\rho^2/\phi)\varepsilon^{\upsilon}_{i,j(t)} + \rho\varepsilon^{\omega}_{it}. \end{aligned} \tag{41}$$

Future earnings probability distribution for the unemployed. As before, we need to address the unobservability of current latent earnings. Consistent with the previous discussion, we can relate these quantities to the last observed earnings. For instance, if the unemployment spells is one period, we can then write:

$$\begin{aligned} \overline{\boldsymbol{\ell}}^{\mathbf{o_{t+1}}}_{\mathbf{kit}} &= (1/\sigma_u)\, r_{kit} + \left(-\rho^2/\sigma_u \right) y^*_{i,t-1} + \left(x_{i,t+1} - \rho^2 x_{i,t-1} \right)' (-\gamma/\sigma_u) \\ &\quad - \mu_i \left(1 - \rho^2 \right)/\sigma_u - \left(\phi^2 - \rho^2 \right)/\sigma_u \upsilon_{i,j(t-1)} - \left(\phi\varepsilon^{\upsilon}_{i,j(t)} + \rho\varepsilon^{\omega}_{it} \right)/\sigma_u. \end{aligned} \tag{42}$$

Following the same substitutions used to obtain equation 41, and more conveniently

Table 7: Summary of notation, part 1

Objects with a data counterpart	
$y_{i,t}$	Log observed individual earnings
$e_{i,t}$	Employment status
$s_{i,t+1}$	Indicator of job change between t and $t+1$
$\bar{\mathbf{y}}_{\mathbf{it}}^{\mathbf{o_{t+1}}}$	Point expectation of annual salary of best offer
τ_k	Multiplicative factor to form bins for probability distributions
r_{kit}	Upper boundary of offer bin k
$\bar{\mathbf{p}}_{\mathbf{kit}}^{\mathbf{o_{t+1}}}$	Reported probability of offer being in bin k
$\bar{\ell}_{\mathbf{kit}}^{\mathbf{o_{t+1}}}$	Log odds ratio of offer being in bin k
$\mathbf{p}_{\mathbf{kit}}^{\mathbf{e0}}$	Reported probability of staying in same job, based on offer k
$\mathbf{p}_{\mathbf{kit}}^{\mathbf{e1}}$	Reported probability of changing job, based on offer k
$\mathbf{p}_{\mathbf{kit}}^{\mathbf{ue}}$	Reported probability of starting work, based on offer k
$\bar{\ell}_{\mathbf{kit}}^{\mathbf{e1}}$	Log odds ratio of changing job, based on offer k
$\bar{\ell}_{\mathbf{kit}}^{\mathbf{e0}}$	Log odds ratio of staying in same job, based on offer k
$\bar{\ell}_{\mathbf{kit}}^{\mathbf{ue}}$	Log odds ratio of starting work, based on offer k
$x_{i,t+1}$	Additional observables (e.g. employment tenure)
Latent objects	
$y_{i,t}^{*}$	Log latent individual earnings
$y_{i,t+1}^{0*}$	Log latent individual earnings if keeping job between t and $t+1$
$y_{i,t+1}^{1*}$	Log latent individual earnings if changing job between t and $t+1$
$ee_{i,t+1}^{0*}$	Value of employment in the current job
$ee_{i,t+1}^{1*}$	Value of employment in a new job
$\upsilon_{i,t+1}^{0*}$	Match-specific effect if keeping job between t and $t+1$
$\upsilon_{i,t+1}^{1*}$	Match-specific effect if changing job between t and $t+1$

Table 8: Summary of notation, part 2

	Model objects
μ_i	Individual ability fixed effect
σ_μ	Standard deviation of μ
η_i	Individual mobility fixed effect
σ_η	Standard deviation of η
$\omega_{i,t}$	Stochastic individual productivity
ρ	Persistence of ω
ϵ^ω	Shock to ω
$\upsilon_{ij(t)}$	Match-specific effect
σ_υ	Standard deviation of υ
ϕ	Persistence of υ across job changes
ϵ^υ	Shock to υ
σ_e	Standard deviation of $\epsilon^\omega_{i,t+1} + \epsilon^\upsilon_{ij(t+1)}$
σ_u	Counterpart of σ_e for the unemployed
δ^0_y	Effect of current earnings on (log) odds of staying
δ^1_y	Effect of potential earnings on (log) odds of changing job
δ^u_y	Effect of potential earnings on (log) odds of working (for unemployed)
b^0_μ	Effect of μ on log odds of staying
b^1_μ	Effect of μ on log odds of changing job
b^u_μ	Effect of μ on log odds of working (for unemployed)
b^0_η	Effect of η on log odds of staying
b^1_η	Effect of η on log odds of changing job
b^u_η	Effect of η on log odds of working (for unemployed)
γ	Coefficient on x in earnings equations
γ^u	Coefficient of x in transition into employment equation
γ^0	Coefficient of x in staying in same job equation
γ^1	Coefficient of x in job change equation
	Simulation objects
α_i	Individual fixed effect in reduced-form regressions
$\zeta_{ij(t)}$	Individual-job fixed effect in reduced-form regressions
$\beta^{A,B,C}_{0,1,2}$	Coefficients in reduced-form regressions
$\hat{\epsilon}^{A,B,C}$	Fitted values in reduced-form regressions
$\hat{\kappa}^{B,C}$	Fitted values in reduced-form regressions of $\hat{\epsilon}^{B,C}$ on y

for the second step estimation we get:

$$\begin{aligned}\overline{\boldsymbol{\ell}}_{\mathbf{kit}}^{\mathbf{o_{t+1}}} &= (1/\sigma_u)\, r_{kit} + \left(-\rho^2/\sigma_u\right) y^*_{it-1} + \left(x_{i,t+1} - \rho^2 x_{i,t-1}\right)' \left(-\gamma/\sigma_u\right) \\ &\quad -\mu_i \left(1-\rho^2\right)/\sigma_u - \left(\phi - \rho^2/\phi\right)/\sigma_u \upsilon_{i,j(t)} - \left(\rho^2/\phi \varepsilon^{\upsilon}_{i,j(t)} + \rho \varepsilon^{\omega}_{it}\right)/\sigma_u.\end{aligned} \tag{43}$$

A.2 Computing future earnings for job stayers

Computing $\widehat{y}^{0*}_{i,t+1}$. We define the data counterpart for future earnings conditional on staying with the same employer (so that $\upsilon_{i,j(t+1)} = \upsilon_{i,j(t)}$) as $\widehat{\mathbf{y}}^{\mathbf{0*}}_{\mathbf{i,t+1}}$. Hence, we can express it as:

$$\begin{aligned}\widehat{\mathbf{y}}^{\mathbf{0*}}_{\mathbf{i,t+1}} &= E\left(y^{0*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1} = e_{i,t} = 1, s_{i,t+1} = 0\right) \\ &= \rho y^*_{it} + \left(x_{i,t+1} - \rho x_{it}\right)' \gamma + (1-\rho)\,\mu_i + (1-\rho)\,\upsilon_{i,j(t)}\end{aligned} \tag{44}$$

Two survey questions provide information on $\widehat{y}^{0*}_{i,t+1}$. The first explicitly asks for a point estimate of earnings, conditional on maintaining the same employer and working the same number of hours, 12 months ahead. Unfortunately, its conditioning on constant hours and its timing mismatch (12 months instead of 4 as in the rest of the questions) make it less than ideal.

The second question asks about unconditional subjective expected earnings in four months' time. We use it to obtain conditional subjective earnings, assuming the individual keeps the same job. To do so, we begin with the unconditional expectation $\overline{y}_{it}$ and derive the conditional expectation as follows

$$\begin{aligned}\overline{y}_{it} &= E\left(y^{0*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1} = e_{i,t} = 1, s_{i,t+1} = 0\right) \Pr\left(e_{i,t+1} = 1, s_{i,t+1} = 0 \mid \Omega_{it}, e_{i,t} = 1\right) \\ &\quad + E\left(y^{1*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1} = e_{i,t} = 1, s_{i,t+1} = 1\right) \Pr\left(e_{i,t+1} = 1, s_{i,t+1} = 1 \mid \Omega_{it}, e_{i,t} = 1\right)\end{aligned} \tag{45}$$

The first term on the right-hand side represents the 4-month conditional expectation of earnings, $\widehat{y}^{0*}_{i,t+1}$, given that the individual remains with the same employer. This is the term we aim at recovering. While $\widehat{y}^{0*}_{i,t+1}$ is unknown, we can compute it since all

other expectations and probabilities are observable in the data. Also, we account for the possibility of self-employment, where the probability is observed, but the income expectation conditional on self-employment is not. To address this, we exclude self-employed individuals from our estimates.

To mitigate potential measurement error arising from our procedure for generating $\widehat{y}^{0*}_{i,t+1}$—specifically, the fact that the probability of staying in the same job appears in the denominator—we will later instrument it using the 12-month conditional question on staying in the same job, which we described earlier.

A.3 Endogenous selection multinomial model: endogenous self-selection in both υ and ω

Employment to employment transitions For the employed workers, the mapping with the multinomial model probabilities (9)-(8) is not straightforward, even with the $\delta^s_2 = \delta^q_2 = 0$ simplification. The reason is that, although the log odd ratios (11)-(10) only depend on one of $y^{1*}_{i,t+1}$ or $y^{0*}_{i,t+1}$, the probabilities will depend on both $y^{1*}_{i,t+1}$ and $y^{0*}_{i,t+1}$. Let us elaborate on this point. For the currently employed we observe:

$$\begin{aligned} & \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid (k-1)\cdot \overline{y}^{of}_{it} < y^{1*}_{i,t+1} \leq k\cdot \overline{y}^{of}_{it}, \Omega_{it}, e_{i,t}=1\right) \\ \approx \; & \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y^{1*}_{i,t+1} = \overline{k}\cdot \overline{y}^{of}_{it}, \Omega_{it}, e_{i,t}=1\right) \equiv p^1_{(k)i,t} \end{aligned} \tag{46}$$

for $k \in \{0.75, 0.85, 0.95, 1.05, 1.15, 1.25\}$. Equation 46 is the same as equation 21 in the main text.

One interpretation of (46) is as an average of the model probabilities taken with respect to the distribution of $y^{0*}_{i,t+1}$ conditioned on $y^{1*}_{i,t+1} = \overline{k}\cdot \overline{y}^{of}_{it}$. Essentially, given Ω_{it}, fixing a value for $y^{1*}_{i,t+1}$ is equivalent to fixing $\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{ij(t+1)}$. Thus, under this interpretation we observe averaged model probabilities with respect to the distribution of $\varepsilon^{\omega}_{i,t+1}$ conditioned on $\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{ij(t+1)}$:

$$E_{\left(y^{0*}_{i,t+1} | y^{*}_{(1)i,t+1} = k\overline{y}^{of}_{it}\right)}\left[\Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y^{0*}_{i,t+1}, y^{1*}_{i,t+1} = \overline{k}\cdot \overline{y}^{of}_{it}, \Omega_{it}, e_{i,t}=1\right)\right] \tag{47}$$

where

$$\begin{aligned} & \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y^{0*}_{i,t+1}, y^{1*}_{i,t+1}=\overline{k}\cdot\overline{y}^{of}_{it}, \Omega_{it}, e_{i,t}=1\right) \qquad (48) \\ = & \frac{e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\overline{k}\cdot\overline{y}^{of}_{it}\right)+b^q_1\mu_i+b^q_2\eta_i}}{1+e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\overline{k}\cdot\overline{y}^{of}_{it}\right)+b^q_1\mu_i+b^q_2\eta_i}+e^{x^{s\prime}_{i,t+1}\gamma^s+\delta^s_1 y^{0*}_{i,t+1}+b^s_1\mu_i+b^s_2\eta_i}}. \end{aligned}$$

The formulation (47)-(48) could be used to estimate the employment to employment transitions, but the required method would not be a simple linear estimator.

A second interpretation is that the respondent interprets the conditioning in the question as being exclusively about the match component. Namely, that a "best offer of $\overline{k}\cdot\overline{y}^{of}_{it}$" is understood by the respondent as meaning that $\varepsilon^{v}_{ij(t+1)}$ is set to:

$$\varepsilon^{v}_{ij(t+1)}=\overline{k}\cdot\overline{y}^{of}_{it}-\overline{y}^{of}_{it}. \qquad (49)$$

The way the question is interpreted matters not only for us (the researchers) but also for the respondent. If she thinks that the statement of a $\overline{k}\cdot\overline{y}^{of}_{it}$ offer is mostly about her own productivity in $t+1$, there is no reason why she would assign higher probabilities of changing jobs to higher values of k.

Under the second interpretation, we would regard (46) as an average of model probabilities with respect to the unconditional distribution of $\varepsilon^{\omega}_{i,t+1}$ as follows:

$$E_{\varepsilon^{\omega}_{i,t+1}}\left[\frac{e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\overline{k}\cdot\overline{y}^{of}_{it}\right)+\delta^q_1\varepsilon^{\omega}_{i,t+1}+b^q_1\mu_i+b^q_2\eta_i}}{1+e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\overline{k}\cdot\overline{y}^{of}_{it}\right)+\delta^q_1\varepsilon^{\omega}_{i,t+1}+b^q_1\mu_i+b^q_2\eta_i}+e^{x^{s\prime}_{i,t+1}\gamma^s+\delta^s_1\widehat{y}^{0*}_{i,t+1}+\delta^s_1\varepsilon^{\omega}_{i,t+1}+b^s_1\mu_i+b^s_2\eta_i}}\right]. \qquad (50)$$

Another complication is that we also need $\Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid k-1<y^{1*}_{i,t+1}\leq k, \Omega_{it}, e_{i,t}=1\right)$, which is not directly observable, but satisfies the expression:

$$\begin{aligned} & \Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid k<y^{1*}_{i,t+1}\leq k-1, \Omega_{it}, e_{i,t}=1\right) \qquad (51) \\ = & \frac{\Pr\left(k-1<y^{1*}_{i,t+1}\leq k \mid s_{i,t+1}=0, e_{i,t+1}=1, \Omega_{it}, e_{i,t}=1\right)\Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid \Omega_{it}, e_{i,t}=1\right)}{\Pr\left(k-1<y^{1*}_{i,t+1}\leq k \mid \Omega_{it}, e_{i,t}=1\right)} \end{aligned}$$

This quantity would be observable if the first term in the numerator could be constructed from the Bernanke question, which is unclear.[17] The second is the "percent chance of

[17] Letting $\Upsilon_{it}=\{s_{i,t+1}=0, e_{i,t+1}=1, \Omega_{it}, e_{i,t}=1\}$ for shortness, the Bernanke question relates to

being employed at the same employer". Finally, the denominator can be constructed from the best offer probability questions.

Granted the above, for the currently employed we would approximately observe:

$$p^{s}_{(k)i,t} \equiv \Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid y^{1*}_{i,t+1} = \overline{k}\cdot \overline{y}^{of}_{it}, \Omega_{it}, e_{i,t}=1\right), \tag{52}$$

which would still require a mapping with the model probabilities similar to (47)-(48) or (50).

As already pointed out in the main text, a simpler version of the multinomial model would be one in which there is endogenous self-selection in $\varepsilon^{\upsilon}_{ij(t+1)}$ but not in $\varepsilon^{\omega}_{i,t+1}$. In such case, $y^{0*}_{i,t+1}$ would be replaced by $\widehat{y}^{0*}_{i,t+1} = E\left(y^{0*}_{i,t+1} \mid \Omega_{it}\right)$ and there would be no need to consider the expectations in (48) or (50). Thus, having the linear equations:

$$\overline{\ell}^{0}_{kit} = \ln\left(\frac{p^{0}_{(k)i,t}}{1-p^{1}_{(k)i,t}-p^{0}_{(k)i,t}}\right) = x^{0\prime}_{i,t+1}\gamma^{0} + \delta^{0}_{y}\widehat{y}^{0*}_{i,t+1} + b^{0}_{\mu}\mu_i + b^{0}_{\eta}\eta_i \tag{53}$$

$$\overline{\ell}^{1}_{kit} = \ln\left(\frac{p^{1}_{(k)i,t}}{1-p^{1}_{(k)i,t}-p^{0}_{(k)i,t}}\right) = x^{1\prime}_{i,t+1}\gamma^{1} + \delta^{1}_{y}\left(\overline{k}\cdot\overline{y}^{of}_{it}\right) + b^{1}_{\mu}\mu_i + b^{1}_{\eta}\eta_i \tag{54}$$

for $k \in \{0.75, 0.85, 0.95, 1.05, 1.15, 1.25\}$.

B Estimation and empirical implementation

B.1 Obtaining the probabilities of staying and transitioning into unemployment conditional on a given offer

As argued in Section 4, one empirical complication is that we do not directly observe p^{e0}_{kit} nor p^{eu}_{kit}, i.e., the conditional probabilities of staying on the same job or going into unemployment given a certain job offer indexed by k.

However, we do observe the unconditional probabilities of transitioning into a new job, staying on the same job and moving into unemployment over the next four months,

$\Pr\left(c_{j-1} < y^{0*}_{i,t+1} \leq c_j \mid \Upsilon_{it}\right)$. The connection between the *cdf* of $y^{1*}_{i,t+1}$ and the *cdf* of $y^{0*}_{i,t+1}$ is given by:

$$\Pr\left(y^{1*}_{i,t+1} \leq r \mid \Upsilon_{it}\right) = E_{\varepsilon^{\upsilon}_{ij(t+1)}}\left[\Pr\left(y^{0*}_{i,t+1} \leq r - \varepsilon^{\upsilon}_{ij(t+1)} + (1-\phi)\,\upsilon_{ij(t)} \mid \Upsilon_{it}\right)\right].$$

which we denote $\overline{p}_{i,t}^{e1}$, $\overline{p}_{i,t}^{e0}$ and $\overline{p}_{i,t}^{eu}$ respectively. We also know that it must hold that $\overline{p}_{i,t}^{x} = \sum_{j=1}^{J} p_{kit}^{x} \overline{p}_{kit}^{o_{t+1}}$ for $x \in \{e1, e0, ue\}$.

We assume that the probability of unemployment is concentrated in the lowest part of the offer distribution. Namely, we assume that there is a threshold $k*$ such that $p_{k*it}^{eu} = 1 - p_{kit}^{e1}$ for offers $k < k*$ and $p_{kit}^{eu} = 0$ for offers $k > k*$. De facto, because the probability for k is discretised, that means that there is usually one intermediate $k*$ such that $p_{k*it}^{eu} = 1 - p_{k*it}^{e1} - \frac{\overline{p}_{it}^{eu} - \sum_{j=1}^{k*-1} p_{kit}^{eu} \overline{p}_{kit}^{o_{t+1}}}{\overline{p}_{k*it}^{o_{t+1}}}$.

B.2 Variance and covariance structure

B.2.1 General minimisation problem

For both the model with exogenous and endogenous selection, in step 2 we estimate a set of parameters based on the variance-covariance matrix implied by the residuals obtained in step 1. This variance-covariance matrix incorporates both covariances (across equations) and autocovariances (across survey waves).

Let y_{ijt} denote the residuals from step 1 for individual $i = 1, ..., N$, equation $j = 1, ..., J$ and survey wave $t = 1, ..., T$. Let us denote the variance-covariance matrix V_{abc}. This matrix has dimensions $J \times J \times C$, where $C > T$ because we deal with job changers separately. For our specific case, $T = 3$ and $C = 6$. We thus define $c = 1, \ldots, C$ such that $c = 1$ refers to the contemporaneous covariances across equations, $c = 2$ refers to first-order autocovariances for people who stay in the same job, $c = 3$ refers to first-order autocovariances for people who change jobs, $c = 4$ second order autocovariances for people who stay in the same job throughout, $c = 5$ second order autocovariances for people who changed job once and $c = 6$ second-order autocovariances for people who changed job twice.

This implies that our estimates of the different elements of this variance-covariance matrix will be, for example:

$$\hat{V}_{ab1} = \frac{1}{N} \sum_{i=1}^{N} y_{iat} y_{ibt} \tag{55}$$

$$\hat{V}_{ab2} = \frac{1}{N}\sum_{i=1}^{N} y_{iat} y_{ibt+1} \mathbb{I}(jc_{i,t+1} = 0) \tag{56}$$

$$\hat{V}_{ab3} = \frac{1}{N}\sum_{i=1}^{N} y_{iat} y_{ibt+1} \mathbb{I}(jc_{i,t+1} = 1) \tag{57}$$

and so on. We will compute both the values of these variances, covariances and autocovariances (i.e., all elements of V_{abc}) implied by our model and those implied by the SCE data, and our objective criterion is to minimise the distance between both.

Because all of these matrices are symmetric, we can just focus on the lower triangle of all six $J \times J$ matrices.

B.2.2 Sample weights

To weigh the different elements of the variance-covariance matrix, we take into account that each of the elements in $\hat{V}_{abc}$ is effectively a sample mean. We are interested in the variance of each of the elements in $\hat{V}_{abc}$. We denote W the matrix that includes all such variances, i.e., $W_{abc} = var(\hat{V}_{abc})$.

Note that this is different from the variance-covariance matrix of $\hat{V}_{abc}$, which would be a higher-dimensional object which would include not only the variance of each element in $\hat{V}_{ab}$, but also its covariances. Our focus on $\hat{W}_{abc}$ is equivalent to restricting the variance-covariance matrix in our objective function to be diagonal. This choice also has the desirable implication that the implied matrix is both invertible and positive semidefinite because it is diagonal and all of its elements are non-negative.

To see this, think of S_i as a vector of all the moments that we aim to match. This is formed, as described before, of the lower triangle of each of the slices of the matrix $\hat{V}_{abc}$, i.e., it contains $J \times (J+2) \times C/2$ elements. The diagonal of the variance-covariance matrix for S_i is the lower triangle of each of the slices of the matrix $\hat{W}_{abc}$, which has the same number of elements.

B.2.3 Using sample fourth moments

Because $\hat{V}_{abc}$ is composed of sample means, the variance of each of its elements is $W_{abc} = \frac{1}{N} var(V_{abc})$. We compute (with slight abuse of notation, treating c as if it was just time

periods rather than the combination of time periods and job changes):

$$\hat{W}_{abc} = \frac{1}{N}\widehat{var(V_{abc})} = \frac{1}{N^2}\sum_{i=1}^{N}(y_{iat}y_{ibt+c} - \hat{V}_{abc})^2 \tag{58}$$

B.2.4 Using fourth moments under normality

For numerical stability, in our current results we do not use sample fourth moments, but rather what those would be under a normality assumption. Namely,

$$\hat{W}_{abc} = 2\hat{V}_{abc}^2 + \sqrt{\hat{V}_{aa1}}\sqrt{\hat{V}_{bbc}} \tag{59}$$