Subjective Earnings and Employment Dynamics

Manuel Arellano, Orazio Attanasio, Margherita Borella,

Mariacristina De Nardi, and Gonzalo Paz-Pardo*

March 30, 2026

Abstract

We develop a new approach to estimating earnings, job, and employment dynamics using subjective expectations data from the NY Fed 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 of the individual productivity component 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.

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

*Arellano: CEMFI, Madrid; Attanasio: Yale University, MeToD@UNICAL, NBER and CEPR; Borella: Università di Torino, CeRP-Collegio Carlo Alberto and CEPR; De Nardi: University of Minnesota, Federal Reserve Bank of Minneapolis, CEPR and NBER; Paz-Pardo: European Central Bank.
Arellano gratefully acknowledges financial support from grant PID2024-158202NB-I00 funded by MICIU/AEI/10.13039/501100011033 and ERDF/EU. We thank Andrew Caplin, Tullio Jappelli, Gizem Kosar, Søren Leth-Petersen, Wilbert van der Klaauw, and seminar participants at several institutions for useful comments and suggestions. 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 identify and estimate, under a minimal set of assumptions, a structural model addressing these questions. These data include information not only on expected future earnings but also on beliefs about potential job offers and the probability of accepting them, opening new identification strategies and estimation methods that overcome the long-recognized limitations of approaches based solely on observed earnings and job transitions.

Relying only on observed earnings realizations and labor market dynamics to give empirical content to the types of models we consider leads to identification and estimation difficulties. First, estimates of earnings risk and persistence may be biased by unobserved heterogeneity, such as 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, standard estimators confound risk with endogenous sorting. Third, realized outcomes, particularly discrete ones, must often be observed over long panels to separately identify latent factors like ability and match quality from transitory noise. This requirement poses practical limitations, especially when the available data span only a few periods per individual.

The existing literature on earnings dynamics tackles these challenges by imposing strong distributional assumptions and exclusion restrictions. We show that a specific type of subjective expectations data can help address all three problems. In particular, we use information on the probability distribution of potential offers individuals might receive to inform the latent earnings distribution, and conditional subjective probabilities of accepting offers to characterize the selection mechanism. Because individuals report their beliefs about earnings and choices under counterfactual scenarios, these data provide direct information on both the latent outcome process and the decision rules governing

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

selection. This information allows us to achieve identification even with short panels by leveraging rich cross-sectional variation in expectations rather than requiring long individual earnings histories. Using these data, we identify and estimate a structural model of earnings, employment, and job transitions, closely related to Altonji et al. (2013), under weaker assumptions and with considerably simpler estimation methods than approaches based solely on realized outcomes.

The model we consider 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's parameters using methods that are both transparent and tractable.

Our identification strategy exploits the timing and content of subjective expectations data on potential job offers and acceptance decisions. Reported expectations are elicited before future outcomes are realized and refer to counterfactual offers that individuals may or may not accept, rather than to earnings conditional on employment. As a result, these expectations reflect individuals' beliefs about 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 potential offers and counterfactual scenarios, including offers they may ultimately reject, the model can condition directly on their information set. This setup avoids selection bias: the residuals in subjective expectations equations we derive from the model do not correlate with unobserved determinants of choices or outcomes.

By contrast, in observational data, selection bias arises when unobserved shocks jointly affect the outcome and the probability that the outcome is observed. For example, a productivity shock may influence both earnings and the decision to work. If such shocks are omitted, the error term in an outcome equation becomes correlated with the selection mechanism, leading to biased estimates. This distinction is central to our identification strategy. By eliciting beliefs about potential offers and acceptance probabilities, rather than earnings conditional on employment, subjective expectations allow us to observe the information that typically drives selection, thereby avoiding the omitted-variable problem that plagues purely observational analyses.

The model we consider is very similar to that proposed by Altonji et al. (2013) and we estimate it 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 about counterfactual scenarios faced by workers 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.

Relating our findings to the existing literature highlights several implications. Our model and estimates clarify the determinants of the high earnings persistence documented using both observed earnings and subjective expectations data. We find that this persistence is driven primarily by substantial individual heterogeneity and employer-match effects, rather than by highly persistent productivity shocks.

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 13% and varies systematically with earnings, consistent with previous studies. Once we remove person- and match-specific heterogeneity, volatility falls to 7.7% and becomes approximately constant across the earnings distribution. This pattern indicates that much of the observed nonlinearity in earnings volatility reflects unobserved heterogeneity and job transitions rather than individual productivity risk.

These results illustrate how subjective expectations data provide new insights into labor market dynamics. While our paper is not the first to use such data to estimate a well-specified economic model, our contribution differs in focus and scope. Faberman et al. (2022) use the same dataset to estimate a search model and study how search effort translates into outcomes for employed and unemployed workers. We focus instead on earnings dynamics and show how subjective expectations data, together with experimentation within the survey, enable identification under weaker assumptions and transparent estimation of models that are difficult to identify using realized outcomes alone.

To summarize, our contributions are three, two methodological and one empirical. First, we establish identification of a model with employment and job risk under weaker assumptions by using subjective expectations on counterfactual offers and acceptance decisions, which provide direct information on latent outcomes and selection. Second, building on this identification strategy, we develop econometric techniques, linking subjective expectations data to structural models, that allow a 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 in short panels, a task that is challenging when only using realized outcome data. Third, empirically, 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.

2 Related Literature

A large empirical literature studies earnings and employment dynamics using panel data on realized outcomes, with the goal of characterizing income risk, persistence, and heterogeneity over the life cycle. Early contributions focused on univariate earnings processes and documented substantial earnings volatility and a pronounced increase in cross-sectional dispersion with age (e.g., MaCurdy (1982); Abowd and Card (1989); Meghir and Pistaferri (2004)). These findings motivated parsimonious representations of earnings dynamics combining persistent and transitory components, which remain widely used in applied work.

Subsequent research emphasized that earnings dynamics are tightly linked to employment histories and job transitions. Wage growth and earnings volatility depend strongly on job-to-job mobility, unemployment spells, and changes in employer–match quality (Topel and Ward (1992); Altonji et al. (2013)). Structural models incorporating job search and employment risk show that explicitly modeling mobility is essential for interpreting earnings variation and measuring income risk (Low, Meghir, and Pistaferri, 2010). Despite this recognition, many empirical analyses continue to estimate earnings processes independently of employment transitions or impose restrictive assumptions on how job changes affect earnings dynamics.

A recurring finding in this literature is that estimates of earnings persistence are sensitive to how unobserved heterogeneity is modeled. Specifications that restrict heterogeneity to low-dimensional or parametric forms tend to generate higher persistence estimates, while allowing for richer forms of individual heterogeneity typically leads to lower persistence (e.g., Guvenen (2009); Browning et al. (2010); Hospido (2015); Alvarez and Arellano (2022)). From an econometric perspective, this sensitivity reflects the difficulty of disentangling persistent shocks from time-invariant heterogeneity and match effects when relying on realized earnings alone, particularly in short panels.

Recent econometric work has made this identification challenge explicit. Arellano and Bonhomme (2012) analyze the identification trade-offs between unobserved heterogeneity and persistence, while Arellano et al. (2017) show how nonlinear income dynamics can generate heterogeneous persistence and risk, even in the absence of ex ante heterogeneity. In such settings, reduced-form linear measures of persistence may partly reflect heterogeneous responses rather than genuinely persistent shocks.

The direct elicitation of expectations through survey instruments has become increasingly common in economics over the past several decades. Early empirical work, such as Dominitz and Manski (1997), demonstrated the feasibility of eliciting subjective probabilities in surveys. Manski (2004) provided a methodological framework for measuring and interpreting probabilistic expectations, clarifying how subjective belief data can be incorporated into economic analysis. A growing literature has since incorporated expectation data across diverse settings. This body of work—surveyed comprehensively by Koşar and O'Dea (2022)—includes studies such as Delavande and Rohwedder (2008), among many others. More recent contributions continue to expand the scope of this approach: Caplin et al. (2024) combine Danish administrative records with survey-based measures of expectations, while Bachmann et al. (2023) employ the same dataset used in our analysis.

Our approach complements this literature by exploiting subjective expectations data from the New York Fed's Survey of Consumer Expectations (SCE). The SCE, in particular, has been used to study topics ranging from inflation expectations to job search behavior (Faberman et al., 2022). We contribute to this literature by using subjective expectations to study earnings dynamics within a framework that explicitly incorporates employment and job transitions.

Methodologically, our approach is closest to Arellano et al. (2024), who use linear moment-based methods to study household income dynamics. That paper shows how models with individual fixed effects can be estimated using short panels while avoiding biases that arise in nonlinear specifications. We extend this line of work to individual earnings with unemployment and job transitions and show how subjective expectations data allow us to estimate earnings persistence, match effects, and individual heterogeneity using simple linear methods, even in the presence of endogenous employment and mobility. A key feature of our framework, which plays a central role in our empirical results, is that it allows for rich fixed effects without imposing parametric assumptions on their distribution or relying on exclusion restrictions.

3 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. 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.[3]

Our sample spans from March 2014 through November 2019. Table 10 in Appendix A provides some demographics and summary statistics from our sample.

We use both observed and subjective variables, including annual labor market income, tenure, job type, part-time or full-time status, sector, and standard demographics (age, state, marital status, education, etc.). Most demographic characteristics are time-invariant over the sample period and are therefore absorbed by individual fixed effects in the estimation; they are reported in Table 10 in Appendix A to describe the sample. While the SCE elicits a broad range of subjective expectations, in what follows we focus on a subset of questions that admit a direct linear representation linking subjective

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

[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.

expectations to potential earnings and acceptance decisions.

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, subjective expectations data from the SCE provide direct information about both the latent outcome process and the decision rule, allowing us to rely on rich cross-sectional variation rather than extended time series data.

3.1 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 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 $Y_{i,t}^{c_{t+1}}$.

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% * $Y_{i,t}^{c_{t+1}}$] dollars: ... %*
- *Between [80% * $Y_{i,t}^{c_{t+1}}$] and [90% * $Y_{i,t}^{c_{t+1}}$]dollars: ... %*
- *Between [90% * $Y_{i,t}^{c_{t+1}}$] and [100% * $Y_{i,t}^{c_{t+1}}$]dollars: ... %*
- *Between [100% * $Y_{i,t}^{c_{t+1}}$] and [110% * $Y_{i,t}^{c_{t+1}}$]dollars: ... %*

- *Between [110% * $Y_{i,t}^{c_t+1}$] and [120% * $Y_{i,t}^{c_t+1}$]dollars: ... %*
- *More than [120% * $Y_{i,t}^{c_t+1}$] dollars: ... %*

Respondents allocate probabilities across six bins defined relative to $Y_{i,t}^{c_t+1}$: less than 80% of $Y_{i,t}^{c_t+1}$, 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%.

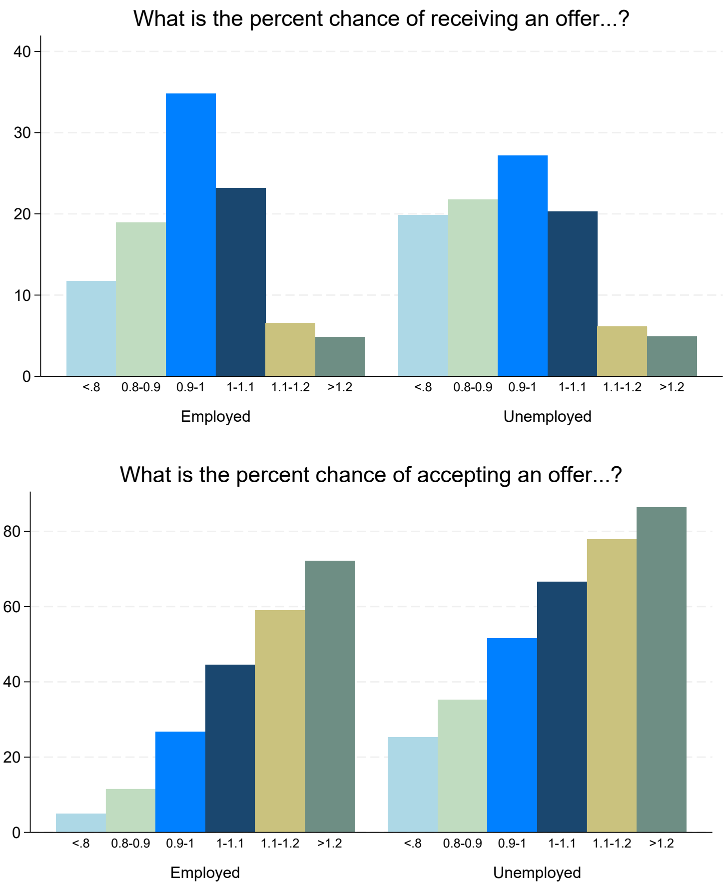
Conditional probabilities of accepting job offers. *Think again about the job offers that you may receive within the coming four months. What do you think is the percent chance that you will accept the job if it offers an annual salary for the first year of...*

- *Less than [80% * $Y_{i,t}^{c_t+1}$] dollars: ... percent chance*
- *Between [80% * $Y_{i,t}^{c_t+1}$] and [90% * $Y_{i,t}^{c_t+1}$]dollars: ... percent chance*
- *Between [90% * $Y_{i,t}^{c_t+1}$] and [100% * $Y_{i,t}^{c_t+1}$]dollars: ... percent chance*
- *Between [100% * $Y_{i,t}^{c_t+1}$] and [110% * $Y_{i,t}^{c_t+1}$]dollars: ... percent chance*
- *Between [110% * $Y_{i,t}^{c_t+1}$] and [120% * $Y_{i,t}^{c_t+1}$]dollars: ... percent chance*
- *More than [120% * $Y_{i,t}^{c_t+1}$] 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 $Y_{i,t}^{c_t+1}$, with the leftmost bar corresponding to offers below 80% of the reference value, the next to 80–90%, and so on. Because $Y_{i,t}^{c_t+1}$ 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 $Y_{i,t}^{c_t+1}$ 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.

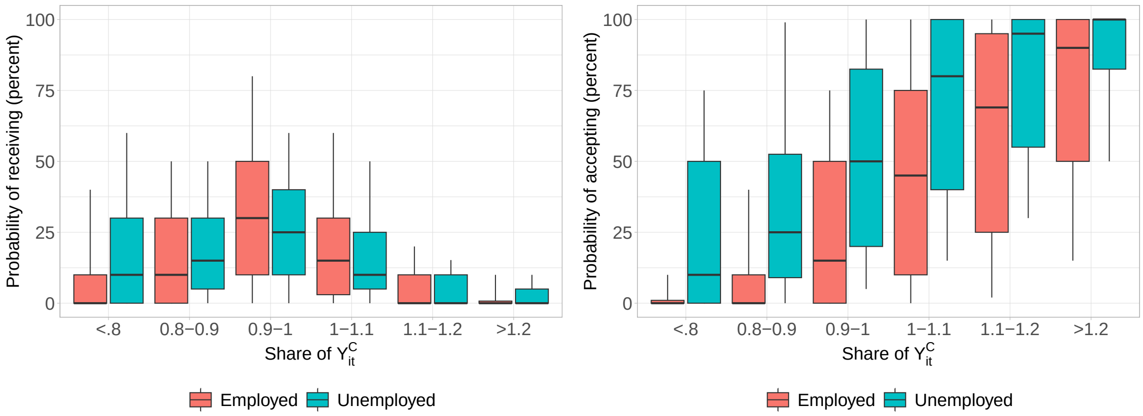
Figure 1: Expected offers (top) and probability of accepting the offer (bottom).



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 $Y_{i,t}^{c_{t+1}}$ (the expected best offer).

Figure 2 plots the cross-sectional distribution of responses across the bins that cover the domain of the offer distribution. 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 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.

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 $Y_{i,t}^{c_{t+1}}$; 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).



3.2 Variation in subjective responses that informs identification

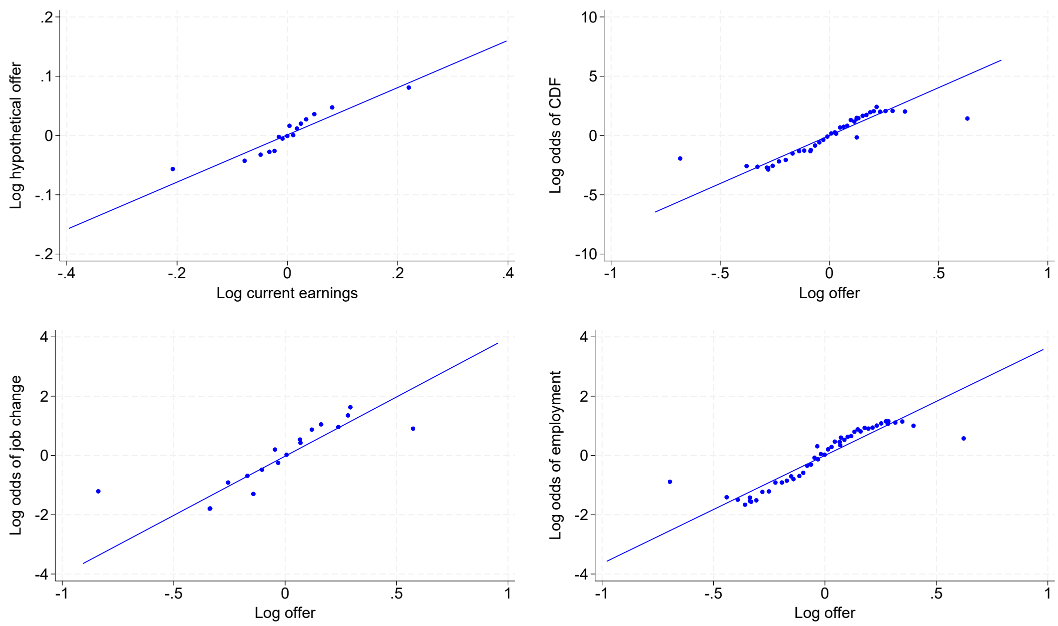
The heterogeneity in responses documented above, across bins and across individuals, is central to identifying the structural components of our model. Figure 3 illustrates four reduced-form relationships in the data, each linking a specific subjective expectation to either current earnings or hypothetical wage offers. As discussed below, these relationships provide identifying variation for the different components of the model.

Before describing the panels in the figure, it is useful to highlight a distinctive feature of our data. The survey not only asks respondents what they expect to happen, but also how they would behave under counterfactual scenarios. For example, both employed and unemployed individuals report the probability that they would accept job offers at various hypothetical wage levels, each of which they typically assign a positive probability of occurring. These questions generate variation similar to that produced by experimental designs and allow us to observe decision rules over a range of counterfactual earnings, rather than inferring them from realized transitions alone.

The top left panel of Figure 3 shows how expected future earnings covaries with 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 vary with the level of the hypothetical offer. The log-odds transformation 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 therefore informative: a steeper slope corresponds to more concentrated beliefs, while a flatter slope reflects greater uncertainty. Accordingly, the inverse of the slope provides a reduced-form measure of *subjective earnings risk*.

Figure 3: Key elements for model identification. Binned scatter plots (randomized quantile bins) with overlaid linear fits estimated on the raw data. All variables are demeaned. Top left: persistence — log expected best offers on log current earnings. Top right: risk — log odds of the subjective CDF of expected offers plotted against log offer. Bottom left: log odds of changing employer (conditional on employment) by log offer. Bottom right: log odds of becoming employed (conditional on unemployment) by log offer. The log offer corresponds to the log midpoint of each elicited offer bin, defined as a multiplicative factor of the individual's expected best offer.



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 report a higher 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 vary with the level of the hypothetical offer. As with employed respondents, unemployed individuals are more likely to accept better offers. This pattern provides information about reservation wage behavior among the unemployed and contributes to identifying *offer-dependent selection into employment*.

Together, these four reduced-form patterns show that individuals report meaningful, monotonic behavioral responses to counterfactual wage offers and that subjective expectations vary systematically with both realized outcomes and hypothetical conditions. This variation provides the identifying information needed to recover the persistence, risk, and selection components of our structural model.

4 A model of earnings and employment transitions

We estimate a version of the framework developed by Altonji, Smith, and Vidangos (2013) (ASV). It combines an earnings determination component with an employment and job transition component. The latter governs both employment selection and employer changes and is central to shaping earnings dynamics.

We depart from ASV along several 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, while 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, which leverages subjective expectations, allows us to estimate employment and job transitions jointly using a multinomial specification in a way that is simple, transparent, and reliable.

Before turning to the details of the model, it is useful to note that the job matching process is a central component. For employed individuals, employment dynamics therefore feature three possible outcomes: remaining in the same job, changing jobs, or becoming unemployed. There are two natural ways to model this process. One is a multinomial specification in which all three outcomes are determined simultaneously. The other is a sequential specification in which individuals first decide whether to remain employed and, conditional on remaining employed, whether to change jobs. We adopt the multinomial specification, which we find more plausible in this context.[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:

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

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

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

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

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 fixed effect, μ_i, and a persistent shock, $\varepsilon^{\omega}_{i,t+1}$ (both unobserved to the econometrician), and a match-specific effect, $\upsilon_{i,j(t+1)}$. 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 $\upsilon_{i,j(t+1)}$ 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, $\varepsilon^{\upsilon}_{i,j(t+1)}$, which becomes operative when the employer-switching indicator $s_{i,t+1}$ equals 1.[5] When there is no employer change, match quality remains fixed. When an employer change occurs,

[4] ASV also view the multinomial specification as conceptually attractive. In their application, however, they focus on a sequential specification because the multinomial model proves numerically unstable. The tractability of our approach allows us to estimate the multinomial model directly.

[5] ASV model hours and wages separately, which we combine into earnings due to data limitations. They also include additional terms in equation (2): a constant when unemployed and a separate constant when previously unemployed, intended to capture human capital depreciation. We omit these terms here for expositional simplicity; however, our methodology readily allows for heterogeneous drifts in the productivity process for employed and unemployed individuals.

match quality evolves with persistence parameter ϕ. The shocks $\varepsilon^{\omega}_{i,t+1}$ and $\varepsilon^{\upsilon}_{i,j(t+1)}$ are independent and identically distributed and independent of each other.

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 $\varepsilon^{\upsilon}_{i,j(t+1)}$, as well as by other exogenous employment and mobility shocks. The match shock, $\varepsilon^{\upsilon}_{i,j(t+1)}$, 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: whether to be employed or not and, if employed, whether to remain with their current employer or switch to a new one.

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

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

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

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 $\varepsilon_{i,t+1} = \left(\varepsilon^{\omega}_{i,t+1}, \varepsilon^{\upsilon}_{i,j(t+1)}\right)$ and $\Omega_{it} = \left(y^{*}_{i,t}, x_{i,s}, \mu_i, \upsilon_{i,j(t)}\right))$ gives rise to the multinomial logit model. We define the probabilities of staying in the same job $p^{0}_{i,t}$ and

[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.

of changing jobs $p^1_{i,t}$ as follows:

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

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

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

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

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

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

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

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 unemployed 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}.

5 Using subjective expectations to identify the model

The model discussed in Section 4 consists of two main blocks. The first characterizes the evolution of latent earnings, which depend on an individual fixed effect, a stochastic productivity component, and employer–match quality. The second block governs employment and job transitions. Employed individuals face three possible outcomes in the next period: remaining with the current employer, switching to a new employer, or becoming unemployed. Unemployed individuals face two outcomes: remaining unemployed or becoming employed.

Our identification and estimation strategy relies on matching model-implied expectations to their empirical counterparts. We first apply this principle to expected values and to the distribution of best (counterfactual) offers, and then extend it to employment and unemployment probabilities.

5.1 Latent earnings and wage offers

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:

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

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 $\upsilon_{i,j(t+1)}$.

In typical survey datasets, realized earnings $y_{i,t+1}$ are observed only for individuals employed in period $t+1$, generating endogenous selection and associated identification challenges. Moreover, dependence on non–strictly exogenous regressors, such as lagged earnings $y_{i,t}$, and the presence of unobserved heterogeneity μ_i complicate the identification of true state dependence and other dynamic effects. Finally, the inclusion of a job-specific component poses additional challenges for estimation.

Potential offers. As discussed in Section 3, the SCE contains information on both point predictions and the full probability distribution of potential yearly salary offers an individual might receive over the next four months, for both employed and unemployed individuals. We begin by describing the implications of these data for point predictions and then extend the analysis to the full distribution, which plays a central role in identification.

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:

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

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

Where we use the fact that, for job changers, $\upsilon_{i,j(t+1)} = \phi\upsilon_{i,j(t)} + \varepsilon^{\upsilon}_{i,j(t+1)}$. Our approach exploits the observability in the SCE of the expected potential (counterfactual) earnings in $t+1$ with a new employer, which in logs we denote with $y^{c_{t+1}}_{i,t}$, as well as several points of its probability distribution. The parameters of interest can be identified using only the information on the probability distribution of potential offers.[7] For expositional purposes we start by equating $y^{c_{t+1}}_{i,t}$ 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:

$$\begin{aligned} y^{c_{t+1}}_{i,t} &= E\left(y^{1*}_{i,t+1} \mid \Omega_{i,t}\right) + \xi^{yc}_{i,t} \\ &= \rho y^{*}_{i,t} + \left(x_{i,t+1} - \rho x_{i,t}\right)' \gamma + (1-\rho)\,\mu_i + (\phi-\rho)\,\upsilon_{i,j(t)} + \xi^{yc}_{i,t}, \end{aligned} \quad (18)$$

where $\xi^{yc}_{i,t}$ is an elicitation error, assumed to be mean-independent of $\Omega_{i,t}$. The term $(1-\rho)\mu_i + (\phi-\rho)\upsilon_{i,j(t)}$ is time invariant over the current job and is therefore observationally equivalent to a composite individual–job fixed effect. As we discuss below, identification of its structural components exploits variation across equations and job transitions.

A key advantage of our approach is that we observe the distribution of latent earnings for all individuals, regardless of whether they work or change employers in the subsequent period. By contrast, the common practice of using realized earnings in $t+1$ restricts the sample to those who are employed both at t and $t+1$. As a result, our specification avoids the selection bias that typically arises when estimation relies solely on realized earnings.

[7] In the empirical application, we use the logarithm of the point expectation, $y^{c_{t+1}}_{i,t} = \log(Y^{c_{t+1}}_{i,t})$, in estimation. However, identification of the parameters of interest does not rely on interpreting it as the mean of the offer distribution and can be achieved using only information on the probability distribution of potential offers, with $y^{c_{t+1}}_{i,t}$ serving as an anchor for that distribution.

This allows us to obtain consistent estimates of the parameters in equation (18) using data on currently employed individuals only, since in our setting $e_{i,t} = 1$ represents an exogenous selection with respect to shocks in $t + 1$.

Estimating equation (18) requires observing $y^*_{i,t}$, which coincides with current earnings for currently employed individuals. The model also has implications for unemployed individuals. However, their *current* latent earnings are unobserved. Furthermore, the SCE contains only a small sample of unemployed respondents.[8] For this reason, we drop the equations for the unemployed from the estimation, while reporting them for completeness in Appendix B.1. This exclusion does not introduce selection bias.

Future earnings probability distribution for the employed. As discussed in Section 3, in addition to point expectations, the SCE reports the probability distribution of the "annual salary of the best offer likely to be received within the next four months," expressed in six 10–percentage-point bins ranging from less than 80% to more than 120% of the expected offer. We interpret these responses as noisy measurements of the conditional cumulative distribution function (CDF) of future latent earnings evaluated at a set of thresholds $R_{ki,t}$. The survey specifies five cutoffs, corresponding to $0.8, 0.9, 1.0, 1.1$, and 1.2 times the expected offer, which define six bins, indexed with k. We express the survey cutoffs in dollars as $R_{ki,t} = \tau_k \exp(y^{c_{t+1}}_{i,t})$, and, since our model is written in logs, we work with the implied log cutoffs $r_{ki,t} \equiv \log R_{ki,t} = y^{c_{t+1}}_{i,t} + \log(\tau_k)$.

Let $p^{c_{t+1}}_{ki,t}$ denote the reported probability that (log) future latent earnings with a new employer in $t + 1$ do not exceed threshold $r_{ki,t}$. We assume additive measurement error in the observed log-odds transformation of the CDF, $\xi^{lc}_{ki,t}$, so that[9]:

$$\text{logit}\big(p^{c_{t+1}}_{ki,t}\big) = \text{logit}\Big(\Pr\big(y^{1*}_{i,t+1} \le r_{ki,t} \mid \Omega_{i,t}\big)\Big) + \xi^{lc}_{ki,t}. \tag{19}$$

Substituting the earnings process in equation (16) into the conditional probability in

[8]For this small sample we can use previous earnings, which slightly changes the latent earnings equation, including a term with higher powers of ρ.

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

(19) yields:

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

Assuming that the composite innovation $\varepsilon_{i,t+1}^{\omega} + \varepsilon_{i,j(t+1)}^{\upsilon}$ follows a logistic distribution with scale parameter σ_e implies that the log-odds transformation of the conditional CDF is linear in the underlying index. As a result, defining $\ell_{ki,t}^{c_{t+1}} \equiv \text{logit}\left(p_{ki,t}^{c_{t+1}}\right)$, we obtain:

$$\begin{aligned} \ell_{ki,t}^{c_{t+1}} = \; & (1/\sigma_e)\, r_{ki,t} + (-\rho/\sigma_e)\, y_{i,t}^{*} + \left(x_{i,t+1} - \rho x_{i,t}\right)' (-\gamma/\sigma_e) \\ & -\mu_i\,(1-\rho)\,/\sigma_e - (1/\sigma_e)\,(\phi-\rho)\,\upsilon_{i,j(t)} + \xi_{ki,t}^{lc} \end{aligned} \tag{20}$$

Here, σ_e denotes the scale parameter of the logistic distribution of $\varepsilon_{i,t+1}^{\omega} + \varepsilon_{i,j(t+1)}^{\upsilon}$.[10] The parameters of interest can be identified using only the probability distribution of best offers (equation 20), with $y_{i,t}^{c_{t+1}}$ anchoring the probability distribution. As with expected latent earnings, we derive the corresponding probability distributions for unemployed individuals in Appendix B.1.

5.2 Employment transitions

In the previous section, we describe how the SCE uses respondents' expected best offers to construct potential offer bins, with respondents reporting the probability of receiving an offer in each bin. The SCE also asks respondents to report the probability, in percent, of accepting an offer in each bin. Together, these responses allow us to estimate transition probabilities between labor market states.

In the model, transitions are determined by comparing the values associated with different labor market statuses, as in equations (6) and (7), where the value of unemployment is normalized to zero. We therefore map the transition probabilities observed

[10] While not pursued in our current implementation of estimators, a sufficiently flexible specification of the coefficients in equation (20) as functions of $r_{ki,t}$ can capture a broad class of probability distributions for $\varepsilon_{i,t+1}^{\omega}+\varepsilon_{i,j(t+1)}^{\upsilon}$—including distributions exhibiting skewness—regardless of the logistic transformation used here (see Arellano et al. (2024)).

in the data to their model counterparts as follows:[11]

$$
\begin{aligned}
p^{ue}_{ki,t} &= \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid r_{k-1,i,t} < y^{1*}_{i,t+1} \le r_{ki,t},\ \Omega_{i,t}, e_{i,t}=0\right), \qquad (21)\\
p^{e1}_{ki,t} &= \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid r_{k-1,i,t} < y^{1*}_{i,t+1} \le r_{ki,t},\ y^{0*}_{i,t+1},\ \Omega_{i,t}, e_{i,t}=1\right),\\
p^{e0}_{ki,t} &= \Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid r_{k-1,i,t} < y^{1*}_{i,t+1} \le r_{ki,t},\ y^{0*}_{i,t+1},\ \Omega_{i,t}, e_{i,t}=1\right),\\
&\qquad k=1,\dots,6,
\end{aligned}
$$

where $r_{0,i,t} \equiv -\infty$ and $r_{6,i,t} \equiv +\infty$.

The term $p^{ue}_{ki,t}$ denotes the probability that an unemployed individual becomes employed (with $1-p^{ue}_{ki,t}$ denoting the probability of remaining unemployed). The terms $p^{e1}_{ki,t}$ and $p^{e0}_{ki,t}$ denote, for currently employed individuals, the probabilities of switching employer and remaining with the same employer, respectively. The probability of becoming unemployed is thus $1-p^{e1}_{ki,t}-p^{e0}_{ki,t}$.[12]

To implement this approach, we approximate the quantities in equation (21) by attributing the probability mass in each interval to the midpoint of the corresponding bin for interior points, and to 0.7 or 1.3 for the boundary bins. Since the midpoints are defined in levels, if $\tilde{R}_{ki,t}$ denotes the midpoint of the interval $(R_{k-1,i,t}, R_{ki,t})$, we define the corresponding log midpoint as $\tilde{r}_{ki,t} = \log(\tilde{R}_{ki,t})$. For example, for the unemployed, we proceed as follows:

$$
p^{ue}_{ki,t} \approx \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y^{1*}_{i,t+1} = \tilde{r}_{ki,t}, \Omega_{i,t}, e_{i,t}=0\right). \qquad (22)
$$

where $\tilde{r}_{ki,t}$ is the representative point for each interval.

Next, we take the log-odds ratios and use our model to obtain the following linear specifications, which are the empirical equivalent of equation (13):

$$
\begin{aligned}
\ell^{ue}_{ki,t} &= \ln\left(\frac{p^{ue}_{ki,t}}{1-p^{ue}_{ki,t}}\right) = x^{u\prime}_{i,t+1}\gamma^u + \delta^u_y \tilde{r}_{ki,t} + b^u_\mu \mu_i + b^u_\eta \eta_i + \xi^{ue}_{ki,t}, \qquad (23)\\
\ell^{e1}_{ki,t} &= \ln\left(\frac{p^{e1}_{ki,t}}{1-p^{e1}_{ki,t}-p^{e0}_{kit}}\right) = x^{1\prime}_{i,t+1}\gamma^1 + \delta^1_y \tilde{r}_{ki,t} + b^1_\mu \mu_i + b^1_\eta \eta_i + \xi^{e1}_{ki,t}, \qquad (24)\\
\ell^{e0}_{ki,t} &= \ln\left(\frac{p^{e0}_{ki,t}}{1-p^{e1}_{ki,t}-p^{e0}_{ki,t}}\right) = x^{0\prime}_{i,t+1}\gamma^0 + \delta^0_y \widehat{y}^{0*}_{i,t+1} + b^0_\mu \mu_i + b^0_\eta \eta_i + \xi^{e0}_{ki,t}. \qquad (25)
\end{aligned}
$$

[11]Subject to additive measurement error in the logit of the observed probabilities, as before. We index bins using the same index k as for thresholds, so that bin k corresponds to the interval $(r_{k-1,i,t}, r_{ki,t}]$.

[12]While $p^{ue}_{ki,t}$ and $p^{e1}_{ki,t}$ are directly observed, $p^{e0}_{ki,t}$ is not, but can be recovered. Further details are provided below.

Here, four points are worth noting. 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) affect the probabilities of switching employers or staying with the current one, the properties of the logistic model and our functional form assumptions imply that only one of the two enters the log-odds ratios.

Second, and very importantly, because in the SCE individuals report the probability of becoming employed if unemployed ($p^{ue}_{ki,t}$) or of changing employer if employed ($p^{e1}_{ki,t}$) conditional on various hypothetical offered earnings, the dependence of transition probabilities on $y^{1*}_{i,t+1}$ is identified thanks to the exogenous variation in the term $\tilde{r}_{ki,t}$ in equations (23) and (24), which is directly observed in the data.

Third, because the SCE does not elicit the probability of staying with the same employer conditional on receiving a job offer, $p^{e0}_{ki,t}$, we recover it using observed unconditional probabilities. Specifically, we use the unconditional probabilities of changing jobs, remaining with the same employer, and transitioning into unemployment. We assume that individuals who reject relatively low offers are more likely to transition into unemployment, whereas those who reject relatively high offers are more likely to remain with their current employer. Further details are provided in Appendix C.1.

Fourth, because the SCE does not elicit future income with the same employer conditional on the receipt of a potential outside offer, in equation (25) we replace $y^{0*}_{i,t+1}$ with $\widehat{y}^{0*}_{i,t+1} = E\left(y^{0*}_{i,t+1} \mid \Omega_{i,t}, e_{i,t+1} = e_{i,t} = 1, s_{i,t+1} = 0\right)$, 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 $\varepsilon^{v}_{i,j(t+1)}$ but not $\varepsilon^{\omega}_{i,t+1}$. In other words, the reported offer reflects future employer-specific shocks in the match, but not shocks to human capital.[13] As a result, the model conveniently simplifies to a system of linear equations.

We use two survey questions to construct $\widehat{y}^{0*}_{i,t+1}$. The first asks respondents to report a point estimate of earnings, conditional on remaining with the same employer and working the same number of hours, 12 months ahead. The second asks for unconditional subjective expected earnings four months ahead. The first question poses two difficulties: it conditions on the same number of hours and involves a slight mismatch in the time

[13] Appendix B.2 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.

horizon. These issues lead us to rely on the second question, together with additional survey information, to construct the conditional subjective expected earnings required for our analysis. We describe the procedure in Appendix B.3.

6 Two-step model estimation

While Section 5 describes how subjective expectations provide the key inputs for identifying earnings and employment transitions, in this section, we outline a two-step estimation procedure for the model's parameters. In the first step, we estimate a subset of the model's parameters from the equations governing latent earnings and job transitions using fixed-effects regressions. In the second step, we exploit the covariance structure of the resulting composite disturbances, that is, the residuals of the first-step in levels, to identify the remaining structural parameters. For ease of reference, we preserve the numbering of the corresponding model equations throughout the estimation section, adding a single (double) apostrophe to denote equations used in the first (second) step.

Step 1. Step 1 estimates the parameters that enter the conditional means of potential earnings and the reduced-form job-transition equations using fixed-effects regressions. Each equation is estimated separately. Throughout Step 1, equations that involve latent offer bins are estimated by pooling observations across bins and exploiting variation in the bin index, allowing us to make full use of the subjective offer distribution.

In Step 1, we use two sets of equations: those governing potential earnings with a new employer and those governing job transitions probabilities. The first set is given by equation (18) and its distributional counterpart in equation (20), the second set by equations (23), (24), and (25), which correspond respectively to transitions from unemployment to employment, and to transitions for employed workers who switch or remain with their employer.

To summarize, the system of equations estimated in Step 1 can be written as:[14]

$$y_{i,t}^{c_{t+1}} = \rho\, y_{i,t}^{*} + (x_{i,t+1} - \rho x_{i,t})'\, \gamma + \alpha_{i,j(t)}^{yc} + \xi_{i,t}^{yc}, \tag{18'}$$

$$\ell_{ki,t}^{c_{t+1}} = \frac{1}{\sigma_e} r_{ki,t} - \frac{\rho}{\sigma_e} y_{i,t}^{*} - (x_{i,t+1} - \rho x_{i,t})'\, \frac{\gamma}{\sigma_e} + \alpha_{i,j(t)}^{\ell c} + \xi_{ki,t}^{\ell c}, \qquad k = 1, ..., 6, \tag{20'}$$

$$\ell_{ki,t}^{ue} = x_{i,t+1}^{u\prime} \gamma^{u} + \delta_y^{u} \tilde{r}_{ki,t} + \alpha_i^{ue} + \xi_{ki,t}^{ue}, \qquad k = 1, ..., 6, \tag{23'}$$

$$\ell_{ki,t}^{e1} = x_{i,t+1}^{1\prime} \gamma^{1} + \delta_y^{1} \tilde{r}_{ki,t} + \alpha_i^{e1} + \xi_{ki,t}^{e1}, \qquad k = 1, ..., 6, \tag{24'}$$

$$\ell_{ki,t}^{e0} = x_{i,t+1}^{0\prime} \gamma^{0} + \delta_y^{0} \widehat{y}_{i,t+1}^{0*} + \alpha_i^{e0} + \xi_{ki,t}^{e0}, \qquad k = 1, ..., 6. \tag{25'}$$

Here, $\alpha_{i,j(t)}^{yc}$ and $\alpha_{i,j(t)}^{\ell c}$ are reduced-form individual-job fixed effects; α_i^{ue}, α_i^{e1} and α_i^{e0} are reduced-form individual fixed effects; and $\xi_{i,t}^{yc}$, $\xi_{ki,t}^{\ell c}$, $\xi_{ki,t}^{ue}$, $\xi_{ki,t}^{e0}$, and $\xi_{ki,t}^{e1}$ denote elicitation errors, which are assumed to be mean independent of the information set available to the respondent at the time of the survey.

Each equation in this system is estimated by ordinary least squares with fixed effects, exploiting the panel structure of the SCE data. The persistence parameter ρ is estimated from both the point-expectation equation and its distributional counterpart, and the resulting estimates are combined by minimum distance, with bootstrap standard errors. Time-invariant heterogeneity is netted out throughout, and time-invariant job-match heterogeneity is additionally netted out in the first two equations. After Step 1 estimation, the following composite disturbances can be consistently estimated:

$$\lambda_{i,t}^{yc} \equiv y_{i,t}^{c_{t+1}} - \rho\, y_{i,t}^{*} - (x_{i,t+1} - \rho x_{i,t})'\, \gamma,$$

$$\lambda_{ki,t}^{\ell c} \equiv \ell_{ki,t}^{c_{t+1}} - \frac{1}{\sigma_e} r_{ki,t} + \frac{\rho}{\sigma_e} y_{i,t}^{*} + (x_{i,t+1} - \rho x_{i,t})'\, \frac{\gamma}{\sigma_e}, \qquad k = 1, ..., 6,$$

$$\lambda_{ki,t}^{ue} \equiv \ell_{ki,t}^{ue} - x_{i,t+1}^{u\prime} \gamma^{u} - \delta_y^{u} \tilde{r}_{ki,t}, \qquad k = 1, ..., 6,$$

$$\lambda_{ki,t}^{e1} \equiv \ell_{ki,t}^{e1} - x_{i,t+1}^{1\prime} \gamma^{1} - \delta_y^{1} \tilde{r}_{ki,t}, \qquad k = 1, ..., 6,$$

$$\lambda_{ki,t}^{e0} \equiv \ell_{ki,t}^{e0} - x_{i,t+1}^{0\prime} \gamma^{0} - \delta_y^{0} \widehat{y}_{i,t+1}^{0*}, \qquad k = 1, ..., 6.$$

The composite disturbances $\lambda(\cdot)$ therefore collect multiple unobserved components by construction. In the first two equations, they include elicitation error, individual heterogeneity, job-match heterogeneity, and the earnings innovation governed by the logit

[14] For notational convenience, we index all threshold- and bin-specific equations by $k = 1, \ldots, 6$. In equation (20'), the case $k = 6$ corresponds to the upper endpoint of the grid, for which the cumulative probability is mechanically equal to one and therefore does not generate an estimable logit observation.

scale parameter σ_e. In the remaining equations, they include elicitation error and two distinct forms of individual heterogeneity. The first form, which we refer to as ability, enters all equations, while the second captures mobility-related heterogeneity and enters only the equations governing job transition probabilities.

The first-step estimates of the composite disturbances $\hat{\lambda}(\cdot)$ obtained from the fixed-effects regressions are used as inputs to Step 2, where their variance and covariance structure is exploited to separately identify individual heterogeneity, job-match heterogeneity, and the scale of earnings risk.

Step 2. We use GMM to estimate the parameters governing unobserved heterogeneity and earnings risk, taking the persistence parameter ρ as given at its first-step estimate and using the first-step estimates $\hat{\lambda}(\cdot)$ of the model-implied composite disturbances.[15] In Step 2 we further assume that elicitation errors are uncorrelated over time and bins.

The model implies that the composite disturbances satisfy the following system of equations:

$$\lambda^{yc}_{i,t} = (1-\rho)\,\mu_i + (\phi-\rho)\,\upsilon_{i,j(t)} + \xi^{yc}_{i,t}, \tag{18''}$$

$$\lambda^{\ell c}_{ki,t} = -\frac{1-\rho}{\sigma_e}\,\mu_i - \frac{\phi-\rho}{\sigma_e}\,\upsilon_{i,j(t)} + \xi^{\ell c}_{ki,t}, \qquad k=1,...,6, \tag{20''}$$

$$\lambda^{ue}_{ki,t} = b^{u}_{\mu}\mu_i + b^{u}_{\eta}\eta_i + \xi^{ue}_{ki,t}, \qquad k=1,...,6, \tag{23''}$$

$$\lambda^{e1}_{ki,t} = b^{1}_{\mu}\mu_i + b^{1}_{\eta}\eta_i + \xi^{e1}_{ki,t}, \qquad k=1,...,6, \tag{24''}$$

$$\lambda^{e0}_{ki,t} = b^{0}_{\mu}\mu_i + b^{0}_{\eta}\eta_i + \xi^{e0}_{ki,t}, \qquad k=1,...,6, \tag{25''}$$

which corresponds to the mapping between the reduced-form fixed effects in (18')–(25') and the latent variables in the structural model: $\alpha^{yc}_{i,j(t)} = (1-\rho)\,\mu_i + (\phi-\rho)\,\upsilon_{i,j(t)}$, etc.

The composite disturbances $\lambda(\cdot)$ are not observed; in Step 2 they are replaced by their first-step estimates $\hat{\lambda}(\cdot)$ when constructing the GMM objective function.

Estimation proceeds by GMM using second moments of the first-step estimated composite disturbances. Specifically, we exploit both the variances of the estimated disturbances $\hat{\lambda}$ and their cross-equation covariances, including contemporaneous compo-

[15] The scale parameter σ_e is also identified in Step 1 from the equations governing potential earnings. We re-estimate σ_e in Step 2 for consistency with the full error-components structure; the resulting estimate is quantitatively very similar to the first-step estimate.

nents and autocovariances computed over the three observation periods available for each worker. When analyzing autocovariances, we explicitly account for whether individuals remain with the same employer or transition across employers between periods.

Using variation in individuals' expectations across different jobs provides information that helps separately identify individual-specific and job-specific components. Because our moments are based on autocovariances of earnings expectations and self-reported job transition probabilities, rather than on realized earnings, they do not condition on realized labor market outcomes. The expectation questions at time t refer to shocks in $t+1$, which, under the assumption that shocks are independent over time conditional on fixed effects, are orthogonal to employment status at t and to job transitions between $t-1$ and t. As a result, job mobility does not induce selection through realized shocks in our estimation strategy. Appendix C.3 provides detailed information on the implementation of the minimization problem, the weighting scheme, and other technical considerations.

The system of equations (18'')–(25'') 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 (Arellano and Bonhomme, 2023).

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

7 Results

In the previous sections, we discussed the available data, the theoretical model, the mapping of some equations to the data, the identification, and the estimation of the structural parameters. We now present our empirical analysis, starting with the criteria that inform the sample selection and then moving on to our parameter estimates. We focus on individuals aged 25 to 60 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. Using an argument analogous to the one we used for the unemployed, this restriction does not introduce selection bias.

Estimated parameters. Table 1 reports the parameters governing the evolution of latent earnings (equations (1) to (4)), obtained from the equations (18) and (20) in the first and second steps. We report the full set of first-step regression coefficients in Appendix D.1

Table 1: Persistence and risk parameters

	Coefficient	Men	Women
Persistence in productivity	ρ	0.511***	0.448***
		(0.087)	(0.153)
Pers. job-specific component	ϕ	0.304	0.000
		(0.318)	(0.181)
SD individual FE	σ_μ	0.578***	0.572***
		(0.022)	(0.027)
Logit scale of $(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{v}_{ij,t+1})$	σ_e	0.111***	0.182***
		(0.003)	(0.003)
SD job-specific component	σ_v	0.953**	0.568**
		(0.380)	(0.256)

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

The estimated persistence parameter ρ equals 0.511 (se 0.087) for men and 0.448 (se 0.153) for women. These values are 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, with more flexible specifications generally yielding lower persistence estimates (e.g., Browning et al. (2010), Hospido (2015), and Alvarez and Arellano (2022)).

Altonji et al. (2013) consider a model similar to ours and estimate it using PSID data, obtaining $\rho \approx 0.91$. Our specification builds on Altonji et al. (2013) by incorporating fixed effects without imposing parametric assumptions on their distribution. The availability of subjective expectations data makes this approach feasible and leads to lower estimates of ρ. As we show in Section 8, a lower structural persistence of general productivity shocks remains consistent with 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 $\phi = 0.304$ (se 0.318) for men and $\phi = 0.00$ (se 0.181) for women. In both cases, the estimates are not statistically different from zero, indicating substantial depreciation of employer-specific human capital following job transitions. These estimates stand in sharp contrast to the value of 0.7 reported in Altonji et al. (2013).

Our estimates of the standard deviation of the "ability" fixed effects, σ_μ, equal 0.578 (se 0.022) for men and 0.572 (se 0.027) for women. These values are substantially larger than those reported by Altonji et al. (2013), who find a value of 0.081 for men.[16] Importantly, our estimates are obtained without imposing functional form assumptions on the distribution of fixed effects.

In our model, latent earnings vary due to innovations in ability, ω_t, and employer–match quality, $\upsilon_{i,j(t)}$. The distribution of the combined innovation $(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{\upsilon}_{i,j(t+1)})$ is governed by a logit model, and the parameter σ_e therefore denotes the scale parameter of the logistic distribution rather than its standard deviation. We estimate this scale parameter with high precision, obtaining $\sigma_e = 0.11$ (s.e. 0.003) for men and $\sigma_e = 0.18$ (s.e. 0.003) for women. Under the logistic assumption, the implied standard deviation equals $\sigma_e \times \pi/\sqrt{3}$, which yields an implied standard deviation of approximately 0.20 for men and 0.33 for women. For men, this implied standard deviation is below the standard deviation of 0.29 reported by Altonji et al. (2013).[17]

[16] We normalize the coefficient on the ability fixed effect μ_i in the latent earnings equation to one and report the standard deviation of the fixed effects. Altonji et al. (2013), instead, normalize the standard deviation of the same component to one. The two normalizations are equivalent; the value 0.081 reported in the text corresponds to the coefficient on μ_i in their latent earnings equation.

[17] Given the data used in our baseline estimation, it is difficult to separately identify $\sigma_{\varepsilon^\upsilon}$ and $\sigma_{\varepsilon^\omega}$. However, an additional survey question elicits the probability distribution of respondents' future salary one year ahead, conditional on remaining with the same employer and working the same number of hours. In this scenario, shocks to employer–match quality are ruled out, allowing us to identify $\sigma_{\varepsilon^\omega}$. We then combine this estimate with the baseline estimate of the composite logistic scale parameter σ_e to obtain an approximate decomposition of the dispersion captured by σ_e, under the assumption that the two components are independent. This yields $\sigma_{\varepsilon^\omega} = 0.044$ and an implied $\sigma_{\varepsilon^\upsilon} = 0.102$ for men, and

The last row of the table reports the cross-sectional standard deviation of the employer-match quality fixed effects. Its point estimate is similar in magnitude to that of σ_μ for women and considerably higher than that for men. It is also higher than the initial cross-sectional standard deviation of υ in Altonji et al. (2013), which is 0.165. As our individuals are observed only for three quarters, our measured heterogeneity is estimated across age groups rather than only at labor market entry. It is therefore not surprising that the variability of this component is larger.

Tables 2 and 3 report estimates of the parameter affecting transitions to new jobs, as specified in equations (23), (24) and (25). Table 2 reports the estimated transition parameters. The term δ_y^u measures the effect of potential offers on the probability of leaving unemployment, δ_y^1 the effect of potential offers on the probability of quitting, and δ_y^0 the effect of current earnings on the probability of staying with the same employer. 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, reporting the effect on each probability of a 1% increase in the expected offer or in earnings.

Table 2: 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***	0.79	1.549***	0.34
on Pr(working)		(0.177)		(0.118)	
Effect of exp. offer	δ_y^1	3.353***	0.54	2.401***	0.37
on Pr(quitting)		(0.032)		(0.031)	
Effect of earnings	δ_y^0	0.377**	0.04	0.538***	0.06
on Pr(staying)		(0.174)		(0.194)	

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

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 effect

$\sigma_{\varepsilon^\omega} = 0.051$ and an implied $\sigma_{\varepsilon^\upsilon} = 0.175$ for women. These results suggest that shocks to employer–match quality are the dominant contributors to the dispersion captured by σ_e. This decomposition, however, is conditional on being employed in both t and $t+1$ and on not changing employers in $t+1$, and is therefore subject to selection, unlike our main results.

of current earnings on the probability of staying is smaller, especially for men. This could reflect the absence of data on how current employers adjust pay in response to outside offers and our need to approximate y_{it}^0, which 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 3: Individual fixed effects and transitions

	Coefficient	Men	Women
Standard deviation of η	$SD(\eta)$	0.519***	0.360***
		(0.074)	(0.053)
Effect of μ	b_μ^u	-4.894***	-2.274***
on Pr(working)		(1.483)	(0.599)
Effect of η	b_η^u	4.577**	-6.111
on Pr(working)		(2.012)	(4.953)
Effect of μ	b_μ^1	-4.038***	-3.027***
on Pr(quitting)		(0.137)	(0.219)
Effect of μ	b_μ^0	-0.365***	-0.321***
on Pr(staying)		(0.045)	(0.058)
Effect of η	b_η^0	-1.380***	-2.010***
on Pr(staying)		(0.525)	(0.339)

Table 3 reports the impact 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 job change for the employed to 1.

For men, the coefficients on the ability fixed effects are all negative and significant across the three probabilities. The coefficient on mobility fixed effects for staying in the same job is negative and significant, as expected, while it is positive for leaving unemployment.

For women, the ability fixed effects follow a similar pattern and are also statistically significant but smaller in magnitude. The impact that the mobility fixed effects have on the probability of staying in the same job is, similarly to men, negative and significant. However, the same fixed effect has, unlike for men, a negative, but insignificant impact on the probability of leaving unemployment. This surprising fact 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 D.2.

Table 4: Variances of the residuals ξ, compared with the variances of the composite disturbances λ

		Men		Women	
Equation	X	$\widehat{\mathrm{Var}(\lambda_{it}^X)}$	$\widehat{\mathrm{Var}(\xi_{it}^X)}$	$\widehat{\mathrm{Var}(\lambda_{it}^X)}$	$\widehat{\mathrm{Var}(\xi_{it}^X)}$
Expected offer	yc	0.138	0.019	0.215	0.050
Offer distribution	lc	11.481	1.813	6.989	2.038
Pr(working)	ue	17.162	3.513	8.921	2.391
Pr(staying)	e0	1.552	0.994	1.582	1.024
Pr(quitting)	e1	6.778	1.058	4.216	1.089

Table 4 shows the size of the variance of the residuals ξ in Equations (18'')–(25''), which remain after removing the contribution of the ability and mobility fixed effects, and the job match components. These residuals may reflect elicitation errors, but also other sources of additional variation such as model misspecification. For most equations, they are relatively small compared with the variance of the composite disturbance λ. The only exception is the equation for job staying $e0$, about which, as discussed earlier, the survey provides less information because there is no experimentation about potential wage offers by the existing employer.[18]

8 The role of risk and unobserved heterogeneity

The estimated model is rich and allows for several channels shaping the dynamics of realized earnings implied by the process *as perceived by the respondents,* [19] including the evolution of ability, the persistence of employer-specific individual-level productivity, and selection into and out of employment and job changes.

[18]In Appendix D.3 we show that our estimates are very similar if we choose to remove this equation from the estimation.

[19]Under the hypothesis of rational expectations of the entire stochastic process, not only the mean, the subjective and objective earning processes coincide. We do not need such an assumption. The short nature of the panel we use does not allow us to test it.

To assess the implications of the estimated parameters for the dynamics of realized earnings, we simulate the model of subjective earnings and employment dynamics for a large number of individuals over many periods, starting from an initial distribution that matches the data. Using the simulated panel, we compute moments commonly used in the literature, without the distortions induced by a short longitudinal dimension. For simplicity, we abstract from unemployment risk and restrict attention to individuals employed in both periods t and $t+1$.

8.1 Decomposing the variance and persistence of earnings

We represent generic heterogeneity using an individual fixed effect, denoted by α_i. These fixed effects differ from the μ_i in the structural model and should be interpreted as *non-structural* heterogeneity in a statistical decomposition of components of variation.

Conditional on $y_{i,t}$, the cross sectional variability of earnings can be decomposed into a risk component and an unobserved heterogeneity component:

$$\begin{aligned}\text{Var}(y_{i,t+1} \mid y_{i,t}) = \; & \mathbb{E}_{\alpha_i|y_{i,t}}\left[\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right] + \\ & \text{Var}_{\alpha_i|y_{i,t}}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right], \end{aligned} \tag{26}$$

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

One can further decompose the first term on the right hand side of Equation 26 by taking into account an additional component of unobserved heterogeneity, $\zeta_{i,j(t+1)}$, which is individual- and employer-specific and thus can vary over time. Therefore, Equation (26) can be re-written as:

$$\begin{aligned}\text{Var}(y_{i,t+1} \mid y_{i,t}) = \; & \mathbb{E}_{\zeta_{i,j(t+1)},\alpha_i|y_{i,t}}\left[\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{i,j(t+1)}) \mid y_{i,t}\right] + \\ & \mathbb{E}_{\alpha_i|y_{i,t}}\left(\text{Var}_{\zeta_{i,j(t+1)}|y_{i,t},\alpha_i}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{i,j(t+1)}) \mid y_{i,t}\right]\right) + \\ & \text{Var}_{\alpha_i|y_{i,t}}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right]. \end{aligned} \tag{27}$$

The $\zeta_{i,j(t+1)}$'s are different from the $\upsilon_{ij(t)}$'s in the model, in the same way as the α_i's are different from the μ_i's. In Equation (27), 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 $\zeta_{i,j(t+1)}$; 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 as in Equation (26).

With a similar approach, we can decompose the total persistence of earnings $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t})}{\partial y_{i,t}}$ into contributions that come from unobserved heterogeneity and/or job specific components α_i and $\zeta_{i,j(t+1)}$:

$$\frac{\partial \mathbb{E}(y_{i,t+1} \mid y_{i,t})}{\partial y_{i,t}} = \mathbb{E}\left[\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\alpha_i)}{\partial y_{i,t}} \mid y_{i,t}\right] + \mathbb{E}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i)\frac{\partial \ln f(\alpha_i|y_{i,t})}{\partial y_{i,t}} \mid y_{i,t}\right]. \tag{28}$$

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 an individual fixed effect and $f(\alpha_i|y_{i,t})$ is the conditional density function of α_i. The left-hand-side of Equation (28) can be estimated by a regression of $y_{i,t+1}$ on $y_{i,t}$. The first term on the right-hand side of the same equation can be approximated by estimating a regression of $y_{i,t+1}$ on $y_{i,t}$ and individual specific fixed effects.

One can further expand the first term in this equation to separate employer-specific and individual-specific time invariant components to obtain:

$$\frac{\partial \mathbb{E}(y_{i,t+1} \mid y_{i,t})}{\partial y_{i,t}} = \mathbb{E}\left[\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\zeta_{i,j(t+1)})}{\partial y_{i,t}} \mid y_{i,t}, \alpha_i\right] + \mathbb{E}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \zeta_{i,j(t+1)})\frac{\partial \ln f(\zeta_{i,j(t+1)}|y_{i,t},\alpha_i)}{\partial y_{i,t}} \mid y_{i,t}, \alpha_i\right] + \mathbb{E}\left[\mathbb{E}(y_{i,t+1} \mid y_{i,t}, \alpha_i)\frac{\partial \ln f(\alpha_i|y_{i,t})}{\partial y_{i,t}} \mid y_{i,t}\right] \tag{29}$$

where now $f(\zeta_{i,j(t+1)}|y_{i,t}, \alpha_i)$ is the density function of $\zeta_{i,j(t+1)}$ conditional on α_i and $y_{i,t}$. In this expression, the first component represents the contribution of persistence in earnings controlling for person- and job-specific fixed effects to overall persistence, the second component represents the contribution of the individual-employer specific component $\zeta_{i,j(t+1)}$ to overall persistence, and the third component represents the contribution of the individual-specific component α_i to overall persistence. The first term of the right-hand-side of Equation (29) can be approximated by estimating a fixed effect regression.

8.2 Implementation and results

We evaluate the persistence and variance decompositions in Equations (26)–(29) using data simulated from our estimated model. The simulation includes a large number of individuals observed over many periods, which avoids biases associated with short panels in the presence of fixed effects, such as Nickell bias. Because the second decomposition relies on regressions with employer-specific fixed effects, we restrict attention to employment spells lasting at least 25 periods to ensure reliable estimation.[20] We begin with persistence and then turn to the variance decomposition.

Persistence. We study the determinants of earnings persistence by estimating the following three linear regressions on the simulated data:

$$y_{i,t+1} = \beta_0^A + \beta_1^A y_{i,t} + \varepsilon_{i,t+1}^A \tag{30}$$

$$y_{i,t+1} = \beta_0^B + \beta_1^B y_{i,t} + \alpha_i + \varepsilon_{i,t+1}^B \tag{31}$$

$$y_{i,t+1} = \beta_0^C + \beta_1^C y_{i,t} + \alpha_i + \zeta_{ij(t+1)} + \varepsilon_{i,t+1}^C. \tag{32}$$

The coefficient β_1^A estimates $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t})}{\partial y_{i,t}}$ in equation (28). The coefficient β_1^B analogously estimates $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\alpha_i)}{\partial y_{i,t}}$ in the same equation. Finally, β_1^C estimates the first term on the right-hand-side of equation (29), $\frac{\partial \mathbb{E}(y_{i,t+1}|y_{i,t},\alpha_i,\zeta_{ij(t+1)})}{\partial y_{i,t}}$. Equation (30) is estimated by OLS, equation (31) by a within estimator that demeans at the individual level; and equation (32) by a within estimator that demeans at the individual and job-spell levels. Table 5 reports the results.

Table 5: Measured persistence in simulated data from different specifications.

Parameter	Value	Controlling for...
β_1^A	0.975	
β_1^B	0.885	+ individual heterogeneity α_i
β_1^C	0.461	+ individual-employer fixed effects $\zeta_{ij(t)}$

Interesting results emerge from estimating a simple autoregressive regression of earnings on data simulated from the model. Although the persistence of the individual-level

[20] Appendix E provides additional details on the simulation procedure.

productivity component is set to 0.511 in the simulation, the estimated persistence from a pooled AR(1) regression equals 0.975, close to unity. This does not imply that the productivity process itself is nearly a unit root. Rather, observed earnings combine several persistent components: permanent individual heterogeneity, the persistent general productivity process, and a match-specific component that is constant within job spells and only partially refreshed upon job changes. In addition, employment and mobility decisions depend on next-period latent earnings and mobility types, so observed earnings are endogenously selected on both permanent and time-varying components. The AR(1) regression therefore loads on the composite persistence of all these forces, attributing to a single coefficient what in the model reflects multiple sources of serial dependence.

Controlling for individual fixed effects lowers the estimated persistence to 0.885 when using a simulated panel of sufficient length to avoid Nickell bias. Worker fixed effects remove permanent individual heterogeneity, but they do not eliminate the persistence generated by match-specific components, which remain constant within job spells and therefore mechanically induce serial correlation across adjacent observations. Moreover, selection through the transition process remains operative: employment and mobility decisions continue to depend on latent earnings and on mobility types, so the sample of observed earnings remains endogenously selected. As a result, an AR regression with worker fixed effects still reflects the joint persistence of the general productivity and match components and therefore remains well above the persistence of the general productivity process alone.

Finally, controlling for individual–job fixed effects reduces the estimated persistence to 0.461, close to the structural persistence parameter of the general productivity process. Job fixed effects absorb the spell-level match component that, if not controlled for, generates strong mechanical serial correlation in earnings within job spells. Although selection through employment and mobility decisions remains present, the dominant non-productivity source of persistence is removed. The remaining serial correlation therefore primarily reflects the dynamics of the general productivity component, and the resulting estimate aligns closely with its structural persistence parameter. Hence, once we control for both individual and job-specific fixed effects, the estimates from simple reduced-form regressions are close to those implied by the semi-structural model with job changes.[21]

In addition, we note that a substantial share of observed earnings persistence and variability is driven by individual and job-specific fixed effects. Considering the individual–job pair (i, j) and further decomposing the job-specific component in equation (32) as $\zeta_{i,j(t+1)} = \kappa_i + \xi_{i,j(t+1)}$, we can distinguish between individual means (including individual average job quality) and match quality innovations. Using the reduced-form persistence estimates (equations 30-32), an approximate calculation suggests that about 80% of the variance of the composite heterogeneous individual-job effects is driven by individual means (including individual ability and mean match quality) and 20% by match quality innovation components.[22]

Variance. We compute the variance measures described above by estimating the following three regressions on simulated data:[23]

$$y_{i,t} = \beta_0^D + \beta_1^D y_{i,t-1} + \beta_2^D y_{i,t-1}^2 + \varepsilon_{i,t}^D \quad (33)$$

$$y_{i,t} = \beta_0^E + \beta_1^E y_{i,t-1} + \beta_2^E y_{i,t-1}^2 + \alpha_i + \varepsilon_{i,t}^E \quad (34)$$

$$y_{i,t} = \beta_0^F + \beta_1^F y_{i,t-1} + \beta_2^F y_{i,t-1}^2 + \alpha_i + \zeta_{ij(t+1)} + \varepsilon_{i,t}^F \quad (35)$$

As before, equation (33) is estimated by OLS, equation (34) by a within estimator that demeans at the individual level; and equation (35) by a within estimator that demeans at the individual and job-spell levels. The corresponding residuals are $\hat{\varepsilon}_{it}^D$, $\hat{\varepsilon}_{it}^E$, and $\hat{\varepsilon}_{it}^F$.

Next, we approximate $\text{Var}(y_{i,t+1} \mid y_{i,t})$ using fitted values from a linear regression of $(\hat{\varepsilon}_{it}^D)^2$ on $(y_{i,t}, y_{i,t}^2)$. Similarly, we approximate $\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i)$ from the fitted values of a linear regression of $(\hat{\varepsilon}_{it}^E)^2$ on $(y_{i,t}, y_{i,t}^2)$ and $\hat{\alpha}_i$, and $\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{ij(t+1)})$ from the fitted values of a linear regression of $(\hat{\varepsilon}_{it}^F)^2$ on $(y_{i,t}, y_{i,t}^2)$ and $\hat{\zeta}_{ij(t+1)}$.

Denote the fitted values of these two latter regressions with $\hat{\kappa}_{it}^E$ and $\hat{\kappa}_{it}^F$ respectively. Regressing $\hat{\kappa}_{it}^E$ on $(y_{i,t}, y_{i,t}^2)$ yields an estimate for $\mathbb{E}_{\alpha_i|y_{i,t}}\left[\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i) \mid y_{i,t}\right]$. Analogously, we approximate $\mathbb{E}_{\alpha_i,\zeta_{ij(t+1)}|y_{i,t}}\left[\text{Var}(y_{i,t+1} \mid y_{i,t}, \alpha_i, \zeta_{ij(t+1)}) \mid y_{i,t}\right]$ by estimating the fitted values of a linear regression of $\hat{\kappa}_{it}^F$ on $(y_{i,t}, y_{i,t}^2)$.

[21] See Appendix E.2 for details.

[22] See Appendix E.3 for details.

[23] Since our model is nonlinear, one could in principle obtain richer decompositions by running more flexible nonlinear regressions on the simulated data. However, we choose to rely on regressions with additive fixed effects, as these map more transparently to popular estimators used in the literature.

Figure 4 plots these objects across the earnings distribution, and Table 6 reports their means. The comparison isolates the variance of next period's earnings conditional on (i) current earnings (red line and first row of the table), (ii) current earnings and individual fixed effects (blue line, second row of the table), and (iii) current earnings together with job- and individual-specific fixed effects (green line and third row of the table). The difference between cases (i) and (ii) corresponds to the unobserved individual heterogeneity component, i.e., the second term on the right-hand side of equation 26.

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

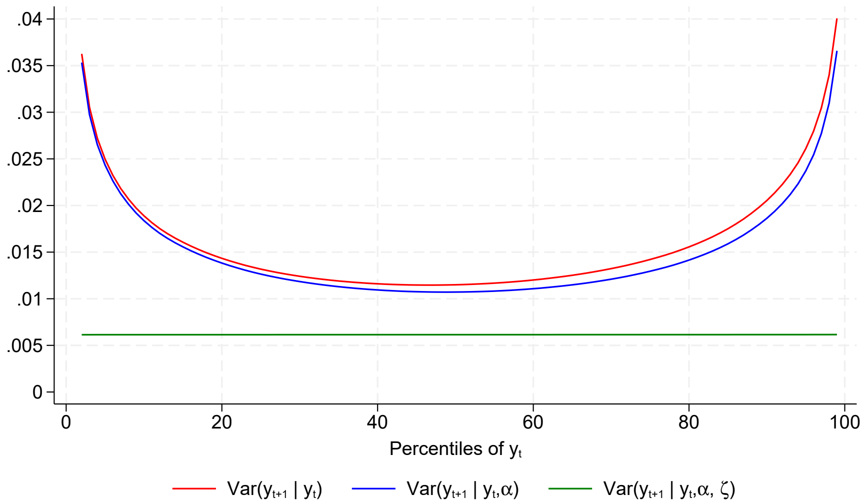


Table 6: Risk decomposition, averages of Figure 4 (top) and contributions of individual fixed effects, employer transitions and employer fixed effects, and remainder (bottom).

Risk Measure	Average	Contribution of...	
$Var(y_{t+1}\|y_t)$	0.017	Risk conditional on α and ζ	36.43%
$Var(y_{t+1}\|y_t,\alpha)$	0.016	Heterogeneity in ζ	57.20%
$Var(y_{t+1}\|y_t,\alpha,\zeta)$	0.006	Heterogeneity in α	6.36%

We first note that the variance of individual earnings in the SCE data displays pronounced non-linearities in lagged income, consistent with recent findings in the literature (see, for example, Guvenen et al. (2021)). In particular, the conditional variance of future earnings is higher at both low and high levels of current earnings.

Our decompositions clarify which components of the model account for the level and shape of conditional earnings variance. Conditioning on individual fixed effects reduces the overall magnitude of the variance only marginally and does not alter the heteroskedastic pattern observed when conditioning solely on current earnings. In contrast, conditioning additionally on job-specific fixed effects substantially reduces the variance and eliminates the heteroskedasticity. This result highlights an important mechanism. Conditional on remaining with the same employer, earnings risk is relatively low. Most earnings risk is instead associated with job changes and is concentrated at low and high levels of current earnings.

We report the results of the simulation for women in Appendix E.4. Although the overall level of earnings variability is higher than for men, the contribution of all three factors displays very similar patterns, with an even higher share of the persistence and variance accounted for by job-specific effects.

9 Conclusions

Expectations play a central role in economic decision-making. Beyond expected values, the perceived distribution of future outcomes, including uncertainty, shapes individual behavior. Although economists were initially reluctant to use subjective expectations, substantial progress over the past decades has improved their measurement and empirical credibility (e.g., Manski (2004)). These advances have been adopted by central banks and other institutions through the systematic collection of expectations surveys, enabling new analyses of labor market dynamics.

The New York Fed's Survey of Consumer Expectations exemplifies these advances. Its data elicit individuals' beliefs about future earnings distributions under counterfactual scenarios, as well as perceived probabilities of discrete events. This information allows researchers to discipline structural models by mapping parameters to subjective probabilities while accounting for fixed effects and measurement error. In this paper, we use these data to estimate a structural model of earnings dynamics with job mobility and unemployment. Combining subjective expectations with realized earnings yields a detailed characterization of labor market risk, decision-making, and unobserved heterogeneity.

We show that subjective beliefs, when appropriately measured, contain sufficient information and variability to estimate both the latent earnings process and the employment and employer transition dynamics. Our framework allows us to use these expectations data to estimate a rich structural model, including ability and mobility fixed effects, using tractable fixed-effects regression and GMM estimation methods.

Using subjective expectations yields a simpler and more transparent identification of the structural parameters of interest without being forced to the strong distributional assumptions often required when using realized outcomes alone. It also leads to empirical results that differ from those obtained using realized earnings alone. Much of the existing literature on earnings dynamics finds that innovations to latent earnings, interpreted as productivity shocks, are highly volatile and persistent. For example, Altonji et al. (2013) estimate an autocorrelation parameter of 0.96, even after controlling for individual unobserved heterogeneity. In contrast, our structural estimate of the same parameter is substantially lower, at 0.51.

Another key parameter is ϕ, which governs the extent to which job-specific human capital carries over following a job change. While Altonji et al. (2013) report a value of about 0.7, our estimate of 0.3 is substantially lower.

Another distinctive difference concerns the variability of unobserved heterogeneity. Our estimates, obtained without imposing functional form assumptions, are substantially larger than those reported by Altonji et al. (2013). For men, we estimate $\sigma_\mu = 0.58$, compared with 0.081 in Altonji et al. (2013). Given the short longitudinal dimension of our data, the estimated fixed effects absorb components such as age and race (while controlling for tenure). It is therefore unlikely that omitted observables alone account for the magnitude of this difference.

Simulations of the model reveal that, despite low structural persistence and risk, realized earnings generated by the model exhibit reduced-form properties similar to those reported in the literature. When conditioning only on current earnings, or on earnings and individual fixed effects, both earnings risk and persistence are high. Once job-specific fixed effects are included, however, both risk and persistence decline substantially.

The key implication of these results is that, abstracting from unemployment, individual earnings risk is primarily associated with job changes. Conditional on remaining with the same employer, earnings risk is relatively low.

We show that the high persistence and volatility commonly reported in the literature largely reflect unaccounted employer transitions and heterogeneity, rather than highly persistent productivity shocks. Likewise, much of the nonlinear relationship between earnings volatility and income disappears once individual and employer-match effects are taken into account. These findings highlight the value of combining subjective expectations data with a rich model of employment dynamics to analyze labor market outcomes.

These results depend on the quality and richness of the subjective expectations data. Although substantial progress has been made in the collection and validation of such data, important challenges remain. Further research and experimentation are needed to establish robust standards for the elicitation and use of subjective expectations.

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. (2003). *Panel Data Econometrics*. Oxford: Oxford University Press.

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.

Arellano, M., R. Blundell, and S. Bonhomme (2017). Earnings and consumption dynamics: A nonlinear panel data framework. *Econometrica 85*(3), 693–734.

Arellano, M. and S. Bonhomme (2012). Identifying distributional characteristics in random coefficients panel data models. *Review of Economic Studies 79*(3), 987–1020.

Arellano, M. and S. Bonhomme (2023). Recovering latent variables by matching. *Journal of the American Statistical Association 118*(541), 693–706.

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.

Bachmann, R., G. Topa, and W. van der Klaauw (Eds.) (2023). *Handbook of Economic Expectations* (1 ed.). Elsevier.

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

Caplin, A., V. Gregory, E. Lee, S. Leth-Petersen, and J. Sæverud (2024). Subjective earnings risk. Working paper.

Cox, D. R. and E. J. Snell (1970). *Analysis of Binary Data*. London: Chapman and Hall.

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

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

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

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021). What do data on millions of 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 C. O'Dea (2022). Expectations data in structural micro models. In R. Bachmann, G. Topa, and W. van der Klaauw (Eds.), *Handbook of Economic Expectations*, Volume 21, pp. 647–675. Elsevier.

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.

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

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

Table 7: Summary of notation in the main text, 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 $t+1$
$y_{i,t}^{c_{t+1}}$	Log point expectation of annual salary of best offer
τ_k	Multiplicative factor to form bins for probability distributions
$r_{ki,t}$	Upper log boundary of offer bin k
$\tilde{r}_{ki,t}$	Log midpoint of offer bin k, used in estimation
$p_{ki,t}^{c_{t+1}}$	Reported CDF at threshold r_k
$\ell_{ki,t}^{c_{t+1}}$	Log odds ratio of CDF at threshold r_k
$p_{ki,t}^{e0}$	Probability of staying in same job, based on offer k
$p_{ki,t}^{e1}$	Reported probability of changing job, based on offer k
$p_{ki,t}^{ue}$	Reported probability of starting work, based on offer k
$\ell_{ki,t}^{e1}$	Log odds ratio of changing job, based on offer k
$\ell_{ki,t}^{e0}$	Log odds ratio of staying in same job, based on offer k
$\ell_{ki,t}^{ue}$	Log odds ratio of starting work, based on offer k
$x_{i,t+1}$	Time-varying observables (e.g. employment tenure)
	Latent model 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
$\upsilon_{i,j(t+1)}^{0*}$	Match-specific effect if keeping job between t and $t+1$
$\upsilon_{i,j(t+1)}^{1*}$	Match-specific effect if changing job between t and $t+1$
$\upsilon_{i,j(t)}$	Match-specific effect associated with job $j(t)$
$\xi_{ki,t}^{\cdot}$	Elicitation (measurement) error in reported probabilities $(yc, lc, ue, e1, e0)$
$\alpha_i^{\cdot}$	Reduced form (and job) fixed effects $(yc, lc, ue, e1, e0)$
$\lambda_{ki,t}$	Composite disturbances $(yc, lc, ue, e1, e0)$
μ_i	Individual ability fixed effect
η_i	Individual mobility fixed effect
$\omega_{i,t}$	Stochastic individual productivity

Table 8: Summary of notation in the main text, part 2

Model parameters	
σ_μ	Standard deviation of μ
σ_η	Standard deviation of η
ρ	Persistence of ω
ϵ^ω	Shock to ω
σ_υ	Standard deviation of υ
ϕ	Persistence of υ across job changes
ϵ^υ	Shock to υ
σ_e	Logit scale in the offer distribution (employed)
σ_u	Logit scale parameter in the offer distribution (unemployed)
δ_y^0	Effect of current earnings on (log) odds of staying
δ_y^1	Effect of potential earnings on (log) odds of changing job
δ_y^u	Effect of potential earnings on (log) odds of working (for unemployed)
b_μ^0	Effect of μ on log odds of staying
b_μ^1	Effect of μ on log odds of changing job
b_μ^u	Effect of μ on log odds of working (for unemployed)
b_η^0	Effect of η on log odds of staying
b_η^1	Effect of η on log odds of changing job
b_η^u	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
Objects from simulated data	
α_i	Individual fixed effect in statistical variance decompositions
$\zeta_{ij(t)}$	Individual-job fixed effect in reduced-form regressions
$\beta_{0,1,2}^{A,\dots,E}$	Coefficients in reduced-form regressions
$\hat{\epsilon}^{D,E,F}$	Residuals in reduced-form regressions
$\hat{\kappa}^{E,F}$	Fitted values in reduced-form regressions of $\hat{\epsilon}^{E,F}$ on y

Supplemental Appendices

A The SCE data

A.1 Sample selection

The data come from the New York Fed's Survey of Consumer Expectations (SCE), a nationally representative, internet-based monthly survey of a rotating panel of approximately 1,300 U.S. household heads, available since 2013. Respondents remain in the panel for up to twelve months and answer a monthly core module as well as periodic supplementary modules. We use the labor market module, fielded every March, July, and November, which collects information on labor market status and expectations about job offers and earnings over the subsequent four months.

Our sample covers the period March 2014 through November 2019. After merging the labor market module and retaining only the March, July, and November waves, the initial sample consists of 20,708 observations (9,842 women and 10,866 men). We then restrict the sample to individuals aged 25 to 60 who are either employed or unemployed, excluding self-employed workers.

To ensure well-defined earnings measures, we keep employed individuals with annual earnings above \$2,500, drop observations with missing or implausible earnings reports, and trim the top and bottom 1% of the distribution of percentage differences between current earnings and (i) the expected job offer and (ii) expected earnings conditional on employment over the next four months. Finally, we exclude individuals who report a positive probability of being self-employed in four months. The final estimation sample consists of 3,121 observations for men and 3,110 observations for women.

Table 9 summarizes the sample selection procedure step by step, separately for men and women.

Table 9: Sample Selection

	Men	Women
Merged labor market module,		
March/July/November waves	10,866	9,842
Keeping age 25–60	6,777	7,085
Keeping employed or unemployed	6,166	5,983
Excluding self-employed	5,464	5,518
Excluding employed with annual earnings below $2,500	5,446	5,499
Excluding employed with missing earnings	5,230	5,196
Excluding observations with standard deviation of log earnings > 1.61	5,215	5,177
Excluding missing values in expected job offers	3,876	3,650
Excluding outliers in expected earnings changes	3,754	3,547
Excluding observations with positive self-employment prob. next period	3,121	3,110

Table 10 provides some demographics and summary statistics from our sample.

A.2 Additional earnings expectation questions

The main body of the paper describes the formulation of the expectation questions that drive identification. This subsection documents additional SCE questions on earnings expectations that we use to inform model objects that are not directly observed.

Unlike the counterfactual job-offer expectations analyzed in the main text, these questions do not admit the simple linear representation exploited in our two-step estimation procedure. In one case, expectations are elicited conditional on a future employment outcome and therefore reflect selection. In the other, expectations are unconditional and combine beliefs about earnings and employment. As a result, we do not use these questions for identification in the main analysis, but exploit them to obtain information on expected earnings in the event of remaining with the same employer, $y^0_{i,t+1}$.

Earnings in the same job. Respondents are asked to consider their earnings twelve months ahead under the hypothetical scenario in which they remain in the same job, at

Table 10: SCE data, sample summary statistics.

	Men N	Mean	Std.Dev.	Women N	Mean	Std.Dev.
Full sample						
Age	3121	42.69	10.06	3110	41.98	10.42
White	3121	0.86	0.35	3110	0.79	0.41
Couple	3121	0.73	0.44	3109	0.59	0.49
HS	3121	0.08	0.27	3110	0.08	0.26
Some college	3121	0.25	0.43	3110	0.32	0.47
College degree	3121	0.39	0.49	3110	0.34	0.47
Post-graduate education	3121	0.28	0.45	3110	0.27	0.44
Unemployed	3121	0.05	0.21	3110	0.08	0.27
Workers						
Current earnings ($1,000s)	2952	88.56	83.84	2857	57.19	84.63
Expected offer in 4 months ($1,000s)	2952	92.75	79.36	2857	61.01	45.59
Prob. receiving offer next 4 months	2952	0.29	0.28	2857	0.31	0.29
Prob. employed in 4 months	2952	0.97	0.09	2851	0.98	0.08
Prob. change job in 4 months	2952	0.11	0.18	2851	0.13	0.20
Prob. unemployed in 4 months	2952	0.03	0.09	2851	0.02	0.08
Unemployed/laid off						
Previous earnings ($1,000s)	164	51.04	49.86	246	31.53	51.87
Expected offer in 4 months ($1,000s)	169	50.39	45.05	253	53.27	220.96
Prob. receiving offer next 4 months	169	0.60	0.31	253	0.58	0.31
Prob. employed in 4 months	169	0.55	0.33	253	0.56	0.32

the same employer, and work the same number of hours. They report both a point expectation for earnings and a subjective distribution over discrete bins describing the probability of earnings increases or decreases of different magnitudes. Because the question conditions on future employment and hours, responses are subject to selection. Additionally, the twelve-month horizon has a timing difference relative to the main questions, which are all about expected offers and employment status in four months' time.

Expected earnings in four months. Respondents are also asked to report their expected annual earnings four months ahead. This question elicits unconditional expectations and therefore reflects both beliefs about future earnings and beliefs about future employment status. We use this information, together with additional survey responses, to construct an empirical counterpart to expected earnings conditional on remaining with the same employer.

B Model extensions and discussion

B.1 Earnings equations for the unemployed

As we discussed in the main text, we do not need to use the data for the unemployed. However, we could 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 $\upsilon_{i,j(t+1)} = \phi\upsilon_{i,j(t)} + \varepsilon^{\upsilon}_{i,j(t+1)}$ and $\upsilon_{i,j(t-1)} = (1/\phi)\upsilon_{i,j(t)} - (1/\phi)\varepsilon^{\upsilon}_{i,j(t)}$ can then re-write equation 18 as follows:

$$\begin{aligned} y^{c_{t+1}}_{i,t} &= \rho^2 y^*_{i,t-1} + \left(x_{i,t+1} - \rho^2 x_{i,t-1}\right)'\gamma + \left(1-\rho^2\right)\mu_i + \left(\phi - \rho^2/\phi\right)\upsilon_{i,j(t)} + \\ &\quad +(\rho^2/\phi)\varepsilon^{\upsilon}_{i,j(t)} + \rho\varepsilon^{\omega}_{i,t}. \end{aligned} \tag{36}$$

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 spell is one period, and denoting with σ_u the scale parameter of the composite innovation associated with the unemployed, similarly to σ_e for the employed, we can then write:

$$\begin{aligned} \ell^{c_{t+1}}_{ki,t} &= (1/\sigma_u)\, r_{ki,t} + \left(-\rho^2/\sigma_u\right) y^*_{i,t-1} + \left(x_{i,t+1} - \rho^2 x_{i,t-1}\right)'(-\gamma/\sigma_u) \\ &\quad -\mu_i\left(1-\rho^2\right)/\sigma_u - \left(\phi^2 - \rho^2\right)/\sigma_u \upsilon_{i,j(t-1)} - \left(\phi\varepsilon^{\upsilon}_{i,j(t)} + \rho\varepsilon^{\omega}_{i,t}\right)/\sigma_u. \end{aligned} \tag{37}$$

Following the same substitutions used to obtain equation 36, and more conveniently for the second step estimation we get:

$$\begin{aligned} \ell_{ki,t}^{c_{t+1}} &= (1/\sigma_u)\, r_{ki,t} + \left(-\rho^2/\sigma_u\right) y_{i,t-1}^{*} + \left(x_{i,t+1} - \rho^2 x_{i,t-1}\right)' (-\gamma/\sigma_u) \\ &\quad -\mu_i \left(1-\rho^2\right)/\sigma_u - \left(\left(\phi - \rho^2/\phi\right)/\sigma_u\right) \upsilon_{i,j(t)} - \left(\rho^2/\phi \varepsilon_{i,j(t)}^{\upsilon} + \rho \varepsilon_{i,t}^{\omega}\right)/\sigma_u. \end{aligned} \tag{38}$$

B.2 Endogenous self-selection in both υ and ω

In our main results, we assume that endogenous self-selection is only driven by $\varepsilon_{i,j(t+1)}^{\upsilon}$ but not $\varepsilon_{i,t+1}^{\omega}$. We now discuss the case in which we relax this assumption and the associated estimation challenges for the employment to employment transitions.

For the employed workers, the mapping with the multinomial model probabilities (8)-(9) is not straightforward. The reason is that, although the log odd ratios (10)-(11) only depend on one of $y_{i,t+1}^{1*}$ or $y_{i,t+1}^{0*}$, the probabilities will depend on both $y_{i,t+1}^{1*}$ and $y_{i,t+1}^{0*}$. Let us elaborate on this point. For the currently employed we observe (Equation 21 in the main text):

$$\begin{aligned} p_{ki,t}^{e1} &= \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid r_{k-1i,t} < y_{i,t+1}^{1*} \leq r_{ki,t},\ y_{i,t+1}^{0*},\ \Omega_{i,t}, e_{i,t}=1\right), \\ &\approx \Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y_{i,t+1}^{1*} = \tilde{r}_{ik,t}, \Omega_{i,t}, e_{i,t}=1\right), \end{aligned} \tag{39}$$

where $\tilde{r}_{ik,t}$ denotes the point assigned to each interval, defined as the log of the mid-point (in levels) of the interval associated with bin k given the reference counterfactual income level $y_{i,t}^{c_{t+1}}$, as described in the main text.

One interpretation of (39) is as an average of the model probabilities taken with respect to the distribution of $y_{i,t+1}^{0*}$ conditioned on $y_{i,t+1}^{1*} = \tilde{r}_{ik,t}$. Essentially, given $\Omega_{i,t}$, fixing a value for $y_{i,t+1}^{1*}$ is equivalent to fixing $\varepsilon_{i,t+1}^{\omega} + \varepsilon_{i,j(t+1)}^{\upsilon}$. Thus, under this interpretation we observe averaged model probabilities with respect to the distribution of $\varepsilon_{i,t+1}^{\omega}$ conditioned on $\varepsilon_{i,t+1}^{\omega} + \varepsilon_{i,j(t+1)}^{\upsilon}$:

$$E_{\left(y_{i,t+1}^{0*} \mid y_{(1)i,t+1}^{*} = \tilde{r}_{ik,t}\right)} \left[\Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y_{i,t+1}^{0*}, y_{i,t+1}^{1*} = \tilde{r}_{ik,t}, \Omega_{i,t}, e_{i,t}=1\right)\right], \tag{40}$$

where

$$\Pr\left(s_{i,t+1}=1, e_{i,t+1}=1 \mid y^{0*}_{i,t+1}, y^{1*}_{i,t+1}=\tilde{r}_{ik,t}, \Omega_{i,t}, e_{i,t}=1\right) \tag{41}$$
$$= \frac{e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\tilde{r}_{ik,t}\right)+b^q_1\mu_i+b^q_2\eta_i}}{1+e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\tilde{r}_{ik,t}\right)+b^q_1\mu_i+b^q_2\eta_i}+e^{x^{s\prime}_{i,t+1}\gamma^s+\delta^s_1 y^{0*}_{i,t+1}+b^s_1\mu_i+b^s_2\eta_i}}.$$

The formulation (40)-(41) 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 $\tilde{r}_{ik,t}$" is understood by the respondent as meaning that $\varepsilon^{\upsilon}_{i,j(t+1)}$ is set to:

$$\varepsilon^{\upsilon}_{i,j(t+1)} = \tilde{r}_{ik,t} - y^{c_{t+1}}_{i,t}. \tag{42}$$

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 $\tilde{r}_{ik,t}$ 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 (39) as an average of model probabilities with respect to the unconditional distribution of $\varepsilon^{\omega}_{i,t+1}$ as follows:

$$E_{\varepsilon^{\omega}_{i,t+1}}\left[\frac{e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\tilde{r}_{ik,t}\right)+\delta^q_1\varepsilon^{\omega}_{i,t+1}+b^q_1\mu_i+b^q_2\eta_i}}{1+e^{x^{q\prime}_{i,t+1}\gamma^q+\delta^q_1\left(\tilde{r}_{ik,t}\right)+\delta^q_1\varepsilon^{\omega}_{i,t+1}+b^q_1\mu_i+b^q_2\eta_i}+e^{x^{s\prime}_{i,t+1}\gamma^s+\delta^s_1\widehat{y}^{0*}_{i,t+1}+\delta^s_1\varepsilon^{\omega}_{i,t+1}+b^s_1\mu_i+b^s_2\eta_i}}\right]. \tag{43}$$

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

$$\Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid \tilde{r}_{ik,t} < y^{1*}_{i,t+1} \leq \tilde{r}_{ik+1,t}, \Omega_{i,t}, e_{i,t}=1\right)$$
$$= \frac{\Pr\left(\tilde{r}_{ik,t} < y^{1*}_{i,t+1} \leq \tilde{r}_{ik+1,t} \mid s_{i,t+1}=0, e_{i,t+1}=1, \Omega_{i,t}, e_{i,t}=1\right)}{\Pr\left(\tilde{r}_{ik,t} < y^{1*}_{i,t+1} \leq \tilde{r}_{ik+1,t} \mid \Omega_{i,t}, e_{i,t}=1\right)}$$
$$\cdot \Pr\left(s_{i,t+1}=0, e_{i,t+1}=1 \mid \Omega_{i,t}, e_{i,t}=1\right) \tag{44}$$

This quantity would be observable if the first term in the numerator could be constructed from the question about the probability distribution of earnings conditional on staying in the same job which we described in Appendix A.2, which is unclear.[24] The second is

[24] Letting $\Upsilon_{i,t} = \{s_{i,t+1}=0, e_{i,t+1}=1, \Omega_{i,t}, e_{i,t}=1\}$ for shortness, this question relates to

the "percent chance of 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:

$$p^{0}_{ki,t} \equiv \Pr\left(s_{i,t+1} = 0, e_{i,t+1} = 1 \mid y^{1*}_{i,t+1} = \tilde{r}_{ik,t}, \Omega_{i,t}, e_{i,t} = 1\right), \tag{45}$$

which would still require a mapping with the model probabilities similar to (40)-(41) or (43).

As already pointed out in the main text, a simpler version of the model would be one in which there is endogenous self-selection in $\varepsilon^{\upsilon}_{i,j(t+1)}$ but not in $\varepsilon^{\omega}_{i,t+1}$. In such case, $y^{0*}_{i,t+1}$ would be replaced by $\widehat{y}^{0*}_{i,t+1} = E\left(y^{0*}_{i,t+1} \mid \Omega_{i,t}\right)$ and there would be no need to consider the expectations in (41) or (43). Thus, having the linear equations:

$$\ell^{e1}_{ki,t} = \ln\left(\frac{p^{e1}_{ki,t}}{1-p^{e1}_{ki,t}-p^{e0}_{kit}}\right) = x^{1\prime}_{i,t+1}\gamma^{1} + \delta^{1}_{y}\left(\tilde{r}_{ik,t}\right) + b^{1}_{\mu}\mu_i + b^{1}_{\eta}\eta_i + \xi^{e1}_{ki,t}, \tag{46}$$

$$\ell^{e0}_{ki,t} = \ln\left(\frac{p^{e0}_{ki,t}}{1-p^{e1}_{ki,t}-p^{e0}_{ki,t}}\right) = x^{0\prime}_{i,t+1}\gamma^{0} + \delta^{0}_{y}\widehat{y}^{0*}_{i,t+1} + b^{0}_{\mu}\mu_i + b^{0}_{\eta}\eta_i + \xi^{e0}_{ki,t}. \tag{47}$$

which correspond to equations (24) and (25) in the main text.

B.3 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 $\widehat{y}^{0*}_{i,t+1}$. Hence, we can express it as:

$$\begin{aligned}\widehat{y}^{0*}_{i,t+1} &= E\left(y^{0*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1} = e_{i,t} = 1, s_{i,t+1} = 0\right) \\ &= \rho y^{*}_{i,t} + \left(x_{i,t+1} - \rho x_{it}\right)'\gamma + (1-\rho)\,\mu_i + (1-\rho)\,\upsilon_{i,j(t).}\end{aligned} \tag{48}$$

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

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

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

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 $\overline{y}_{i,t}$ and derive the conditional expectation as follows

$$\begin{aligned}\overline{y}_{i,t} &= E\left(y^{0*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1}=e_{i,t}=1, s_{i,t+1}=0\right) \Pr\left(e_{i,t+1}=1, s_{i,t+1}=0 \mid \Omega_{it}, e_{i,t}=1\right) \\ &\quad + E\left(y^{1*}_{i,t+1} \mid \Omega_{it}, e_{i,t+1}=e_{i,t}=1, s_{i,t+1}=1\right) \Pr\left(e_{i,t+1}=1, s_{i,t+1}=1 \mid \Omega_{it}, e_{i,t}=1\right).\end{aligned} \tag{49}$$

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. Apart from staying in the same job and changing jobs, this measure incorporates the fact that individuals can transition into unemployment and hence have zero earnings. Individuals who are self-employed or expect to be self-employed with positive probability are not included in our sample.

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 estimate equation 25 instrumenting it using the 12-month conditional question on staying in the same job, which we described in Appendix A.2.

C Estimation and empirical implementation

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

As argued in Section 5, one empirical complication is that, while we observe $p^{e1}_{ki,t}$, i.e., the probability of changing jobs conditional on receiving an offer in the kth bin, we do not directly observe $p^{e0}_{ki,t}$ nor $p^{eu}_{ki,t}$, i.e., the conditional probabilities of staying on the same job or going into unemployment given that same job offer. These probabilities are necessary for the estimation of Equations 24 and 25.

To approximate them, we use additional information available in the survey, namely, the unconditional probabilities of transitioning into a new job, staying on the same job and moving into unemployment over the next four months, which we denote $p^{e1}_{i,t}$, $p^{e0}_{i,t}$ and $p^{eu}_{i,t}$ respectively. Recalling that we denote $p^{c_{t+1}}_{ki,t}$ the probability of receiving a job offer in a certain bin, it must hold that $p^{e0}_{i,t} = \sum_{k=1}^{K} p^{e0}_{ki,t} p^{c_{t+1}}_{ki,t}$ and $p^{ue}_{i,t} = \sum_{k=1}^{K} p^{ue}_{ki,t} p^{c_{t+1}}_{ki,t}$

However, we still need an assumption to recover $p^{e0}_{k,it}$ and $p^{eu}_{k,it}$ from these equalities. We proceed by assuming that the probability of unemployment is concentrated in the lowest part of the offer distribution. Namely, we assume that there is a *threshold offer* $R'_{i,t}$ such that $p^{eu}_{ki,t} = 1 - p^{e1}_{ki,t}$ for bins k such that their upper bound $R_{k+1i,t} < R'_{i,t}$ and $p^{eu}_{ki,t} = 0$ for bins k such that their lower bound $R_{ki,t} > R'_{i,t}$. In intuitive terms, we assume that if people reject a high offer, it is because they plan to stay in the same job, while moves to unemployment may be associated with the receipt of low outside offers. For the bin $k*$ for which $R_{k*i,t} < R'_{i,t} < R_{k*+1i,t}$, we set $p^{eu}_{k^*i,t} = 1 - p^{e1}_{k^*i,t} - \frac{p^{eu}_{i,t} - \sum_{k=1}^{k^*-1} p^{eu}_{ki,t} p^{c_{t+1}}_{ki,t}}{p^{c_{t+1}}_{k^*i,t}}$.

C.2 Keeping zero and one probabilities

We follow the same procedure as in Arellano et al. (2024) to accommodate elicited probabilities that take values equal to zero or one. The logit transformation used in the main analysis requires probabilities $p_{ki,t}$ to lie strictly between zero and one, but boundary responses naturally arise in the presence of rounding and elicitation error, which we capture through the measurement-error term $\xi_{ki,t}$. To retain these observations, we apply a regularized logit transformation,

$$\ell_{ki,t} = \text{logit}(\tilde{p}_{ki,t}), \qquad \tilde{p}_{ki,t} = \frac{p_{ki,t} + \frac{k}{2m}}{1 + \frac{K}{2m}},$$

which maps boundary probabilities into the interior of the unit interval. This transformation generalizes the modified logit proposed by Cox and Snell (1970) and admits an interpretation in which $\xi_{ki,t}$ reflects finite-sample noise from a hypothetical sample of size m, so that $m = O(1/\sigma^2_\xi)$. As shown in Arellano et al. (2024), this adjustment provides a coherent way to account for elicitation error in subjective probability data.

C.3 Variance and covariance structure

C.3.1 General minimisation problem

In Step 2 we estimate a set of parameters using the variance–covariance structure implied by the composite residuals obtained in Step 1. These moments incorporate both covariances across equations and autocovariances across survey waves.

Let $\hat{\lambda}_{ikt}$ denote the composite residual from Step 1 for individual $i = 1, \ldots, N$, equation $k = 1, \ldots, K$, and survey wave $t = 1, \ldots, T$. We consider moments of the form

$$V_{k\ell c} = E[\lambda_{ikt}\lambda_{i\ell,t+c}],$$

where k and ℓ index equations and c indexes the type of covariance considered.

In our application $T = 3$ and $C = 6$. The index c distinguishes the type of covariance according to the job mobility pattern between survey waves:

- $c = 1$ refers to contemporaneous covariances across equations,
- $c = 2$ refers to first–order autocovariances for individuals who stay in the same job,
- $c = 3$ refers to first–order autocovariances for individuals who change jobs,
- $c = 4$ refers to second–order autocovariances for individuals who stay in the same job throughout,
- $c = 5$ refers to second–order autocovariances for individuals who change jobs once, and
- $c = 6$ refers to second–order autocovariances for individuals who change jobs twice.

For each individual we define the moment contribution

$$m_{i,k\ell c} = \hat{\lambda}_{ikt}\hat{\lambda}_{i\ell,t+c},$$

with the appropriate job–transition indicators included where relevant. The corresponding sample moments are

$$\hat{V}_{k\ell c} = \frac{1}{N}\sum_{i=1}^{N} m_{i,k\ell c}.$$

Because each $K \times K$ covariance matrix is symmetric, only the lower triangular elements are used in estimation.

Stacking all distinct elements of $\hat{V}_{k\ell c}$ across k, ℓ, and c yields the vector of empirical moments

$$\hat{S} = \frac{1}{N}\sum_{i=1}^{N} m_i,$$

where m_i collects all individual moment contributions $m_{i,k\ell c}$. The dimension of this vector is

$$M = \frac{K(K+1)}{2}C.$$

Let $S(\theta)$ denote the vector of moments implied by the model. Parameters are estimated by minimizing the quadratic form

$$Q(\theta) = (\hat{S} - S(\theta))' W^{-1} (\hat{S} - S(\theta)),$$

where W is a diagonal weighting matrix.

C.3.2 Sample weights

Each element of $\hat{S}$ is a sample mean. If m_{ik} denotes the k-th element of m_i, then

$$Var(\hat{S}_k) = \frac{1}{N} Var(m_{ik}).$$

Let W denote the diagonal weighting matrix collecting these sampling variances:

$$W = diag\left(Var(\hat{S}_1), \ldots, Var(\hat{S}_M)\right).$$

Restricting the weighting matrix to be diagonal amounts to ignoring the covariances between different moments. This choice ensures that the matrix is invertible and positive semidefinite while still capturing the different sampling variability of the individual moments.

C.3.3 Using sample fourth moments

An estimator of the sampling variance of each moment can be obtained using sample fourth moments. Specifically,

$$\widehat{Var}(\hat{S}_k) = \frac{1}{N^2}\sum_{i=1}^{N}(m_{ik} - \hat{S}_k)^2. \tag{50}$$

C.3.4 Using fourth moments under normality

For numerical stability, in the results reported in the paper we replace the empirical fourth moments in equation (50) with the values implied by the assumption that the residuals are jointly normally distributed.

Fix (k, ℓ, c) and define

$$X_i \equiv \lambda_{ikt}, \qquad Y_i \equiv \lambda_{i\ell,t+c}.$$

Since $V_{k\ell c} = E[X_i Y_i]$ and $\hat{V}_{k\ell c} = N^{-1} \sum_i X_i Y_i$, we have

$$Var(\hat{V}_{k\ell c}) = \frac{1}{N} Var(X_i Y_i) = \frac{1}{N}\Big(E[X_i^2 Y_i^2] - E[X_i Y_i]^2\Big).$$

Under joint normality and zero means,

$$E[X_i^2 Y_i^2] = E[X_i^2]E[Y_i^2] + 2\,E[X_i Y_i]^2,$$

so that

$$Var(X_i Y_i) = E[X_i^2]E[Y_i^2] + E[X_i Y_i]^2.$$

Replacing population moments with their sample counterparts yields

$$\widehat{Var}(\hat{V}_{k\ell c}) = \frac{1}{N}\left(\hat{V}_{kk1}\hat{V}_{\ell\ell 1} + \hat{V}_{k\ell c}^2\right).$$

D Additional results and robustness

D.1 First stage estimation

Table 11: First step - Men

	(1) $y_{i,t}^{c_{t+1}}$	(2) $l_{ki,t}^{c_{t+1}}$
$y_{i,t}^{*}$	0.500***	-4.637***
	(0.0906)	(0.782)
Job tenure	-0.00211	0.0123
	(0.00204)	(0.0194)
$r_{ki,t}$		8.889***
		(0.387)
Constant	2.108***	-17.25***
	(0.372)	(3.335)
ρ	0.500	0.522
S.E.(ρ)	0.0906	0.0870
Observations	2948	14600
R^2	0.133	0.573

Standard errors in parentheses

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

Table 12: First step - Women

	(1) $y_{i,t}^{c_{t+1}}$	(2) $l_{ki,t}^{c_{t+1}}$
$y_{i,t}^{*}$	0.446***	-2.349***
	(0.141)	(0.795)
Job tenure	0.00547	-0.0229
	(0.00427)	(0.0267)
$r_{ki,t}$		5.208***
		(0.560)
Constant	1.833***	-8.886***
	(0.521)	(3.439)
ρ	0.446	0.451
S.E.(ρ)	0.141	0.157
Observations	2856	14075
R^2	0.0659	0.331

Standard errors in parentheses

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

D.2 Alternative assumptions for individual fixed effects

In this Appendix, we study the possibility that the mobility fixed effect η_i that enters the job-to-job transition equations is different to the fixed effect that enters the unemployment-to-employment transition equations.

More specifically, we modify Equation 12 in our baseline model as follows:

$$ue^{*}_{i,t+1} = x^{u\prime}_{i,t+1}\gamma^{u} + \delta^{u}_{y} y^{1*}_{i,t+1} + b^{u}_{\mu}\mu_i + b^{u}_{\theta}\theta_i + \varepsilon^{u}_{i,t+1}. \tag{51}$$

while keeping Equation 6 and 7 unchanged. As a result, we impose that the trait of individual heterogeneity that determines how mobile an individual is across jobs is different from that which determines how mobile an individual is between unemployment and employment. Such a situation can arise, for example, if there is large individual heterogeneity in the taste or distaste for work.

Therefore, the residuals we use for the GMM procedure in this case look as follows:

$$\lambda^{yc}_{i,t} = (1-\rho)\,\mu_i + (\phi-\rho)\,\upsilon_{i,j(t)} + \xi^{yc}_{i,t}, \tag{18''}$$

$$\lambda^{\ell c}_{ki,t} = -\frac{1-\rho}{\sigma_e}\,\mu_i - \frac{\phi-\rho}{\sigma_e}\,\upsilon_{i,j(t)} + \xi^{\ell c}_{ki,t}, \tag{20''}$$

$$\lambda^{ue}_{ki,t} = b^{u}_{\mu}\mu_i + b^{u}_{\theta}\theta_i + \xi^{ue}_{ki,t}, \tag{23'''}$$

$$\lambda^{e0}_{ki,t} = b^{0}_{\mu}\mu_i + b^{0}_{\eta}\eta_i + \xi^{e0}_{ki,t}, \tag{24''}$$

$$\lambda^{e1}_{ki,t} = b^{1}_{\mu}\mu_i + b^{1}_{\eta}\eta_i + \xi^{e1}_{ki,t}. \tag{25''}$$

We report the results in Table 13. As in our baseline case, we normalize $b^{1}_{\eta} = 1$ and $b^{u}_{\theta} = 1$.

We find that, both for the men and the women samples, the estimates are very similar to those of our baseline model which we reported in Table 3. For women, the standard deviation of θ is highly significant and of a similar magnitude to that of men, which contrasts with the negative and insignificant result for β^{u}_{η} in our main specification. This result suggests that there might be substantial heterogeneity in terms of a (dis)taste for work in the women sample.

For men, instead, the coefficient on $SD(\theta)$ is very similar to the product of $b^{u}_{\eta}SD(\eta)$

Table 13: Individual fixed effects and transitions, separating mobility fixed effect in UE transitions (θ) and mobility fixed effect in quitting and staying transitions (η)

	Coefficient	Men	Women
Effect of μ	b_μ^u	-4.943***	-1.637**
on Pr(working)		(1.606)	(0.725)
Standard deviation of θ	$SD(\theta)$	2.375***	2.438***
		(0.478)	(0.389)
Effect of μ	b_μ^1	-4.043***	-3.022***
on Pr(quitting)		(0.259)	(0.358)
Standard deviation of η	$SD(\eta)$	0.509**	0.364**
		(0.200)	(0.158)
Effect of μ	b_μ^0	-0.361***	-0.324***
on Pr(staying)		(0.065)	(0.068)
Effect of η	b_η^0	-1.409***	-1.996***
on Pr(staying)		(0.275)	(0.399)

in our main results, reflecting that this new factor affects transitions to employment in a very similar way to η in our main results. Although both factors are found to be independently significant, which suggests that there is no strong evidence in the data that they are correlated, it should be noted that the identification of this correlation relies on individuals that actually transition between employment and unemployment in our sample, which is a relatively small set. At the same time, in this specification the estimation of b_η^0 is less precise than in our main results, suggesting that grouping both factors into one adds efficiency to the estimation.

D.3 Estimation excluding the questions on the probability of staying

Table 14: Persistence and risk parameters

	Coefficient	Men	Women
Persistence in productivity	ρ	0.511***	0.448***
		(0.087)	(0.153)
Pers. job-specific component	ϕ	0.015	0.000
		(0.272)	(0.125)
SD individual FE	σ_μ	0.618***	0.565***
		(0.052)	(0.032)
Logit scale of $(\varepsilon^{\omega}_{i,t+1} + \varepsilon^{v}_{ij,t+1})$	σ_e	0.111***	0.182***
		(0.026)	(0.002)
SD job-specific component	σ_v	0.339	0.578**
		(0.428)	(0.236)

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

Table 15: Individual fixed effects and transitions

	Coefficient	Men	Women
Standard deviation of η	$SD(\eta)$	0.962**	0.236***
		(0.448)	(0.041)
Effect of μ	b^u_μ	-3.793**	-1.959***
on Pr(working)		(1.479)	(0.646)
Effect of η	b^u_η	2.999	-9.999
on Pr(working)		(7.241)	(9.583)
Effect of μ	b^1_μ	-3.539***	-3.102***
on Pr(quitting)		(0.543)	(0.126)

E Simulations and decompositions

E.1 Implementation of the simulation

To decompose the relative contribution of ability, job-specific human capital and selection into and out of unemployment and job changes, we simulate our earnings process using our estimated parameters and estimate a number of regressions in the resulting simulated data as described in Section 8.

Specifically, we simulate the earnings process in Equations 1 to 4, the latent values of different employment options (Equations 8, 9 and 12) and the associated log odds ratios (Equations 10, 11, 13) which determine employment transitions.

To do this, we use our estimated parameters, which we report in Tables 1, 2 and 3 , but we also need to take a stance about some additional objects and, importantly, about the distributions of the different components. We highlight that these assumptions, while necessary for our simulation, are not needed for our main results, which are therefore more general.

We simulate the individual fixed effects μ_i, η_i and the initial $\upsilon_{i,j(0)}$ from independent normal distributions with standard deviations σ_μ , σ_η and σ_υ. We also assume that $\omega_{i,0}$ has a normal distribution. As for σ_ω, we estimate it directly from observed data in the SCE on non-job changers and find it to be 0.163.[25] We separate $\sigma_{\varepsilon^\omega}$ and $\sigma_{\varepsilon^\upsilon}$ in the manner described in footnote 17. Following the assumptions described in Section 5, we assume that all shocks follow a logistic distribution. We fix the initial number of unemployed in the simulation at 3%.

We simulate $N = 10,000$ individuals over $T = 100$ time periods (4-month long, equivalent to the periods in our SCE sample). We need to simulate a relatively large T because the equations we estimate in our simulated data are subject to Nickell (1981) bias, unlike our main results. Results are not sensitive to increasing N.

The procedure for the simulation is as follows for each worker i:

1. We take a draw of μ_i and η_i, of the initial conditions $\omega_{i,0}$, $\upsilon_{i,j(0)}$, and of the initial

[25] In our realized earnings data, we look at people who do not change jobs between $t = 0$ and $t = 1$ and compute $Var(y_0)$ and $Cov(y_0, y_1)$. For non-job changers, the υ component gets subsumed into the fixed effect μ and we have $Var(y_0) = Var(\mu + \upsilon) + Var(\omega_0)$ and $Cov(y_0, y_1) = Var(\mu + \upsilon) + \rho Var(\omega_0)$, from which we can easily obtain $Var(\omega_0)$ given our estimated ρ.

employment status e_{i0}.

2. We simulate the $\omega_{i,t}$ process, which is independent of employment transitions.

3. In each period t:

   (a) We randomly generate a new *potential* $v^1_{i,j(t+1)}$ in the event of job change/transition into employment.

   (b) We use it to obtain the values $ee^{0*}_{i,t+1}$, $ee^{1*}_{i,t+1}$, $ue^{*}_{i,t+1}$

   (c) We use these values to obtain the probabilities of accepting the offer and thus $p^0_{i,t}$, $p^1_{i,t}$, $p^{ue}_{i,t}$, depending on the employment status.

   (d) For each worker, we draw shocks $\varepsilon^0_{i,t+1}$, $\varepsilon^1_{i,t+1}$ and $\varepsilon^u_{i,t+1}$ from a uniform distribution. These determine whether, given a certain probability of accepting the offer, a certain worker actually does it or not.

   (e) The worker's choices determine their employment status in period $t+1$ $e_{i,t+1}$, their job change status $jc_{i,t+1}$ and their new $v_{i,j(t+1)}$.

To reflect that, in the observed data, the probability 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.[26]

As a result of the simulation, we obtain a panel of latent earnings $y^*_{i,t}$ on which we estimate the regressions in Equations 33 to 32.

E.2 Interpreting Autoregressive Estimates in Simulated Earnings Data

In this appendix we discuss how to interpret the reduced-form AR(1) regressions on simulated realized earnings in terms of the structural model. Recall that for employed

[26] 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.

individuals ($e_{i,t+1} = 1$), latent log earnings satisfy

$$y^*_{i,t+1} = x'_{i,t+1}\gamma + \mu_i + \omega_{i,t+1} + \upsilon_{i,j(t+1)}, \tag{52}$$

with persistent idiosyncratic productivity

$$\omega_{i,t+1} = \rho\, \omega_{i,t} + \varepsilon^{\omega}_{i,t+1}. \tag{53}$$

The match component $\upsilon_{i,j(t+1)}$ is constant within a spell and evolves only upon switching:

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

Observed earnings satisfy $y_{i,t+1} = y^*_{i,t+1} \times e_{i,t+1}$, so the distribution of observed earnings is conditional on the endogenous employment and mobility choices at $t+1$.

To interpret reduced-form AR(1) regressions on simulated realized earnings, it is useful to separate two aspects: (i) *persistent components* in $y^*_{i,t+1}$ beyond $\omega_{i,t+1}$, and (ii) *endogenous selection* into employment and job transitions due to the dependence of transition utilities on $y^{0*}_{i,t+1}$ and $y^{1*}_{i,t+1}$ through the coefficients δ^0_y, δ^1_y, and δ^u_y.

Recall the reduced form equations that we consider in the main text are:

$$\begin{aligned} y_{i,t+1} &= \beta^A_0 + \beta^A_1 y_{i,t} + \varepsilon^A_{i,t+1} \\ y_{i,t+1} &= \beta^B_0 + \beta^B_1 y_{i,t} + \alpha_i + \varepsilon^B_{i,t+1} \\ y_{i,t+1} &= \beta^C_0 + \beta^C_1 y_{i,t} + \alpha_i + \zeta_{i,j(t)} + \varepsilon^C_{i,t+1}. \end{aligned}$$

Pooled AR(1) specification. Consider the reduced-form regression (estimated on observed earnings),

$$y_{i,t+1} = \beta^A_0 + \beta^A_1 y_{i,t} + \varepsilon^A_{i,t+1} \tag{54}$$

Abstracting from selection, β^A_1 captures all persistent components of the structural earnings equations (52)–(53), including permanent ability μ_i and the match component $\upsilon_{i,j(t)}$ in addition to the individual-level productivity $\omega_{i,t}$.

Selection through both time-invariant heterogeneity and time-varying earnings components further amplifies persistence in (54). First, transition utilities depend directly on permanent ability and mobility type:

$$ee^{0*}_{i,t+1} = \cdots + b^0_\mu \mu_i + b^0_\eta \eta_i + \varepsilon^0_{i,t+1},$$
$$ee^{1*}_{i,t+1} = \cdots + b^1_\mu \mu_i + b^1_\eta \eta_i + \varepsilon^1_{i,t+1},$$
$$ue^{*}_{i,t+1} = \cdots + b^u_\mu \mu_i + b^u_\eta \eta_i + \varepsilon^u_{i,t+1},$$

so the event of observing earnings at $t+1$ (i.e., $e_{i,t+1} = 1$) is positively associated with (μ_i, η_i) whenever the relevant b_μ and b_η coefficients are nonzero. Because μ_i and η_i are perfectly persistent over time, this induces dynamic selection on permanent types: individuals with particular values of μ_i and η_i are disproportionately represented among employed observations in consecutive periods, strengthening the observed serial correlation between $y_{i,t}$ and $y_{i,t+1}$.

Second, choices at $t+1$ also depend on next-period latent earnings via $\delta^0_y y^{0*}_{i,t+1}$, $\delta^1_y y^{1*}_{i,t+1}$, and $\delta^u_y y^{1*}_{i,t+1}$. Because $y^{0*}_{i,t+1}$ and $y^{1*}_{i,t+1}$ are driven by $\omega_{i,t+1} = \rho\omega_{i,t} + \varepsilon^\omega_{i,t+1}$ and (upon switching) by $\upsilon_{i,j(t+1)}$, conditioning on being employed at $t+1$ and on the realized mobility outcome further selects on realized productivity and match components.

Taken together, omitted persistent components and endogenous selection on (μ_i, η_i, $\omega_{i,t+1}$, $\upsilon_{i,j(t+1)}$) push $\widehat{\beta}^A$ upward relative to the structural persistence parameter ρ.

AR(1) with individual fixed effects. With individual fixed effects, the estimated regression is

$$y_{i,t+1} = \beta^B_0 + \beta^B_1 y_{i,t} + \alpha_i + \varepsilon^B_{i,t+1},$$

or:

$$y_{i,t+1} - \bar{y}_i = \beta^B_1 \big(y_{i,t} - \bar{y}_i\big) + \varepsilon^B_{i,t+1}, \tag{55}$$

where $\bar{y}_i$ denotes the within-individual mean. The within transformation removes the permanent ability component μ_i from the earnings equation. Using (52), and again ignoring selection, within-individual deviations can be written as

$$y^*_{i,t+1} - \bar{y}_i = \big(\omega_{i,t+1} - \bar{\omega}_i\big) + \big(\upsilon_{i,j(t+1)} - \bar{\upsilon}_i\big) + \big(x'_{i,t+1}\gamma - \overline{x'_i\gamma}\big). \tag{56}$$

Thus, although μ_i is purged from the regression, residual match heterogeneity $(\upsilon_{i,j(t+1)} - \bar{\upsilon}_i)$ remains; because $\upsilon_{i,j(t+1)}$ is constant within spells, this residual behaves as a spell-level component over adjacent observations within spells, which mechanically inflates persistence in within-worker earnings. In addition, individual fixed effects do not eliminate selection driven by persistent individual heterogeneity. While μ_i drops out of the within regression, both μ_i and η_i continue to enter the transition utilities, affecting the probability of being employed and of switching at $t+1$.

Hence, even after conditioning on worker fixed effects, the sample of observed earnings remains endogenously selected on persistent heterogeneity through η_i, as well as on the realized time-varying components that enter $y^*_{i,t+1}$ (notably $\omega_{i,t+1}$ and, depending on the mobility outcome, $\upsilon_{i,j(t+1)}$). This residual selection, combined with the match-level persistence of $\upsilon_{i,j(\cdot)}$, helps explain why $\widehat{\beta}^B$ can remain substantially above the structural persistence ρ.

AR(1) with match (individual–job) fixed effects. In the third specification, we absorb heterogeneity with a single fixed effect that is specific to the individual–job match, denoted $\tilde{\alpha}_{ij(t+1)} = \alpha_i + \zeta_{ij(t+1)}$. The reduced-form regression can be written as

$$y_{i,t+1} = \beta_0^C + \beta_1^C y_{i,t} + \tilde{\alpha}_{ij(t+1)} + \varepsilon^C_{i,t+1},$$

and the within-transformed form is

$$y_{i,t+1} - \bar{y}_{ij} = \beta^C \big(y_{i,t} - \bar{y}_{ij}\big) + \varepsilon^C_{i,t+1}, \tag{57}$$

where $\bar{y}_{ij}$ denotes the mean within the individual–job match.

Under the model, within a spell the match component $\upsilon_{i,j(\cdot)}$ is constant. Hence, and abstracting for the moment from selection, the match fixed effect absorbs the level component

$$\tilde{\alpha}_{i,j(t+1)} \approx \mu_i + \upsilon_{i,j(t+1)} + (\text{job-level mean of } x'_{i,\cdot}\gamma).$$

As a result, within a spell we have the approximation

$$y^*_{i,t+1} - \bar{y}_{ij} \approx \big(\omega_{i,t+1} - \bar{\omega}_{ij}\big) + \big(x'_{i,t+1}\gamma - \overline{x'_{ij}\gamma}\big), \tag{58}$$

so that the remaining serial correlation in the transformed outcome is primarily driven by the dynamics of $\omega_{i,t+1}$ in (53), rather than by permanent ability or job-specific match heterogeneity. Consequently, the estimated persistence $\widehat{\beta}_1^C$ is substantially closer to the structural persistence parameter ρ.

Selection remains present because employment and mobility decisions are endogenous. However, once job-level match heterogeneity is absorbed by the fixed effect, most non-ω sources of persistence are captured by the specification. As a result, the autoregressive coefficient primarily reflects the persistence of $\omega_{i,t}$ and is therefore much closer to the structural parameter ρ.

E.3 Quantifying the Heterogeneous-Mean Contribution

To assess the magnitude of the reduced-form persistence coefficients, it is useful to apply directly the heterogeneous-mean bias formula as explained in Arellano (2003).

Consider the abstract AR(1) model with generic heterogeneous mean Γ_i:

$$y_{i,t} = \Gamma_i + u_{i,t}, \qquad u_{i,t} = \rho u_{i,t-1} + \varepsilon_{i,t},$$

where $|\rho| < 1$, $\varepsilon_{i,t}$ is i.i.d. with variance σ_ε^2, and independent of Γ_i. Let $\sigma_\Gamma^2 = \text{Var}(\Gamma_i)$ and $\sigma_u^2 = \text{Var}(u_{i,t}) = \sigma_\varepsilon^2/(1-\rho^2)$.

Let ρ_r denote the population pooled regression coefficient of $y_{i,t+1}$ on $y_{i,t}$ and a constant. Then (Arellano, 2003, eq. 5.16):

$$\rho_r - \rho = \frac{(1-\rho)\lambda_r}{1+\lambda_r}, \qquad \lambda_r = \frac{\sigma_\Gamma^2}{\sigma_u^2}. \tag{59}$$

In our setting, this decomposition is useful because it allows us to infer the relative contribution of heterogeneous means to the differences in persistence measures that we reported in Section 8.

Pooled regression (Step 1). To do so, we begin by looking at Equation 30, in which we run a reduced-form regression of earnings on their lag to obtain their persistence. Based on these estimates, we are interested in separating the contribution of the structural persistent component $\omega_{i,t}$ to persistence from that of all other factors, including individual heterogeneity and job-specific effects. Therefore, we map the heterogeneous mean in the reduced-form representation to:

$$\Gamma_i \equiv \tilde{\alpha}_{i,j(t+1)} = \alpha_i + \zeta_{i,j(t+1)}.$$

so that the reduced-form persistent component $u_{i,t}$ summarizes the serially correlated part of earnings once the heterogeneous mean $\tilde{\alpha}_{i,j(t+1)}$ is removed; in our structural model, this component corresponds primarily to $\omega_{i,t}$ (abstracting from transition dynamics and selection) and its persistence ρ is equal to our structural estimate for the persistence of $\omega_{i,t}$.

Using the estimated values for the reduced form parameter $\beta_1^A = 0.975$ and for the structural parameter $\rho = 0.512$, equation (59) implies

$$0.975 - 0.512 = \frac{(1 - 0.512)\lambda_A}{1 + \lambda_A}.$$

Solving yields $\lambda_A = \frac{\text{Var}(\tilde{\alpha})}{\text{Var}(u)} = 18.52$.

Thus, in the pooled regression, equation (59) implies that the variance of the heterogeneous mean $\alpha_i + \zeta_{i,j(t+1)}$ is approximately 18.5 times the variance of the persistent reduced form component u_{it}.

Individual fixed effects (Step 2). We now turn to the estimates from Equation 31. In the fixed effects regression, the within-transformation removes effects that are fixed for an individual over time. To describe exactly what this entails, we further decompose the job specific component as:

$$\zeta_{i,j(t+1)} = \kappa_i + \xi_{i,j(t+1)},$$

where κ_i is the projection of $\zeta_{i,j(t+1)}$ onto the individual dimension and $\xi_{i,j(t+1)}$ is orthogonal to it. Therefore,

$$\tilde{\alpha}_{i,j(t+1)} = \alpha_i + \kappa_i + \xi_{i,j(t+1)}.$$

In Equation 31, α_i and κ_i are removed by the within-transformation and $\Gamma_i = \xi_{i,j(t+1)}$, while $u_{i,t}$ still corresponds roughly to $\omega_{i,t}$. Let ρ_B denote the population coefficient corresponding to regression (31), i.e. β_1^B. Using the same formula as in (59), with

$$\lambda_B = \frac{\mathrm{Var}(\xi)}{\mathrm{Var}(u)},$$

we have

$$\rho_B - \rho = \frac{(1-\rho)\lambda_B}{1+\lambda_B}.$$

Using the estimated values $\beta_1^B = 0.885$ and $\rho = 0.512$, we obtain

$$0.885 - 0.512 = \frac{(1-0.512)\lambda_B}{1+\lambda_B},$$

which implies $\lambda_B = 3.24$. Hence,

$$\frac{\lambda_B}{\lambda_A} = \frac{\mathrm{Var}(\xi)}{\mathrm{Var}(\tilde{\alpha})} = \frac{3.24}{18.52} = 0.175.$$

This implies that less than a fifth of the variance of the heterogeneous mean $\alpha_i + \zeta_{i,j(t+1)}$ is due to the residual match-specific component $\xi_{i,j(t+1)}$, while approximately 80% is captured by the worker-level component $\alpha_i + \kappa_i$.

Equivalently, as a crude measure of persistence in the heterogeneous mean,

$$\frac{\mathrm{Var}(\alpha+\kappa)}{\mathrm{Var}(\tilde{\alpha})} = 1 - \frac{\mathrm{Var}(\xi)}{\mathrm{Var}(\tilde{\alpha})} = 0.825.$$

This variance decomposition provides a quantitative interpretation of the decline in persistence from β_1^A to β_1^B: most of the heterogeneous-mean variance is absorbed by worker fixed effects, while a non-negligible residual match component continues to inflate the autoregressive coefficient relative to the structural persistence parameter.

E.4 Results for women

Table 16: Measured persistence in simulated data from different specifications.

Parameter	Value	Reduction in persistence controlling for...
β_1^A	0.961	
β_1^B	0.837	+ individual heterogeneity α_i
β_1^C	0.400	+ individual-employer fixed effects $\zeta_{ij(t)}$

Figure 5: Risk decomposition. The term ζ represents individual-employer fixed effects, while α represents individual fixed effects

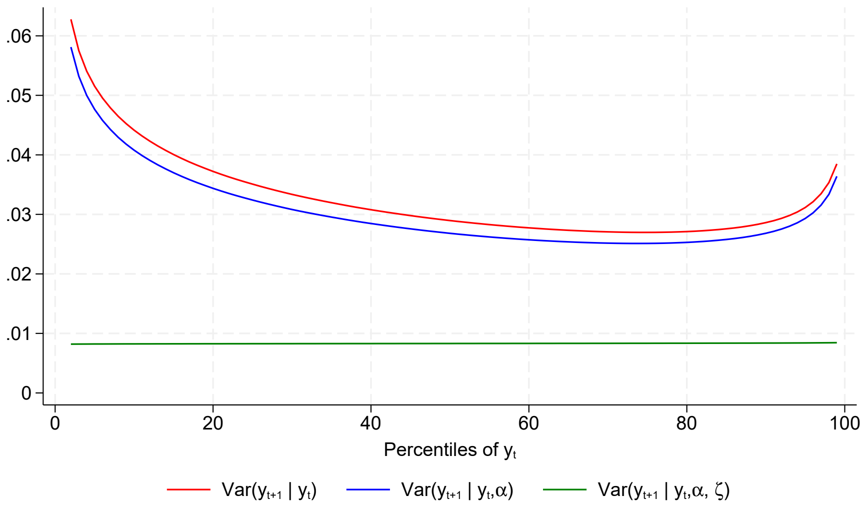


Table 17: Risk decomposition, averages of Figure 5 (top) and contributions of individual fixed effects, employer transitions and employer fixed effects, and remainder (bottom).

Risk Measure	Average	Contribution of...	
$Var(y_{t+1}\|y_t)$	0.033	Risk conditional on α and ζ	24.86%
$Var(y_{t+1}\|y_t,\alpha)$	0.031	Heterogeneity in ζ	67.97%
$Var(y_{t+1}\|y_t,\alpha,\zeta)$	0.008	Heterogeneity in α	7.17%