A Hidden Markov Model of Wages and Employment Mobility with Worker and Firm Heterogeneity: Evidence from Italian Register Data

Long Hong* Rasmus Lentz† Jean-Marc Robin‡

August 5, 2026

Abstract

We develop a joint wage-mobility model with two-sided heterogeneity, extending previous discrete mixture models by letting worker types evolve via an employer-dependent hidden Markov process. Estimation uses a Variational EM algorithm that jointly classifies worker and firm types, improving latent type classification over existing methods. Using matched employer–employee data from Veneto, Italy (1982–2001), we find that worker heterogeneity contributes more to log-wage variance, and firm heterogeneity less, than in a static two-way fixed-effect model, with a greater worker-firm covariance. Sorting rises over careers, driven by firm-type-dependent worker dynamics rather than mobility. Workers experience stronger wage-type growth when employed at high wage-type firms. Worker heterogeneity explain about 60% of wage variation at labor market entry and decline to roughly 50 percent after twenty years. Worker-type dynamics accounts for more than 85% of wage growth with experience. Non-employment generates persistent scarring effects, particularly for workers previously in high-type firms.

Keywords: Heterogeneity; Wage distributions; Employment and job mobility; Matched employer-employee data; Finite mixtures; EM algorithm; Classification algorithm; Sorting; Decomposition of wage inequality
JEL codes: E24; E32; J63; J64

---

*W. P. Carey School of Business, Arizona State University. E-mail: long.hong@asu.edu.
†University of Wisconsin-Madison. E-mail: rlentz@wisc.edu.
‡SciencesPo, Paris. Email `jeanmarc.robin@sciencespo.fr`. Jean-Marc Robin gratefully acknowledge the support from the European Research Council (grant reference ERC-2020-ADG-101018130).

1 Introduction

This paper develops a unified model of wage dynamics, worker and firm heterogeneity, and employment mobility. Workers and firms have latent types, and worker types evolve according to a hidden Markov process that may depend on the current firm type. Wages therefore reflect both sorting across firms and changes in latent worker productivity within and across matches.

The paper connects two largely separate literatures. The matched employer–employee literature identifies worker and firm wage heterogeneity from wages and mobility across employers, as in Abowd, Kramarz, and Margolis (1999), Card, Heining, and Kline (2013), Sorkin (2018), Kline, Saggio, and Sølvsten (2020), Bonhomme, Lamadon, and Manresa (2019), and Lentz, Piyapromdee, and Robin (2023). Related work incorporates labor market frictions, including Postel-Vinay and Robin (2002), Lise, Meghir, and Robin (2016), Hagedorn, Law, and Manovskii (2017), and Bagger and Lentz (2019). Because worker heterogeneity is usually fixed, wage dynamics arise mainly from shocks and mobility across firms.

A second literature models earnings as reduced-form stochastic processes with permanent and transitory components, often using random-walk or ARMA specifications. Contributions include Moffitt and Gottschalk (1995), Meghir and Pistaferri (2004), Browning, Ejrnæs, and Alvarez (2010), Altonji, Smith, and Vidangos (2013), Lochner and Shin (2014), Arellano, Blundell, and Bonhomme (2017), Hoffmann (2019), and Guvenen, Karahan, Ozkan, and Song (2021). These models flexibly describe the persistence, volatility, heteroskedasticity, and life-cycle covariance structure of earnings, but abstract from worker–firm matching, sorting, and employer-dependent earnings innovations.

Our framework combines the strengths of both approaches. It introduces persistent within-worker dynamics through latent worker-type transitions while retaining two-sided heterogeneity, sorting, and endogenous mobility. Earnings evolve through job changes and through systematic changes in worker productivity that depend on employment histories and firm types. In this approach, the paper is related to Bagger, Fontaine, Postel-Vinay, and Robin (2014), Lise and Postel-Vinay (2020), Taber and Vejlin (2020).

The emphasis on firm-specific learning environments also connects with Lentz and Roys (2024), Borovičková and Macaluso (2025), and Gregory (2026), that develop structural equilibrium models in which matches determine both current output and worker skill accumulation, estimated by simulated minimum distance. Gregory (2026) notably uses direct information on training from German matched employer–employee data. Also with a focus on firm-specific learning environments, Arellano-Bover and Saltiel (2026) and Gualdani, Pastorino, de Paula, and Salgado (2026) take more reduced-form approaches.

Methodologically, we extend the finite-mixture framework of Bonhomme, Lamadon, and Manresa (2019) and the classification method of Lentz, Piyapromdee, and Robin (2023). We introduce hidden Markov worker-type dynamics and jointly estimate worker and firm classifications from the full likelihood of wages and mobility. A variational expectation–maximization algorithm approximates the likelihood while jointly recovering firm types and worker-type trajectories.[1] Unlike pre-clustering approaches, both classifications use the model's full dynamic structure.

We establish identification from wages and employment transitions across consecutive spells. Under mild rank conditions and with at least three observed spells per worker, the parameters governing worker-type transitions, employment mobility, and wage distributions are identified up to label permutations. The argument exploits differences between wage dynamics within firms and across firms of the same type, together with transitions through non-employment.

We estimate the model using administrative matched employer–employee data from Italy's Veneto region for 1982–2001. The model reproduces employment rates, wage-growth profiles, and the life-cycle evolution of wage dispersion. Joint estimation also improves latent worker and firm classification relative to procedures that estimate the two sides separately.

The estimates identify three worker learning types with low, medium, and high wage-type growth. High-growth workers are disproportionately employed in white-collar occupations, while low-growth workers are predominantly in blue-collar occupations. Firms also differ in their learning environments, and these differences are positively related to firm wage types. For high-growth workers, annual wage-type growth is 4 percentage points greater at the highest- than at the lowest-wage firm type. For low- and medium-growth workers, the difference is about 1 percentage point.

This positive association between firm wage types and worker growth is consistent with Arellano-Bover and Saltiel (2026), who classify firms using k-means on firm wage-growth distributions and then estimate worker wage equations containing experience by firm type. Lentz and Roys (2024) interpret the same association as a natural implication of supermodular match surplus in worker skill and firm productivity, which gives more productive firms stronger incentives to provide training. Borovičková and

[1] See Balke, Bonhomme, and Lamadon (2026) for another recent application of variational methods. Balke et al. develop an indirect inference procedure to correct for a possible inconsistency of variational EM. However, in the case of a finite mixture model, we will argue that consistency has been shown in many network applications and is likely to hold here also.

Macaluso (2025) and Gregory (2026) find smaller correlations between firm productivity and learning environments, positive in the former and negative in the latter.

Worker heterogeneity accounts for the largest share of wage dispersion, although positive sorting between worker and firm wage types is also important. Sorting rises with experience through both worker-type evolution and mobility. Because worker types grow faster at high-wage firms, worker–firm type correlations increase over the life cycle even without substantial reallocation. Conditional on this firm-dependent growth, however, mobility dampens sorting at the margin by interrupting the type growth associated with high-type matches.

Four additional findings highlight the importance of worker-type evolution. First, the initial allocation of workers across jobs explains about 60 percent of wage variation at labor-market entry and about 50 percent after twenty years. Second, non-employment produces substantial and persistent scarring by slowing or reversing wage-type growth, especially for workers previously employed at high-wage firms. The average wage-type loss reaches 1.6 percent after seven years and remains 1.5 percent after twenty years. Third, worker wage-type growth accounts for more than 85 percent of wage growth with experience, with improved selection into firm types accounting for the remainder. Fourth, worker wage variance nearly triples over twenty years, almost entirely because of increased worker wage-type variance.

Thus, a worker's own type is the primary determinant of current wages, but that type evolves substantially and depends importantly on the type of firm with which the worker is matched.

Section 2 presents the model. Section 3 establishes identification. Section 4 presents the estimator's likelihood criterion and develops a variational expectation-maximization (VEM) algorithm for its maximization. Section 5 presents the estimation on and fit to Italian matched employer-employee data. Section 6-10 present the results of the estimation including wage variance decompositions, learning dynamics, wage sorting, and wage growth mean and variance decompositions. Section 11 concludes.

2 The model

We start by describing the matched employer-employee data used in this study. Then we construct a model for the data.

2.1 Matched employer-employee data

The panel starts in year $t = 1$ and ends in year $t = T$, a rather small, fixed value. If a worker quits her job within a year, the employment sequence is recorded. Say that the first 4 months are spent with the incumbent employer, the next 3 months the worker is unemployed until she is hired by another firm for the remaining 5 months. In this case, we say that the year contains three spells, two employed and one unemployed, and two different total earnings are recorded–one for each employment spell.[2]

We thus construct for each worker $i \in \{1, ..., I\}$ in the panel a sequence of observations indexed by the spell index $s \in \{1, ..., S_i\}$, comprising the following information: the unique employer ID, total earnings $EARN$, the spell length $D \in \{1, ..., 12\}$ in months, together with starting calendar time $TIME$, potential experience EXP (cumulated number of months spent working since leaving school) and job tenure TEN (cumulated number of months with the current employer) at the beginning of the spell. We write $W = \ln(EARN/D)$ for log monthly earnings, that we call log wages (or simply wages).[3]

Workers are drawn in the stock in the first survey year $t = 1$. They enter the panel with a non zero tenure and experience in this case. For $t > 1$, there is a flow of workers entering the labor market after school and an initial search period. For these workers we record trajectories from the first job, and tenure and experience are initially equal to zero.

2.2 Initial network

The initial network of employment matches is modeled as a Bipartite Degree-Corrected Stochastic Block Model (DC-SBM) (see Dasgupta, Hopcroft, and McSherry, 2004; Karrer and Newman, 2011; Larremore, Clauset, and Jacobs, 2014). Workers only match with firms. Workers and firms are initially clustered into a finite number of groups. The biclustering and initial matches are governed by three independent multinomial draws.

Firm nodes. We divide J employers indexed by $j \in \{1, ..., J\}$ into L groups $\ell \in \{1, ..., L\}$. Let $Z_j^f = \ell$ if firm j draws type ℓ and define the indicator variables $Z_{j\ell}^f \in \{0, 1\}$, with $\sum_{\ell=1}^{L} Z_{j\ell}^f = 1$ and $Z_{j\ell}^j = 1$ if and only if $Z_j^f = \ell$. Firms draw a type

[2]The choice of the one month period length is in anticipation of the implementation of the estimator on Italian data, as described in Section 5.

[3]Wages are subject to measurement error which is tied to start and end date of jobs. We could allow the wage density to depend on a dummy for spells lasting the entire year ($D = 12$). This would take care of both first and last year of spell.

$Z_j^f = \ell$ with probability π_ℓ^f. Lastly, we add a special "firm" node $j = 0$ that denotes unemployment and a corresponding firm type $\ell = 0$.

Worker nodes. We divide I workers indexed by $i \in \{1, ..., I\}$ into K groups $k \in \{1, ..., K\}$. Workers initially draw a type $Z_i^w(1) = k$, independently, with probability π_k^w. We also define indicator variables $Z_{ik}^w(1)$, with $\sum_{k=1}^K Z_{ik}^w(1) = 1$ and $Z_{ik}^w(1) = 1$ if $Z_i^w(1) = k$.

Initial matching. The initial block-structure of the network depends on the probabilities $m(k, \ell) := m_{k\ell} \geq 0$, with $\sum_{\ell=0}^L m_{k\ell} = 1$. The probabilities $m_{k\ell}$ are called the initial connectivity parameters. In addition, a parameter $\theta_{j\ell}$, that we treat like a firm fixed effect, controls the expected degree (size) of firm $j \neq 0$. Specifically, $\theta_{j\ell}$ is the probability that having drawn a firm group ℓ, we draw firm j from all the firms in this group. Hence, we impose the normalization: $\sum_{j=1}^J \theta_{j\ell} Z_{j\ell}^f = 1$ (via a softmax function). We also normalize $\theta_{00} = 1$. This is a generalization of the uniform sampling of firms given type that we assumed in LPR.

Let $Y_i(1) = j$ and $Y_{ij}(1) = 1$ if worker i matches with firm j initially; $Y_{ij}(1) = 0$ otherwise. Moreover, $Y_i(1) = 0$ or $Y_{i0}(1) = 1$ indicates unemployment. We have $\sum_{j=0}^J Y_{ij}(1) = 1$, as employment states are mutually exclusive. We suppose that $Y_i(1) = j$ with probability $m_{k\ell}\theta_{j\ell}$ given $Z_i^w = k$ and $Z_j^f = \ell$.[4]

2.3 Dynamics

The network of worker-firm matches is exposed to two separate dynamics experienced by workers: employment transitions and worker type changes. We assume that further realizations of worker wages and employment transitions are subject to the following Markovian conditional independence assumption between wages, employment transitions, and type changes: for $s = 1, ..., S_i - 1$,

$$W_i(s) \perp\!\!\!\perp (D_i(s), Y_i(s+1)) \perp\!\!\!\perp Z_i^w(s+1) \mid Z^f, Z_i^w(s), Y_i(s),$$

where past observations beyond $Z^f, Z_i^w(s), Y_i(s)$ are irrelevant.

Matching dynamics. Employment transitions occur between employment states (to and from unemployment, or employer change) with probability $M(\ell'|k, \ell) := M_{k\ell\ell'} \geq 0$

[4]In a standard SBM, the typical assumption is that (i, j) are drawn independently from a Bernoulli distribution (or Poisson) with parameter $m_{k\ell}$ given $Z_i^w = k$ and $Z_j^f = \ell$. So there is no normalization $\sum_{\ell=0}^L m_{k\ell} = 1$. If there are many firms, the dependence in the first case is negligible. It is also customary to replace the Bernoulli by a Poisson distribution to simplify mathematical expression.

per month. Note that $M_{k\ell\ell}$ is the probability of changing employer within the same group $\ell \geq 1$. We set $M_{k00} = 0$ as there are no transitions from unemployment to unemployment. Unemployment may however persist for more than two consecutive months because no job offer has been received or accepted. Let $M(\neg|k,\ell) := M_{k\ell\neg} = 1 - \sum_{\ell'=0}^{L} M_{k\ell\ell'}$ denote the probability of staying one additional month in the same employment state. Specific employers are drawn within firm groups as in the initial spell with probabilities $\theta_{j\ell}$. We denote as $D_i(s)$ the length of the sth spell in months. All durations $D_i(s)$ are censored by the number of months left until the end of the current year. Suppose that the period starts in month 5. Then $D_i(s) \leq 12 - 4 = 8$ and the minimum duration is one month. For all $1 \leq d \leq 12$, the joint probability of the recorded duration and new destination (d, j', ℓ') given the origin (k, j, ℓ) is $(M_{k\ell\neg})^{d-1} M_{k\ell\ell'}\theta_{j'\ell'}$. The probability of censoring — $D_i(s) = 12$ — is $(M_{k\ell\neg})^{12}$. These probabilities do not depend on the current employer ID j beyond its type ℓ.

The renewal process of worker types. The type of worker i may change with the period index s. Let $Z_i^w(s)$ denote the sequence of worker i's types. The probability that $Z_i^w(s) = k'$ given $Z_i^w(s-1) = k$, $Y_i(s-1) = j$ and $Z_j^f = \ell$ is only function of k, ℓ, k' and is denoted $A(k'|k,\ell) := A_{k\ell k'}$. We allow types to change during unemployment ($j = \ell = 0$). Various constrained specifications are possible. For example, it could be that $A_{k\ell k'}$ only depends on whether $\ell = 0$ (unemployment) or $\ell \neq 0$ (employment). We could also assume that only incremental type changes are possible: $A_{k\ell k'} = 0$ if $|k' - k| > 1$.

Wages. As stated above, a unique wage is observed for each separate period of employment within a year. We keep the main job in case of several simultaneous jobs. We assume that the wage in spell s is Gaussian:

$$W_i(s) \mid Z_i^w(s) = k, Z_j^f = \ell \sim \mathcal{N}\left(\mu_{k\ell}, \sigma_{k\ell}^2\right).$$

We could have a different equation for within-job wages and hiring wages, add autoregressive effects, etc. We prefer to focus on the main innovation of this paper: worker type dynamics. Also, errors need not be Gaussian. For example, Brault, Keribin, and Mariadassou (2020) consider the case of the regular univariate exponential family. The normal is bivariate exponential, which also greatly simplifies the maximum likelihood calculations. We can write the density of observation $W_i(s) = w$ conditional on $Z_i^w(s) = k, Z_j^f = \ell$ as $f(w|k,\ell) = \varphi(T(w), \eta_{k\ell})$ with

$$\varphi(T, \eta) = h(w) \exp\left(\eta^\top T - \psi(\eta)\right),$$

where $T(w) = (w, w^2)^\top$ is the sufficient statistic, $\eta = (\eta_1, \eta_2)^\top = \left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)^\top$ is the natural parameter and $\psi(\eta) = -\frac{\eta_1^2}{4\eta_2} - \frac{1}{2}\ln(-2\eta_2)$. Using ψ we can derive moments of the sufficient statistics by differentiation:

$$\mathbb{E}T = \frac{\partial \psi(\eta)}{\partial \eta}, \quad \mathbb{V}T = \frac{\partial^2 \psi(\eta)}{\partial \eta \partial \eta^\top}.$$

3 Identification of parameters given firm types

In this section we show identification using at least three workers per firm and at least three consecutive periods of observation per worker. We present a simple proof of identification for the discrete time case with no control variables and where the timings of wage observations and employment transition coincide. We also suppose to simplify that the wage distributions are discrete. The identifiability of the parameters of SBM were first obtained by Allman, Matias, and Rhodes (2009). Our model is a lot more complex because of the joint dynamic of (worker) types and states (network connections). Our identification argument relies heavily on the wage observations as well as on the specific bipartite structure with fixed firm types. If graph communities existed beyond the clusters identified from wage observations, we do not guarantee that the model would remain identified. The recent paper by Longepierre and Matias (2019) allows for type dynamics but rules out state-dependence in network connections. Online Appendix A details the argument.

3.1 Identification of firm types

Suppose that for each firm j we observe three wages. Then, assuming independent wage draws given firm type, we can identify the proportion of each firm type, and the distributions of observed firm characteristics and wages given firm types in each year, using standard arguments for the identification of random mixtures. This was the argument in BLM and it still applies in the present context. So let us now assume that we know the type of each firm in every year. We assume wage discrete for simplicity.

3.2 Identifying matrices

We then construct a collection of identifying matrices of probabilities. Fix one firm type $\ell = 1, ..., L$, and let us stack the model parameters into the following matrices,

$$\Pi_\ell = \text{diag}[m(k,\ell)]_k, \quad F_\ell = [f(w|k,\ell)]_{w\times k}, \quad A_\ell = [A(k'|k,\ell)]_{k\times k'},$$
$$D_{\ell\neg} = \text{diag}[M(\neg|k,\ell)]_k, \quad D_{\ell\ell'} = \text{diag}[M(\ell'|k,\ell)]_k, \quad D_{\ell 0} = \text{diag}[\delta_{k\ell}]_k,$$
$$A_0 = [A(k'|k,0)]_{k\times k'}, \quad D_{\ell 0} = \text{diag}[M(0|k,\ell)]_k, \quad D_{0\ell'} = \text{diag}[M(\ell'|k,0)]_k.$$

In the identification appendix, we show that the following algebraic structure holds.

Within spell wage dynamics. For the probability $\mathbb{P}(\ell, w, \neg, w')$ of two wages w and w' in periods 1 and 2, we have

$$Q_{\ell\neg} \equiv [\mathbb{P}(\ell, w, \neg, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\neg} A_\ell F_\ell^\top.$$

Between spell wage dynamics. For the probability $\mathbb{P}(\ell, w, \ell', w')$ of wages w and w' in periods 1 and 2 with a job change from ℓ to ℓ', we have

$$Q_{\ell\ell'} \equiv [\mathbb{P}(\ell, w, \ell', w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\ell'} A_\ell F_{\ell'}^\top.$$

Transition to unemployment. For the probability $\mathbb{P}(\ell, w, 0)$ of wage w in period 1 in a firm of type $\ell = 1, ..., L$ followed by a transition to non-employment in period 2, we have

$$Q_{\ell 0} = [\mathbb{P}(\ell, w, 0)]_w = F_\ell \Pi_\ell D_{\ell 0} e_K,$$

where e_K is the K-vector of ones.

Wage dynamics with an unemployment interruption. For the probability $\mathbb{P}(\ell, w, 0, \ell', w')$ of wages w and w' and employer types ℓ and ℓ' in periods 1 and 3 with an intermediate non-employment spell in period 2, we have

$$Q_{\ell 0\ell'} = [\mathbb{P}(\ell, w, 0, \ell', w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell 0} A_\ell D_{0\ell'} A_0 F_{\ell'}^\top.$$

3.3 Assumptions and main identification result

We make the following identifying restrictions.

Assumption 1. *The following conditions hold.*

1. *For all* $(k, \ell) \in \{1, ..., K\} \times \{1, ..., L\}$*, all matches are initially possible, i.e.* $m(k, \ell) \neq 0$*, making the matrices* $\Pi_\ell = \mathrm{diag}[m(k, \ell)]_k$ *non singular for all* $\ell \in \{1, ..., L\}$.

2. *The matrices* $F_\ell = [f(w|k, \ell)]_{w\times k}$ *have rank* K *for all* $\ell \in \{1, ..., L\}$.

3. *The matrices* $A_\ell = [A(k'|k, \ell)]_{k\times k'}$ *are square-invertible for all* $\ell = 1, ..., L$.

4. *For all* $(k, \ell) \in \{1, ..., K\}\times\{1, ..., L\}$*, no mobility is always possible, i.e.* $M(\neg|k, \ell) \neq 0$*, making the matrices* $D_{\ell\neg} = \mathrm{diag}[M(\neg|k, \ell)]_k$ *non singular for all* $\ell \in \{1, ..., L\}$.

5. *For all* (k, k') *there exists* $\ell \in \{1, ..., L\}$ *such that* $\frac{M(\ell|k,\ell)}{M(\neg|k,\ell)} \neq \frac{M(\ell|k',\ell)}{M(\neg|k',\ell)}$*, which means that the diagonal matrices* $D_{\ell\neg}^{-1}D_{\ell\ell} = \mathrm{diag}[M(\ell|k, \ell)/M(\neg|k, \ell)]$ *have distinct diagonal entries.*

6. *Positive probability of not changing type in unemployment:* $A(k|k, 0) > 0$.

Conditions 1-4 are rank restrictions implying that the matrices $Q_{\ell\neg}$, $\ell \in \{1, ..., L\}$, have rank K. Hence, the number of worker types is identified as the rank of an observable matrix. This is because we are making more assumptions than Kasahara and Shimotsu (2014) who only identify a lower bound of the number of groups. In particular, we restrict the conditional wage distributions $(f(w|k, \ell))_k$ to be linearly independent across worker types for all firm types and the wage supports is large enough for the row rank to be greater that K.

Condition 5 is not a rank restriction. It ensures that certain matrices have all their eigenvalues simple, making eigenspaces unidimensional. Condition 6 allows to recover the parameters corresponding to unemployment spells once all other parameters have been recovered. Conditions 1-6 are far from being minimal, but they facilitate the proof of identification. And drawing all parameters from a continuous distribution unrestrictedly, they are satisfied with probability one. In other words, given K and L, they are generic.

The following proposition shows that it suffices to observe worker wages and employer types over three consecutive periods to identify all parameters.

Theorem 2. *The model is identified under Assumption 1 up to a group label permutation given matrices* $Q_{\ell\neg}, Q_{\ell\ell'}, Q_{\ell 0}, Q_{\ell 0\ell'}$*, for* $\ell, \ell' = 1, ..., L$.

The proof is in Appendix 1. It is constructive. It requires two consecutive employments spells with and without intermediate period of unemployment. The latter case is the only reason why we need three consecutive spells. The main idea of the proof, which is specific to our setup and greatly simplifies the argument, is that we can learn a lot on the parameters by contrasting wage changes within firms and between firms of the same type.

We conclude this section by a remark on model symmetry, a notion introduced and discussed by Brault, Keribin, and Mariadassou (2020). Model symmetry occurs when switching group labels does not change parameters. This happens for example in a SBM when groups have equal probabilities and the connection probabilities manifest certain symmetries. With continuous wage distributions, model symmetry is highly unlikely to occur, and it is ruled out by Assumption 1, Condition 2.

4 Estimation method

4.1 Complete information likelihood

Let $X = (Y, W, D)$ denote the observation sample. The latent worker and firm types are gathered in $Z = (Z^f, Z^w)$. The log-likelihood of the complete data $\left(X, Z^f, Z^w\right)$ is

$$\ln \mathcal{L}\left(X, Z^f, Z^w\right) = \ln \mathcal{L}\left(Z^f\right) + \ln \mathcal{L}\left(X, Z^w \mid Z^f\right),$$

with

$$\ln \mathcal{L}\left(Z^f\right) = \sum_{j=1}^{J} \sum_{\ell=1}^{L} Z^f_{j\ell} \ln \pi^f_\ell, \quad \ln \mathcal{L}\left(X, Z^w \mid Z^f\right) = \sum_{i=1}^{I} \ln \mathcal{L}_i\left(X_i, Z^w_i \mid Z^f\right).$$

To calculate $\mathcal{L}_i\left(X_i, Z^w_i \mid Z^f\right)$, let us split the whole sequence of individual observations (or emissions) as

$$\begin{aligned} X_i(1) &= (Y_i(1), W_i(1), D_i(1), Y_i(2)), \\ X_i(s) &= (W_i(s), D_i(s), Y_i(s+1)), \ s = 2, ..., S_i - 1, \\ X_i(S_i) &= (W_i(S_i), D_i(S_i)). \end{aligned}$$

Given firm types Z^f, the probability that the worker is initially of type k is

$$\alpha_k(1) = \mathbb{P}\left(Z^w_i(1) = k\right) = \pi^w_k.$$

The probability of the first emission $X_i(1)$ given that $Z_i^w(1) = k$ is

$$\beta_k(1|Z^f) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(1) \sum_{\ell=0}^{L} Z_{j\ell}^f \Bigg(\ln m_{k\ell} + \ln\theta_{j\ell} + \mathbf{1}\{j \neq 0\} \ln\varphi\left(T\left[W_i(1)\right], \eta_{k\ell}\right)$$
$$+ D_i(1)\ln M_{k\ell\neg} + \sum_{j'=0}^{J} Y_{ij'}(2) \sum_{\ell'=0}^{L} Z_{j'\ell'}^f \mathbf{1}\{j' \neq j\}\left[\ln M_{k\ell\ell'} + \ln\theta_{j'\ell'}\right]\Bigg)\Bigg].$$

Then, for all following spells $s = 2, ..., S_i$, the probability that $Z_i^w(s) = k'$ given Z^f and $Z_i^w(s-1) = k$, irrespective of the whole past, is

$$\alpha_{kk'}(s|Z^f) = \exp\left[\sum_{j=1}^{J} Y_{ij}(s-1) \sum_{\ell=0}^{L} Z_{j\ell}^f \ln A_{k\ell k'}\right].$$

The likelihood of emissions $X_i(s)$ is, for $s = 2, ..., S_i - 1$,

$$\beta_k(s|Z^f) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(s) \sum_{\ell=1}^{L} Z_{j\ell}^f \Bigg(\mathbf{1}\{j \neq 0\} \ln\varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right)$$
$$+ D_i(s)\ln M_{k\ell\neg} + \sum_{j'=0}^{J} Y_{ij'}(s+1) \sum_{\ell'=0}^{L} Z_{j'\ell'}^f \mathbf{1}\{j' \neq j\}\left[\ln M_{k\ell\ell'} + \ln\theta_{j'\ell'}\right]\Bigg)\Bigg],$$

and for the last spell,

$$\beta_k(S_i|Z^f) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(S_i) \sum_{\ell=1}^{L} Z_{j\ell}^f \Bigg(\mathbf{1}\{j \neq 0\} \ln\varphi\left(T\left[W_i(S_i)\right], \eta_{k\ell}\right) + D_i(s)\ln M_{k\ell\neg}\Bigg)\Bigg].$$

We finally have,

$$\ln\mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right) = \sum_{k=1}^{K} Z_{ik}^w(1) \ln\left[\alpha_k(1)\beta_k(1|Z^f)\right]$$
$$+ \sum_{s=2}^{S_i}\sum_{k=1}^{K} Z_{ik}^w(s-1) \sum_{k'=1}^{K} Z_{ik'}^w(s) \ln\alpha_{kk'}(s|Z^f) + \sum_{s=2}^{S_i}\sum_{k=1}^{K} Z_{ik}^w(s) \ln\beta_k(s|Z^f)$$
$$= \ln\left[\alpha_{k_1}(1)\beta_{k_1}(1|Z^f)\alpha_{k_1k_2}(2|Z^f)\beta_{k_2}(2|Z^f)\alpha_{k_2k_3}(3|Z^f)\beta_{k_3}(3|Z^f)...\right],$$

for $Z_i^w = (k_1, k_2, k_3, ...)$.

Let b denote the parameter of a parametric specification of $\mathcal{L}(X, Z)$. It comprises the parameters of the likelihood of the latent variables: group shares π_k^w, π_ℓ^f and the worker type transition probabilities $A_{k\ell k'}$; and the parameters of the emission probabilities: the initial match probabilities $m_{k\ell}$, the match transition probabilities $M_{k\ell\ell'}$, the firm sampling probabilities $\theta_{j\ell}$, the wage parameters $\eta_{k\ell}$. We derive the Maximum Likelihood Estimator of the complete information model, maximizing $\mathcal{L}(X, Z; b)$, in Online Appendix B.

4.2 Variational EM

The complete information MLE is a useful benchmark, but is unfortunately not feasible because the types $Z = (Z^w, Z^f)$ are unobserved. Let us gather observed emissions in the array $X_i = (W_i, Y_i, D_i)$ and let $X = (X_i)$. Averaging $\mathcal{L}(X, Z)$ over Z to calculate the likelihood of the observed data $\mathcal{L}(X)$ is infeasible because of the large dimension of the vector of firm types Z^f. The EM algorithm comes to mind. Unfortunately, EM requires the computation of the conditional probability $\mathcal{L}(Z \mid X)$ which is also intractable.[5] We can calculate $\mathcal{L}(Z^w \mid X, Z^f)$ because worker trajectories are independent conditional on firm types, but not $\mathcal{L}(Z^f \mid X, Z^w)$ because the network of firms connected by common employees is essentially one big connected component.

Bonhomme, Lamadon, and Manresa (2019) (or BLM) solve this difficulty by pre-clustering firms using a k-means algorithm. Then, they estimate the latent worker model using a standard EM algorithm. Lentz, Piyapromdee, and Robin (2023) (or LPR) follow BLM and produce a hard classification of firms and a random clustering of workers. However, they imbed the firm classification inside the M-stage of the EM algorithm for workers, using the Classification EM algorithm of Celeux and Govaert (1992).

The problem with a hard classification of firms is that the estimation of the model can be used for simulation with only the firms used for estimation. The model remains incomplete as it does not say how firm types are drawn. In addition, some firms will be clustered accurately because they have many employees, but there are also many small firms with little information to group them accurately. Abowd, McKinney, and Schmutte (2019) develop a model of worker-firm matching very similar to the one above, using a Bayesian approach to estimation. Drawing firm types from the conditional distributions with the Gibbs sampler is a formidable computing task that Abowd et al. speed up a bit by making use of the possibility that firms may belong to disconnected sub-networks, which allows for some degree of parallelization. But this parallelization is not massive enough — the firm network is too connected — to make a big difference. In the end, despite having access to significant computing capacity, they reduce the

[5]To simplify notations, we use the same notations for the likelihood (function of parameters) and the density function or probability mass of observed and latent variables.

estimation sample to a random subsample of about 400,000 person-year observations with 80,000 workers, 60,000 firms and 130,000 matches. Although a very active research exists and has been successful at increasing the efficiency of the moves in MCMC algorithms (see Lee and Wilkinson, 2019), an alternative to Monte Carlo methods is the class of variational expectation-maximization (VEM) methods, that aim at maximizing a lower bound of the observed log-likelihood $\ln \mathcal{L}(X)$ (see Jaakkola (2001), Govaert and Nadif, 2008, Daudin, Picard, and Robin, 2008) which we now describe.

For any probability distribution $R(Z)$ (summing to one), define the so-called ELBO (Evidence Lower Bound) as

$$\mathcal{J}(R,X) = \ln \mathcal{L}(X) - D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right)$$

where $D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right)$ is the Kullback-Leibler divergence

$$D_{KL}\left(R \parallel \mathcal{L}(Z \mid X)\right) = \sum_Z R\left(Z\right) \ln\left(\frac{R(Z)}{\mathcal{L}(Z \mid X)}\right) \geq 0.$$

Hence, the ELBO is a lower bound of the log-likelihood that is attained when $R(Z) = \mathcal{L}(Z \mid X)$.

Moreover, as $\ln \mathcal{L}(X) = \sum_Z R(Z) \ln \mathcal{L}(X)$,

$$\begin{aligned}
\mathcal{J}(R,X) &= \sum_Z R(Z) \ln \mathcal{L}(X) - \sum_Z R\left(Z\right) \ln\left(\frac{R(Z)}{\mathcal{L}(Z \mid X)}\right) \\
&= \sum_Z R(Z) \ln \mathcal{L}(X,Z) - \sum_Z R(Z) \ln R(Z) \\
&= \sum_Z R(Z) \ln \mathcal{L}(X \mid Z) - \sum_Z R\left(Z\right) \ln\left(\frac{R(Z)}{\mathcal{L}(Z)}\right),
\end{aligned}$$

where $\mathcal{H}(R) = -\sum_Z R(Z) \ln R(Z) = -\mathbb{E}_R \ln R(Z)$ is the entropy of distribution R. Hence, a good distribution $R(Z)$ is one that maximizes the expected complete log-likelihood while keeping the entropy $\mathcal{H}(R)$ as large as possible (no small groups). The last equality gives another interesting interpretation. The ELBO is the expected conditional log-likelihood $\ln \mathcal{L}(X \mid Z)$ with respect to the guessed probability distribution $R(Z)$ of the latent variable Z, penalized by the KL divergence of $R(Z)$ from the true prior $\mathcal{L}(Z)$.

Variational EM is similar to standard EM. Given $\mathcal{L}(Z \mid X)$ and data X, a pseudo

E-step optimizes on $R(Z)$:

$$\hat{R} = \arg\max_{R\in\mathcal{R}} \mathcal{J}(R, X) = \arg\max_{R\in\mathcal{R}} \sum_Z R(Z) \ln \mathcal{L}(X \mid Z) - \sum_Z R(Z) \ln \frac{R(Z)}{\mathcal{L}(Z)},$$

where $\mathcal{R}$ is a class of distributions making this optimization problem tractable. If $\mathcal{L}(Z \mid X)$ is computable, then we do not need to restrict the feasibility set $\mathcal{R}$, and we get $\hat{R}(Z) = \mathcal{L}(Z \mid X)$ as in standard EM. Otherwise, the optimal solution can only be approximated within a class of manageable distributions. VEM is then looking for a feasible posterior update $R(Z)$ that maximizes the expected conditional log-likelihood penalized for any strong departure from the prior distribution $\mathcal{L}(Z)$. Notice also that, as long as $\mathcal{R}$ contains the Dirac mass δ_{Z^*},

$$\ln \mathcal{L}(X, Z^*) = \sum_z \delta_{Z^*}(z) \ln \mathcal{L}(X, z) = \mathcal{J}(\delta_{Z^*}, X) \leq \max_{R\in\mathcal{R}} \mathcal{J}(R, X) \leq \ln \mathcal{L}(X). \quad (1)$$

The Variational Estimator of b is obtained in the pseudo M-step:

$$\hat{b} = \arg\max_b \max_{R\in\mathcal{R}} \mathcal{J}(R, X; b) = \arg\max_b \sum_Z \hat{R}(Z) \ln \mathcal{L}(X, Z; b).$$

Consider for example the case of a finite mixture model:

$$\mathcal{L}(X, Z; b) = \prod_{i=1}^{I} \sum_{k=1}^{K} Z_{ik} \pi_k f(X_i; \beta_k),$$

where $Z_{ik} = 1$ if i is in group k. We could restrict $R(Z)$ to be a hard classification of observations: $R(Z) = \delta_\tau(Z) = \prod_{i=1}^{I} \prod_{k=1}^{K} \chi_{ik}^{Z_{ik}}$, where $\chi = (\chi_i)$ and $\chi_{ik} = 1$ if $\chi_i = k$. Then, using the convention $0 \ln(0) = 0$ and since $1 \ln(1) = 0$,

$$\sum_Z R(Z) \ln \mathcal{L}(X, Z) - \sum_Z R(Z) \ln R(Z) = \sum_{i=1}^{I} \sum_{k=1}^{K} \chi_{ik} \pi_k f(X_i; \beta_k).$$

Thus, given $b = (\pi_k, \beta_k)$, the pseudo E-step selects for each i the k that maximizes $\pi_k f(X_i; \beta_k)$. This is the Classification EM algorithm of Celeux and Govaert (1992) that was used in Lentz, Piyapromdee, and Robin (2023).

4.3 VEM for our model

We first approximate $\mathcal{L}(Z \mid X)$ as $R(Z) = R^w(Z^w)R^f(Z^f)$ with forced independence between worker and firm types. The integrated or expected log-likelihood is

$$\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w) = \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^f, Z^w).$$

The M-step maximizes $\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w; b)$ with respect to b given current R^f, R^w. The E-step is decomposed into a sequence of two steps. First, we maximize $\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w; b)$ with respect to R^w given the updated b and current R^f, with no restriction on R^w. In Appendix C, we show that the optimal value of R^w is of the form $R^w(Z^w) = \Pi_{i=1}^I R_i^w(Z_i^w)$, with independence between worker types. We also calculate the marginal distributions:

$$\mathbb{P}_{R^w}\left\{Z_i^w(s) = k\right\} = \xi_{ik}(s), \quad \mathbb{P}_{R^w}\left\{Z_i^w(s) = k, Z_i^w(s+1) = k'\right\} = \xi_{ikk'}(s+1),$$

which are sufficient to calculate the expected log-likelihood. In practice, we use the Forward-Backward algorithm for workers, which is standard EM and not Variational EM, and use an approximation of the posterior distribution only for firms.[6]

Then we maximize the expected log-likelihood with respect to R^f given updated b and R^w subject to the following mean-field restriction:

$$R^f(Z^f) = \Pi_{j=1}^J R_j^f(Z_j^f) = \Pi_{j=1}^J \chi_{j\ell}^{Z_{j\ell}^f}.$$

This amounts to an assumption of posterior firm type independence. On examining the complete log-likelihood, we see that it linearly depends on the indicators $Z_{j\ell}^f$, except for the contributions of employment transitions where we have $Z_{j\ell}^f Z_{j'\ell'}^f$ for a transition of (j, ℓ) to (j', ℓ'). The optimal R^f requires estimating the marginal probability of $Z_{j\ell}^f$, say $\chi_{j\ell}$, and the joint probabilities of $Z_j^f, Z_{j'}^f$. Under independence between firms, the optimization problem becomes manageable.

All this was done assuming a value for the numbers of latent groups K and L. There is no consensual method to estimate the numbers of groups when the true data likelihood is not easily computed. Daudin, Picard, and Robin (2008) and Matias and Miele (2017) suggest to use the Integrated Classification Likelihood (ICL) of Biernacki,

[6] Matias and Miele (2017) develop a Variational Estimator for a dynamic SBM where node types change according to a Markov chain as in our model, and Longepierre and Matias (2019) show consistency and derive convergence rates. They observe that the usual Forward-Backward (FB) algorithm for Hidden Markov Models is not feasible here because the links (and joint match outcomes) depend on the types of both nodes for each match (p. 1130). However, with a bipartite network, it is possible to deal with row nodes conditional on column nodes, which restores the usefulness of the FB algorithm. See Zhao, Hao, and Zhu (2024) for a similar point.

Celeux, and Govaert (2000). There is, however, no known result on the convergence of the ICL procedure.

In Appendix C we describe in detail the implementation of the VEM algorithm that serves to determine the model parameter estimates as well as the worker and firm classifications that maximize the likelihood of the data. As is typical for EM algorithms, we perform many restarts with different initial guesses to ensure that the parameter estimate represents a global maximum. We have found it helpful to make new initial parameter guesses to be perturbations around the current best estimate to speed up the search for improvements. The estimate we are showing is a result of 500 restarts.

4.4 Consistency of the ML and Variational estimators

Celeux and Govaert (1992) consider the particular case of VEM for standard finite mixtures and show numerical convergence and consistency for certain initial conditions. Recent work derives conditions proving the identification and the consistency of VEM for stochastic block models (SBMs) and latent block models (LBMs) depending on the network density. Celisse, Daudin, and Pierre (2012) shows consistency of Maximum Likelihood and Variational Estimators for the Stochastic Block Model (SBM) model. Longepierre and Matias (2019) consider the case of a Dynamic SBM where the latent types of the network's nodes change like the worker types in our paper. They prove consistency of the connectivity parameters (the probabilities of a link given node types), but to prove the consistency of the type probabilities (static or dynamic) they assume that the connectivity parameters converge at a rate at least equal to n^{-1}, where n is the number of nodes. This is a strong assumption that Bickel, Choi, Chang, and Zhang (2013) (for SBMs) and Brault, Keribin, and Mariadassou (2020) (for LBMs) were able to circumvent by developing a Bernstein-type concentration inequality for sub-exponential variables. Brault et al. show consistency and asymptotic normality of all estimators by showing that both the observed MLE and the VEM are asymptotically equivalent to the complete information MLE. In other words, the double inequality (1) becomes an equality as the number of nodes tends to infinity. Zhao, Hao, and Zhu (2024) study the case of a static degree-corrected bipartite SBM. They show consistency by showing that the consistency of the clustering (R converges to δ_{Z^*}, where Z^* denotes the true clustering).

In this perspective, one can view the VEM as a particular way of searching over hard classifications — as in the CEM estimation in Lentz, Piyapromdee, and Robin (2023) — that is aided by allowing "soft" classifications in the interim iterations, and the benefit that comes from the associated search over continuous parameters.

4.5 Moment simulation

Finally, we explain how we generate worker trajectories using the model.

In every month, to predict the worker's state $(k', \ell') \in \{1, ..., K\} \times \{0, 1, ..., L\}$ in the next month, we need to know the current state (k, ℓ), whether a job change has occurred ($J \in \{0, 1\}$), and whether the current month is the last of the year or not ($D \in \{0, 1\}$). This latter distinction is made because of the assumption that worker types change either upon a job-to-job transition or at the end of the year. The model allows us to calculate the monthly transition probability $P_D(J, k', \ell' \mid k, \ell)$ as

$$P_0(J = 0, k', \ell' \mid k, \ell) = M_{k\ell\neg}\delta_{\ell\ell'}\delta_{kk'},\ P_1(J = 0, k', \ell' \mid k, \ell) = M_{k\ell\neg}\delta_{\ell\ell'}A(k' \mid k, \ell),$$
$$P_0(J = 1, k', \ell' \mid k, \ell) = M_{k\ell\ell'}A(k' \mid k, \ell),\ P_1(J = 1, k', \ell' \mid k, \ell) = M_{k\ell\ell'}A(k' \mid k, \ell),$$

where $\delta_{\ell\ell'}$ is the Kronecker delta. Let $P_{DJ} := [P_D(J, k', \ell' \mid k, \ell)]$ be the matrix collecting these probabilities for a given D and a given J, with (k, ℓ) in rows and (k', ℓ') in columns. Note that P_{00} is diagonal.

Because earnings are measured one single time in a given year for each employment record, it makes sense to consider employment and wage dynamics at yearly frequency. The one-year forward Markov transition matrix disregarding job changes (ie with one job change or without in any month) is obtained as

$$Q = [Q(k', \ell' \mid k, \ell)] = (P_{00} + P_{01})^{11} (P_{10} + P_{11}).$$

$Q(k', \ell' \mid k, \ell)$ is the probability of being in state (k', ℓ') in December of year $\tau + 1$ given that the worker is in state (k, ℓ) in year τ.

The one-year forward Markov transition matrix with no job change in the year (say $JJ = 0$) and with at least one job change (say $JJ = 1$) are, respectively,

$$Q_0 = P_{00}^{11}P_{10} = \left[M_{k\ell\neg}^{12}A(k' \mid k, \ell)\delta_{\ell\ell'}\right] \text{ and } Q_1 = Q - Q_0.$$

Because we will want to condition on the job change event, we also define the normalized matrices as

$$\overline{Q}_0 = \frac{Q_0}{\Pr(JJ = 0 \mid k, \ell)}, \quad \overline{Q}_1 = \frac{Q_1}{\Pr(JJ = 1 \mid k, \ell)},$$

where

$$\Pr(JJ = 0 \mid k, \ell) = \sum_{k',\ell'} Q_0(k', \ell' \mid k, \ell) = M^{12}_{k\ell\neg},$$
$$\Pr(JJ = 1 \mid k, \ell) = 1 - \Pr(JJ = 0 \mid k, \ell) = 1 - M^{12}_{k\ell\neg}.$$

A cohort's initial distribution in year of entry $t_e(\zeta)$ is $p_{0,\zeta}(k, \ell)$ over states (k, ℓ) given by,

$$p_{0,\zeta}(k, \ell) = \pi^w_k(\zeta)\, m_{k\ell}(\zeta).$$

Then, we calculate at yearly frequency the distribution of match type $(k_{\tau_s}, \ell_{\tau_s})$ in experience-in-panel year $\tau_s > 0$ as

$$p_{\tau_s,\zeta} = p_{0,\zeta} Q^{\tau_s},$$

with $p_{\tau_s,\zeta} = [p_{\tau_s,\zeta}(k, \ell)]$ specified as row-vectors. The probabilities $p_{\tau_s,\zeta}(k, \ell)$ allow us to calculate wage moments at any experience level τ_s.

Finally, using $\tau_s(t, \zeta) = t - t_e(\zeta)$, we deduce $p_t(k, \ell) = \sum_\zeta \pi_t(\zeta) p_{\tau_s(t,\zeta),\zeta}(k, \ell)$, where $\pi_t(\zeta)$ is the cohort distribution in calendar year $t = 1982, ..., 2001$, and we aggregate over all years to obtain $p(k, \ell) = \frac{1}{T} \sum_{t=1}^{T} p_t(k, \ell)$.

5 Estimation on Italian data

In this section, we describe the data used for the empirical implementation of the model and its estimation. We will also show how the model fits the data and how worker and firm types can be transformed into more easily interpretable worker and firm effects..

5.1 Data

We implement the estimator on data from the Veneto region in Italy for the 20 year period, 1982-2001.[7] The main sampling criterion in the data set is to include full employment worker histories for all Italian workers who at some point during the 1982-2001 period have an employment relationship with a firm in the Veneto region. This means that the data include employment relationships in firms outside the Veneto region, but only if the worker involved happened to have a job with a Veneto firm at some other time during 1982-2001.

The data record a worker's birth year. To facilitate comparison with the existing

[7]We are grateful to the Fondazione Rodolfo Debenedetti at Bocconi University for the data access.

literature, we restrict attention to men aged 25-50 in full-time employment relationships. For these workers, the data record their employment spells at monthly frequency. A employment spell is a match between a worker ID and a firm ID, and we thus observe the start month and year of the match, and the end month and year of the match, for all job transition occurring between 1982-2001. Full-time status is inferred through the employer's reported hours. Based on the annual earnings, we construct the daily wage through the inferred number of days the worker has been employed during the given calendar year. Earnings are inflation-adjusted to the price level of the year 2003. It is important to note that we observe annual earnings separately for each employment relationship a worker may have in a given calendar year. Thus, if a worker switches job in a year, we will have separate daily earnings within that year for the worker in the two different jobs.

Non-employment is not directly observed. We pad an employer's employment spell history with non-employment spells in between employment spells. In addition, we pad worker histories in front and end to have full 25-50 year old histories, only restricted by the left and right censoring of the calendar time window of the data. As an example, if the last employment spell we observe for a given worker ends in June 1999 and the worker is 45 years old at the time, we pad the worker's history with a non-employment spell that runs from July 1999 through December 2001, which is the last month of the data observation window.

Furthermore, we require that the data be strongly connected to allow implementation of the leave-one-out estimator (see Kline, Saggio, and Sølvsten (2020), or KSS), which implies in particular that each firm has at least two employment relationships during the observation window. Finally, to economize on the number of required worker and firm types in the estimation, we discard job spells that have duration 2 months or less. These spells likely have distinct measurement noise to them and the estimation tends to deal with them by dedicating particular types to them.

5.2 Specification

The estimation is performed on the full 20-year panel (1982-2001). We generally refer to calendar year by t. All parameters are constant, independent of calendar year, except the initial worker type distribution $\pi^w_k(\zeta)$ and the initial match distribution $m_{k\ell}(\zeta)$, which depend on the cohort $\zeta \in \{1957, \dots, 2001\}$, defined by the year of the cohort's 25$^{\text{th}}$ birthday. The panel's initial 1982 cross-section is composed of the stacking of the $\zeta \in \{1957, \dots, 1982\}$ cohorts. Each of these cohorts have year-of-entry into the panel, $t_e(\zeta) = 1982$ for $\zeta \in \{1957, \dots, 1982\}$. A cohort ages out of the panel upon the year of its 51st birthday. The $\zeta \in \{1983, \ldots, 2001\}$ cohorts enter the panel according to their cohort year, $t_e(\zeta) = \zeta$ for $\zeta > 1982$. A cohort's age beyond 25 is defined as $\tau_a(t, \zeta) = t - \zeta$ and its experience-in-panel is referred to as $\tau = t - t_e(\zeta)$.

There are 45 cohorts in the panel. To economize on the number of parameters dedicated to $\pi_k^w(\zeta)$ and $m_{k\ell}(\zeta)$ in the estimation, we collect the cohorts into 8 groups, $\overline{\zeta}_z$, $z = 1, \ldots, 8$. The first four groups are collections over the initial 1982 cross-section; $\overline{\zeta}_1 = \{1957, \ldots, 1963\}$, $\overline{\zeta}_2 = \{1964, \ldots, 1969\}$, $\overline{\zeta}_3 = \{1970, \ldots, 1975\}$, and $\overline{\zeta}_4 = \{1976, \ldots, 1981\}$. The remaining four groups are collections over cohorts that enter the panel at age 25; $\overline{\zeta}_5 = \{1982, \ldots, 1986\}$, $\overline{\zeta}_6 = \{1987, \ldots, 1991\}$, $\overline{\zeta}_7 = \{1992, \ldots, 1996\}$, and $\overline{\zeta}_8 = \{1997, \ldots, 2001\}$. We estimate the objects $\hat{\pi}_k^w(z)$ and $\hat{m}_{k\ell}(z)$, $z = 1, \ldots, 8$, and represent the individual $\pi_k^w(\zeta)$ and $m_{k\ell}(\zeta)$ by interpolation between the cohort group mid-points.[8]

The estimation is done with the number of worker and firm types set to $K = 24$ and $L = 6$. These numbers are the results of different trials after setting up on a particular structure for the worker type transition probability matrix. To ease interpretability and identifiability, we restrict the matrix $A(k' \mid k, \ell)$, for each ℓ, to be block-diagonal with tridiagonal blocks. Thus, the worker type space is a hybrid of three permanent groups where, within each group, a worker's type evolves by transitions through neighboring types.[9] Specifically, we impose 3 disconnected groups with 8 types each. That is,

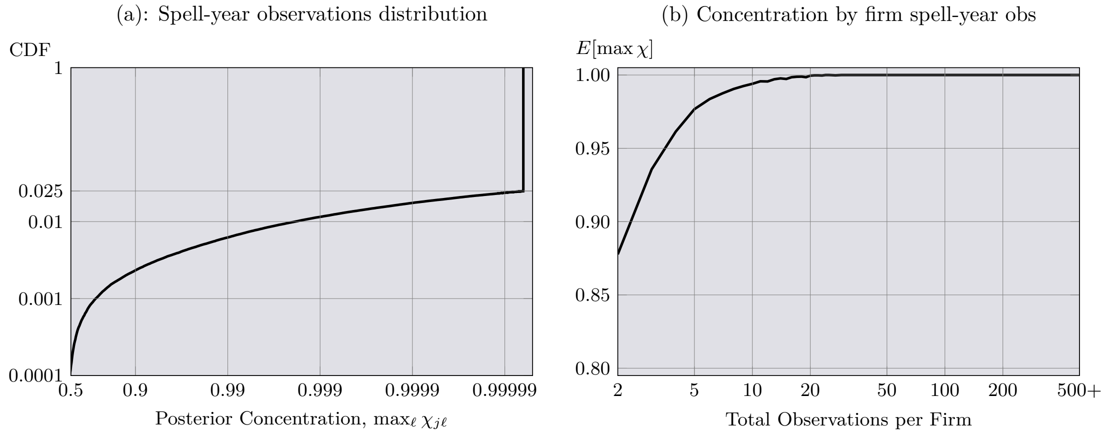
$$A_\ell = \begin{pmatrix} A_\ell^{(1)} & 0 & 0 \\ 0 & A_\ell^{(2)} & 0 \\ 0 & 0 & A_\ell^{(3)} \end{pmatrix}, \quad A_\ell^{(\cdot)} \sim \begin{pmatrix} b_{1\ell} & c_{1\ell} & 0 & 0 & \cdots & 0 \\ a_{2\ell} & b_{2\ell} & c_{2\ell} & 0 & \ddots & \vdots \\ 0 & a_{3\ell} & b_{3\ell} & c_{3\ell} & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{7\ell} & b_{7\ell} & c_{7\ell} \\ 0 & \cdots & 0 & 0 & a_{8\ell} & b_{8\ell} \end{pmatrix},$$

where "$\sim$ " means "similar to". In order to allow potentially large worker type changes associated with even brief non-employment spells, we make an exception to the tri-diagonal restriction for the case of within group type changes in non-employment, where we allow $A_0^{(\cdot)}$ to be fully flexible. We refer to the disconnected worker type groups as blocks indexed by $\kappa = 1, 2, 3$, that partition the worker type space into sets, $\bar{k}_\kappa$ such that $\bar{k}_1 = \{1, \ldots, 8\}$, $\bar{k}_2 = \{9, \ldots, 16\}$, and $\bar{k}_3 = \{17, \ldots, 24\}$.

[8] As an example, $\pi_k^w(\zeta = 1963) = \frac{1963-1960}{1966.5-1960}\hat{\pi}_k^w(z = 1) + \frac{1966.5-1963}{1966.5-1960}\hat{\pi}_k^w(z = 2)$.

[9] We have experimented with other structures. A prominent alternative has been a single large tridiagonal block, where the estimation has responded by generating multiple "local" ladders in the tridiagonal structure in a manner that resembles the separate block types.

Figure 1: How hard is the firm classification?



Notes: Panel (a) shows the CDF of the estimated maximum posterior type probability $\mathbb{E}\left[\max_\ell \chi_{j\ell}\right]$ across firms j. Panel (b) shows $\mathbb{E}\left[\max_\ell \chi_{j\ell}\right]$ by the number of firm observations.

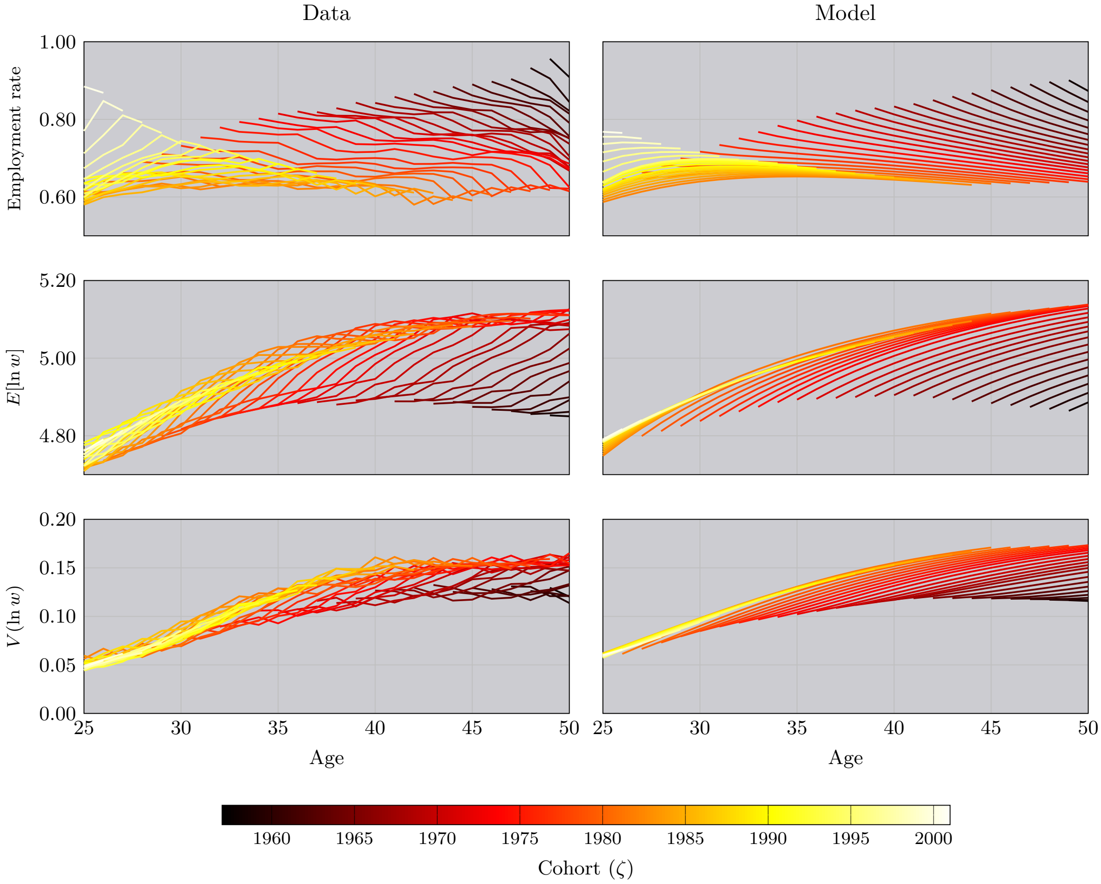
5.3 How hard is the firm classification?

We have adopted a mean-field approximation of $\Pr\left(Z^f \mid X\right)$, the posterior probability of the firm classification. Specifically, we apply the EM algorithm under the restriction of posterior independence across firm types. Consistency of the estimator rests on the argument that as the number of observations per firm increases, the posterior marginal probability of each firm type Z_j^f concentrates on the true type. The practical concern amounts to whether the data contain enough observations per firm for the firm posteriors to be well concentrated. We obtain an average concentration measure of $\mathbb{E}\left[\max_\ell \chi_{j\ell}\right] = .98$, where $\chi_{j\ell}$ is the estimated approximate posterior probability that firm j has type ℓ.

Figure 1, panel (a), shows the fraction of the spell-year observations in the data that are associated with firms of a given posterior concentration measure or less. As can be seen only about 0.01% of the spell-year observations in the data are associated with firms that have a posterior concentration of 0.5 or less. 99% of the data are associated with firms that have concentration measures of 0.999 or greater. Thus, the vast majority of the likelihood is evaluated at fully concentrated firm posteriors.

Figure 1, panel (b), shows how $\mathbb{E}\left[\max_\ell \chi_{j\ell}\right]$ varies with the number of firm-year observations. As can be seen, concentration is increasing in the number of firm observations. For firms with very few observations some uncertainty arises, but even for low firm sizes, average concentration is quite high. Partly, we are aided by the restriction that a firm must have at least two worker matches over its history imposed so as to make our sample directly applicable to the Kline, Saggio, and Sølvsten (2020) leave-one-out estimator. In addition, the six firm types are sufficiently differentiated that even a

Figure 2: Model Fit by Cohort and Age: Employment and wage levels.



Notes: Each line is a different cohort ζ. The first row shows the fraction of employed workers by age, the second row the mean log wage and the third row the variance of log wages (actual and predicted).

handful of firm-year observations on average generate significant concentration. Thus, as a practical matter, the estimation has arrived at a "hard" firm type classification.

5.4 Fit of employment and wage levels by age and cohort

In Figure 2 we show how the model fits cohort wage and employment dynamics in the data. Data are shown in the left column and model estimates are shown in the right column. The data paths are obtained by collecting all workers in the same cohort ζ, and then follow them from their year of entry in the panel and forward until the panel ends or the cohort ages out, whichever comes first. We show the profiles by the age of the cohort. Each graph shows the 45 cohorts in the panel where each cohort is color coded according to the colorbar. The estimated employment rate is $n_{\tau_a,\zeta} = \sum_{k,\ell>0} p_{\tau_s(\tau_a+\zeta,\zeta),\zeta}(k,\ell)$, and the mean and variance of the log-wage at age τ_a

follow from

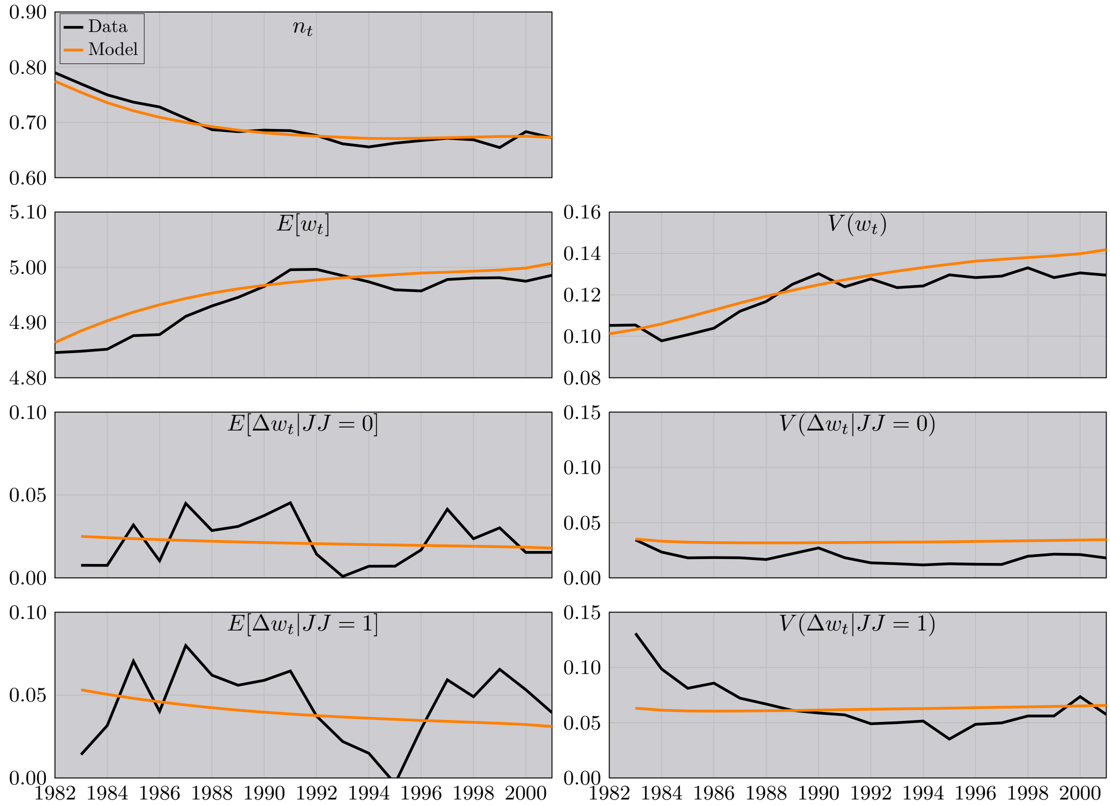
$$\mathbb{E}w_{\tau_a,\zeta} = \frac{1}{n_{\tau_a,\zeta}} \sum_{k,\ell>0} p_{\tau_s(\tau_a+\zeta,\zeta),\zeta}\mu_{k\ell}, \quad \mathbb{E}w^2_{\tau,\zeta} = \frac{1}{n_{\tau_a,\zeta}} \sum_{k,\ell>0} p_{\tau_s(\tau_a+\zeta,\zeta),\zeta}(k,\ell)\left[\sigma^2_{k\ell} + \mu^2_{k\ell}\right].$$

The first row of Figure 2 shows the cohort employment rate by age. Between 25 and 30 years of age, the employment rate tends to be increasing the first few years of a cohort's experience. Past 30 years of age, the employment rate of any cohort is steadily declining with age. We would yet expect that for the young cohort, early non-employment may for some types reflect ongoing education spells. We see substantial cohort effects in the data. Employment rates by age are increasing as we move towards the ends of the cohort spectrum. Sampling selection likely plays some role for these result: The earlier the cohort, the narrower the window to be included in the data set, which requires at minimum one employment spell in Veneto during 1982-2001. Workers from older cohorts that are non-employed during their older years are more likely to be excluded from the sample. A similar effect is present toward the latter cohort birth years as well: For example, an individual from the $\zeta = 2001$ cohort must have had an employment spell in Veneto prior to turning 25 in order to be included in the sample. This type of selection is less of an issue for the cohorts in the $\zeta = 1982$ vicinity, where the inclusion criterion into the data coincides with a 20 year employment window around the ages 25-45. These cohorts show fairly steady employment rates over age. These are selection issues that affect the employment rate profiles, only. All the wage related profiles are employment conditional, which eliminates the selection issue. As can be seen, our model is sufficiently flexible in the initial conditions by cohort that it captures the cohort employment effects and fits the data patterns remarkably well.

The second row shows the cohort average log wage age profiles. Average wages are increasing in age and increase by about 30-35% over a 20 year experience horizon depending on the cohort. Growth is faster at younger ages. The upper envelope of the age profiles coincides with the middle of the cohort spectrum. There are substantial cohort effects in the $\zeta < 1982$ group of cohorts with older cohorts having lower wages at a given age than the cohorts born in later years. Again, we see that the model is able to fit these patterns remarkably well through the cohort specific initial conditions.

The third row shows the cohort log wage variance by age profiles. Within cohort variance is increasing with age, about a tripling over 20 years of experience. The same cohort effects present in the average wage profiles manifest in variance dynamics as well, and once again, the model provides a strong fit.

Figure 3: Model fit by calendar time.

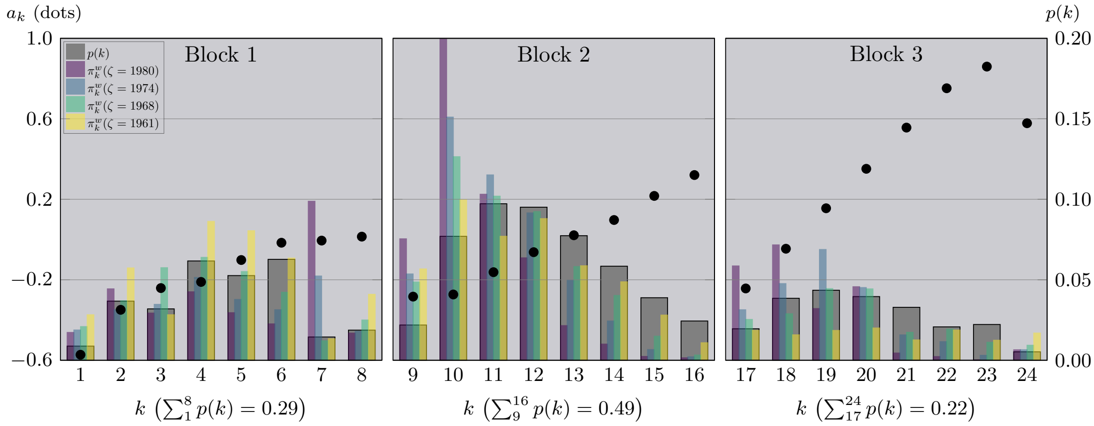


Notes: Actual (black) and predicted (red) statistics. First rows: employment share, second row: mean and variance of log wages, third row: mean and variance of log wage growth for the job stayers, fourth row: mean and variance of log wage growth for the job movers.

5.5 Fit by calendar time

We end this subsection by showing how cohort aggregate by calendar time (Figure 3). The data show a declining employment rate and increasing average wages and wage variances. These trends are particularly strong pre 1990. The model estimate replicates these patterns. In particular, the employment rate is matched very tightly. It is a notable model achievement that it can match the time variation patterns purely by variation in the cohort initial conditions while the rest of the model's parameters are calendar time invariant. The calendar time specific variation in average wage growth is apparent. Growth conditional on a job change is greater than the unconditional growth and there is tendency towards lower growth over time. The model estimates match these patterns without including the data's variation around the trend. The model overestimates the unconditional wage growth variance while matching the overall average conditional on job change.

Figure 4: Worker wage types and initial 1982 cohort entry distribution



Notes: Each plot corresponds to a different fixed worker group, and each k to one of the dynamic types within fixed groups. The wide-bar, gray histogram is the estimated average proportion of each worker type, $p(k)$. The different thin-bar, colored histograms are the initial type probabilities $\pi_k^w(\zeta)$ indexed by the age cohort $\zeta = 1980, 1974, 1968, 1961$ (younger to older). The black circles are the estimated worker fixed effects a_k.

6 Wage projection

Having established that the model fits the data well, we proceed to its interpretation. To ease interpretation and comparison with standard two-way fixed-effect estimations on linked employer-employee data, we use a simple linear projection of the estimated conditional wage means:

$$\mu_{k\ell} = \overline{\mu} + a_k + b_\ell + \tilde{\mu}_{k\ell}. \tag{2}$$

The projection is obtained by OLS, weighing each match type (k, ℓ) by its employment share $p(k, \ell)$, for all $k, \ell > 0$. We will in the following refer to a_k and b_ℓ as the worker and firm wage fixed effects. They are normalized so that $\mathbb{E}a_k = \mathbb{E}b_\ell = \mathbb{E}\tilde{\mu}_{k\ell} = 0$ (where the expectations are relative to $p(k, \ell), \ell > 0$).

6.1 Worker effects

The worker effect is fixed given the worker type, but the worker type changes over time. Firm types remain fixed over time, but job mobility induces a dynamics of firm effects as well. Firm types are labelled according to b_ℓ sorted in ascending order. We do not sort worker types. The diagonal restrictions on the type transition matrix itself encourages an ordering in the estimation as it seeks to explain the wage dynamics in the data.[10]

[10]We have arranged the blocks in order of increasing average a_k by block and we have also flipped order of a block if it has an obvious ladder so as to make it increasing in the type index.

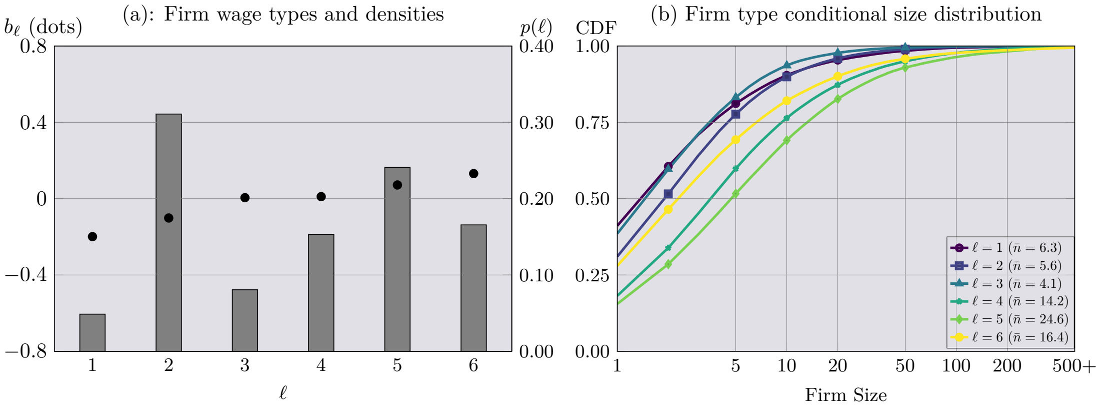
Figure 4 shows the worker effects a_k (black circles) and the marginal distribution of worker types in wide, gray bar. In addition, the figure shows the initial cohort distributions for $\zeta = 1980, 1974, 1968, 1961$, a selection of the 1982 entrants by age. Each panel represents a block. The estimation has responded to the restrictions on the type transition matrix by making neighboring types close to each other in wages. Furthermore, the estimation has constructed different wage advancements ladders across the blocks. The three blocks are naturally ordered by the progressivity of the wage ladder a_k. Starting wage types are fairly similar. The first block, $\bar{k}_1$ has little nominal wage growth. The third block, $\bar{k}_3$, displays the steepest nominal wage growth. Young workers are over-represented in the low indices and older workers are over-represented in the upper indices, indicating entry at the lower end of the ladder and subsequent advancement with experience.

It is important to allow for permanent worker heterogeneity as the data provide little information on worker and firm heterogeneity, such as education or industry. The broad occupation of the worker is recorded though. Block 1 is mostly comprised of blue-collar workers (88.9%), Block 2 is 65% blue-collar and 34% white-collar, while Block 3 is mixed with 21% blue-collar, 66% white-collar and 13% managers. Blue-collar workers spread themselves evenly between Blocks 1 (52%) and 2 (15%), white-collars are mixed with a majority (52%) in Block 2 and 33% in Block 3, and managers are nearly all in Block 3 (87%). It is interesting that the block structure is not exactly superimposed on the occupational structure. This means that the wage and mobility data reveal communities that are not directly determined by administrative occupational categories.

6.2 Firm effects

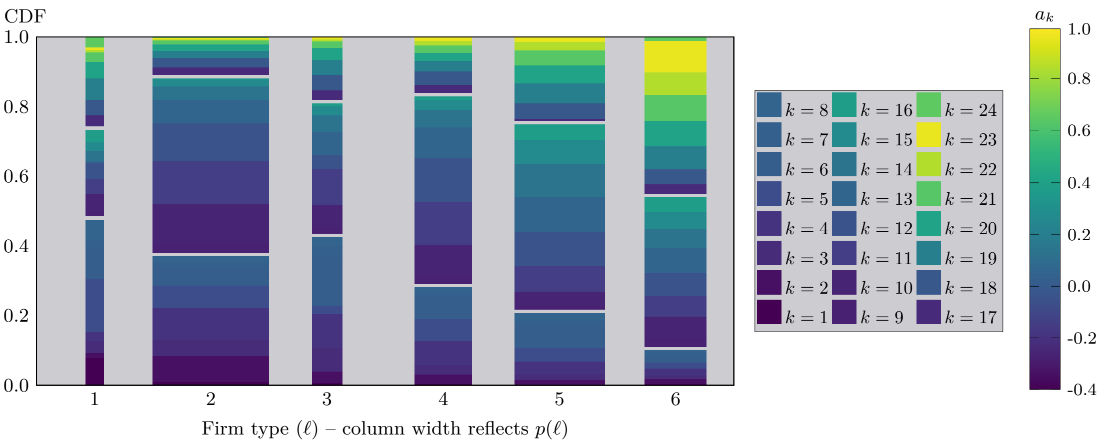
Figure 5 summarizes the results for firms. Panel (a) shows the distribution $p(\ell) = \sum_k p(k, \ell)$ and firm fixed effects b_ℓ, and panel (b) displays the CDF of firm sizes by type. There are four main firm types, $\ell = 2, 4, 5, 6$, that account for 87% of all matches. There is substantially greater dispersion in the worker fixed effects a_k than in firm fixed effects b_ℓ. It is therefore not surprising that the wage variance decomposition in the next section will attribute a greater variance contribution share to $\mathbb{V}a_k$ than to $\mathbb{V}b_\ell$. Firm effects tend to be positively linked to firm size. The largest average firm size type is type 5. Firm type 6 is on average a little larger than firm type 4, and has substantially greater size dispersion than firm type 4.

Figure 5: Firm wage types and size



Notes: The left panel displays the firm type shares $p(\ell)$ and firm fixed effects $b(\ell)$. The right panel displays the CDF of firm sizes by type ℓ (in brackets the average firm size).

Figure 6: Labor force composition by firm type



Notes: Each bar corresponds to a different firm type. The width of bar ℓ is proportional to the marginal share $p(\ell)$. A heat map is used for worker fixed effects a_k. The height of each hue in each bar is proportional to $p(k, \ell)$. Horizontal white lines separate fixed worker groups.

6.3 Type sorting

Finally, Figure 6 shows the distribution of worker types by firm type, darker hues representing worker types with bigger wage fixed effects. The marginal distribution $p(\ell)$ of firm types is also reflected in the width of the bars. The overall picture is one of positive wage sorting, where high-wage workers are more prevalent in high-wage firms. Firm type 1 is something of an outlier with its high representation of the high wage block workers, but also represents a small part of the overall match population.

7 Wage variance decomposition

Next, we reconsider the usual decomposition of the variance of log wage into its usual worker, firm and sorting components. We will emphasize the role played by learning.

7.1 Static decomposition

The linear projection of $\mu_{k\ell}$ in equation (2) allows for a decomposition of the log wage variance by the law of total variance:

$$\mathbb{V}w = \mathbb{E}\,\mathbb{V}\left(w \mid k, \ell\right) + \mathbb{V}\,\mathbb{E}\left(w \mid k, \ell\right) = \mathbb{E}\sigma^2 + \mathbb{V}a + \mathbb{V}b + 2\text{Cov}\left(a, b\right) + \mathbb{V}\tilde{\mu}. \tag{3}$$

The first component, $\mathbb{E}\sigma^2$, is the within-group variance, and the following four components decompose the between-group variance into the specific contributions of worker and firm fixed effects, their covariance, and the total variance of all interactions between worker and firm types. In the following section, we discuss the decomposition along with a comparison with the standard Abowd, Kramarz, and Margolis (1999) (AKM) and Kline, Saggio, and Sølvsten (2020) (KSS) estimators relative to the variance decomposition obtained using the linear projection (2).

Table 1, column 1, shows the variance decomposition for the full sample using the model predicted distribution $p(k, \ell)$, as described in Section 4.5. Within-group variation, $\mathbb{E}\sigma^2$ contributes 12.2% of overall log-wage variation. The worker fixed effect variance $\mathbb{V}a$ is the single most dominant contribution at 60.3% of overall variation. The firm FE variance $\mathbb{V}b$ and the FE covariance $2\text{Cov}(a, b)$ contribute 7.3% and 13.6%, respectively. Finally, the linear projection of $\mu_{k\ell}$ leaves only 6.5% of the total wage variance to be explained by worker-firm interactions, $\mathbb{V}\tilde{\mu}_{k\ell}$. The implied wage sorting expressed by the correlation coefficient between worker and firm effects $\text{Corr}(a, b)$ is 0.33. The HMM-predicted total $\mathbb{V}w$ of 0.124 closely matches the empirical $\mathbb{V}w$ of 0.123 on December observations---the model accounts for essentially all of the observed wage variance.

This decomposition restitutes what we would calculate using a standard OLS regression of log wages on worker and firm types from many trajectories simulated using the model. Alternatively, we can use the Viterbi algorithm to recover the most likely worker-type sequence for each worker and the most likely firm type for each firm, and append these to our analysis sample as new observables, which we refer to as the *completed data.*[11]

[11]For the given model estimates, firm types are assigned as the most likely firm prior $\ell_j = \arg\max_\ell \tau_{j\ell}$ (see section 4.3). With this firm classification, Z^f, separately for each worker i, the Viterbi algorithm

Table 1: Full Sample variance decomposition (shares of variance)

	Full panel				3-yr panel avg	
	Model	Completed data	AKM	KSS leave-pers-yr	Completed data	KSS
Var (a)	0.603	0.587	0.544	0.515	0.659	0.732
Var (b)	0.073	0.101	0.239	0.227	0.087	0.125
2Cov (a, b)	0.136	0.139	0.050	0.067	0.106	0.051
$\mathbb{E}[\sigma^2]$	0.122					
Var$(\tilde{\mu})$	0.065					
Residual		0.172	0.167	0.190	0.149	0.091
Var (w)	0.124	0.123	0.123	0.123	0.115	0.115
$\mathrm{Corr}(a, b)$	0.325	0.286	0.069	0.098	0.222	0.085

Notes: Column 1 reports the HMM-predicted $\mathbb{V}w$ decomposition; columns 2–4 report the empirical $\mathbb{V}w$ computed on December wages. Completed data means that we hard classify workers (Viterbi algorithm) and firms. Column 2 uses hard classified workers and firms. Column 3 uses hard-classified workers and firm fixed effects. The KSS algorithm (leave-person-year) is used to reduce second-order biases.

Table 1, column 2 (completed data), shows the same variance decomposition as column 1 (HMM) but using the hard Viterbi paths. The HMM and completed-data shares are close. The contribution of the firm effect is slightly overestimated and the worker effect is correspondingly underestimated (e.g., firm effect: 0.101 vs. 0.073, worker effect: 0.587 vs. 0.603), thus confirming that some of the randomness of worker type trajectories is lost in the hard classification. Nevertheless, hard classifying latent types works remarkably well, even for conditionally Markovian, stochastic worker types. Note also that the residuals explain the same same of the variance (0.172) compared with $\mathbb{V}\tilde{\mu} + \mathbb{E}\sigma^2$.

By contrast, the AKM and KSS decompositions attribute a much larger share to the firm component (0.23 vs. 0.07) and find substantially lower sorting correlations — an order of magnitude lower (OLS: 0.07 and KSS: 0.10 vs. model: 0.33). However, this is not the usual limited-mobility bias. For such a long panel (20 years), the KSS bias correction is relatively minor.[12] The fixed worker type assumption over the long panel likely introduces some bias. To assess this, we estimate KSS on short 3-year panels and report

finds the most likely sequence of states that could have produced those observations. The tools to solve this problem are easily available as by-products of the estimation algorithm. We are making the data completion available through the online resources.

[12]Our OLS/KSS decomposition differs from that published in Kline, Saggio, and Sølvsten (2020). There are some differences in the construction of the sample, and we recoded the algorithm in FORTRAN for speed, but the main difference is that we use 20 years of data while KSS use the 1999-2001 subsample. On this subsample, we find nearly exactly the same shares (0.67, 0.11, 0.14 and 0.08, respectively, for worker variance, firm variance, twice the covariance and residual variance; and a FE correlation of 0.25).

in column 6 a simple average over the full period of their variance decompositions. To implement the KSS bias correction, the 3-year panels discard a substantial fraction of observations (roughly half of the spell-years). To facilitate comparison, column 5 shows the hard classification decomposition done on the 3-year panels. The gap between KSS and our results do indeed narrow in support that the fixed worker type assumption biases results in the long panel, but there remains a significant gap. We will emphasize in the next section, that the gap is primarily a result of a divergence for the early years of the panel.

The sum $\mathbb{V}\tilde{\mu} + \mathbb{E}\sigma^2$ and the AKM/KSS residuals are close. Therefore, we want to compare the model projection $a_{it} + b_{it}$, where $a_{it} = a(k_{it})$ and $b_{it} = b(\ell_{j(i,t)})$, with the AKM projection $\alpha_i + \psi_{it}$, where $\psi_{it} = \psi_{j(i,t)}$ (with obvious notations). The worker type k_{it}, despite living in a smaller space than index i, varies over time. Hence, k_{it} is **not** a classification of i. This is potentially an additional source of correlation between worker and firm types. The results in Table 1 imply that Cov(a_{it}, b_{it}) is about twice as large as Cov(α_i, ψ_{it}). The next section will study the mechanisms of this correlation further. For the moment, we take the greater model covariance as given.

Second, the variance of the projection of log wages on fixed firm types $b_{it} = b(\ell_{j(i,t)})$ is necessarily smaller than the variance of $\psi_{it} = \psi_{j(i,t)}$, because the firm type ℓ_j is a classification of the firm ID j. As already said, this argument does not apply to a_{it}. Thus, $\mathbb{V}a_{it}$ can be greater that $\mathbb{V}\alpha_{it}$.

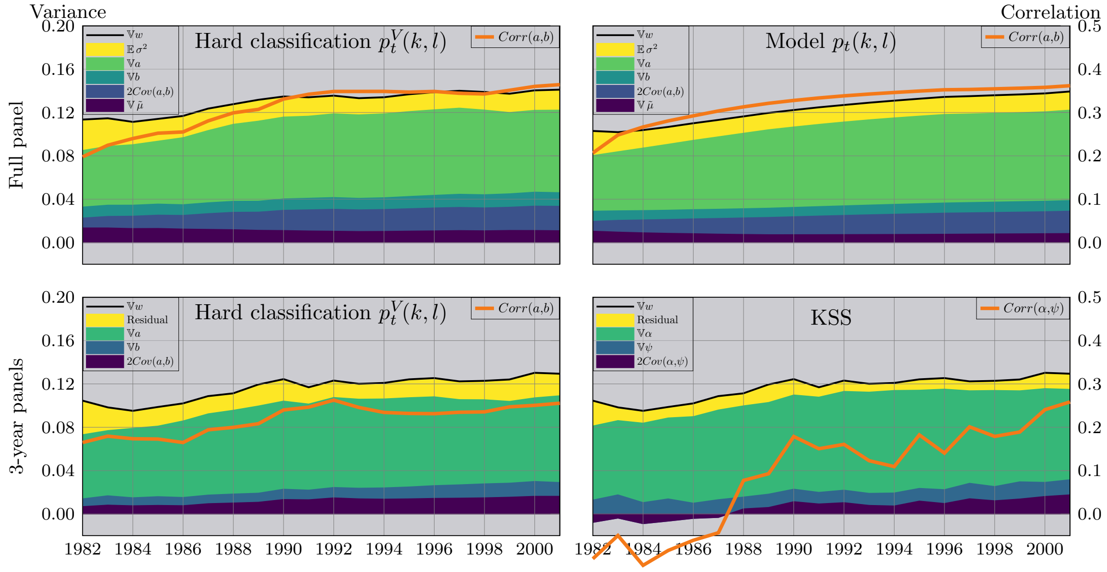
7.2 Evolution over time

In Figure 7, we show the variance decomposition for all panel years. The top row shows results for the full panel where the linear projection is done according to the model predicted match distribution in year t, $p_t(k, \ell)$ in the right panel and according to the hard Viterbi classification match distribution in year t, $p_t^V(k, \ell)$. The bottom row shows results for an overlapping series of 3-year panels. The left panel shows the $\mathbb{V}w$ decomposition using the hard classification types from the full panel applied to the relevant 3-year panel sample. The right hand panel shows the re-estimated KSS $\mathbb{V}w$ decomposition.

First, we find confirmation that the Viterbi algorithm does a good job at representing the data compared to the empirical $\mathbb{V}w$ shown in Figure 3. However, the variance of log wages that is predicted by the model (black solid line) misses the kink in 1990, and prefers a smoother trend. Again, we confirm the ability of the model to capture calendar trends with only initial cohort effects.

Second, We find that the observations in Bagger, Sørensen, and Vejlin (2013) and

Figure 7: Wage variance decomposition and wage sorting over panel years



Notes: Left scale: variance decomposition; right scale: correlation between the equation (2) worker and firm effects. The black solid line is the expected total log wage variance predicted by the model (right panel) and by the hard classification (left panel).

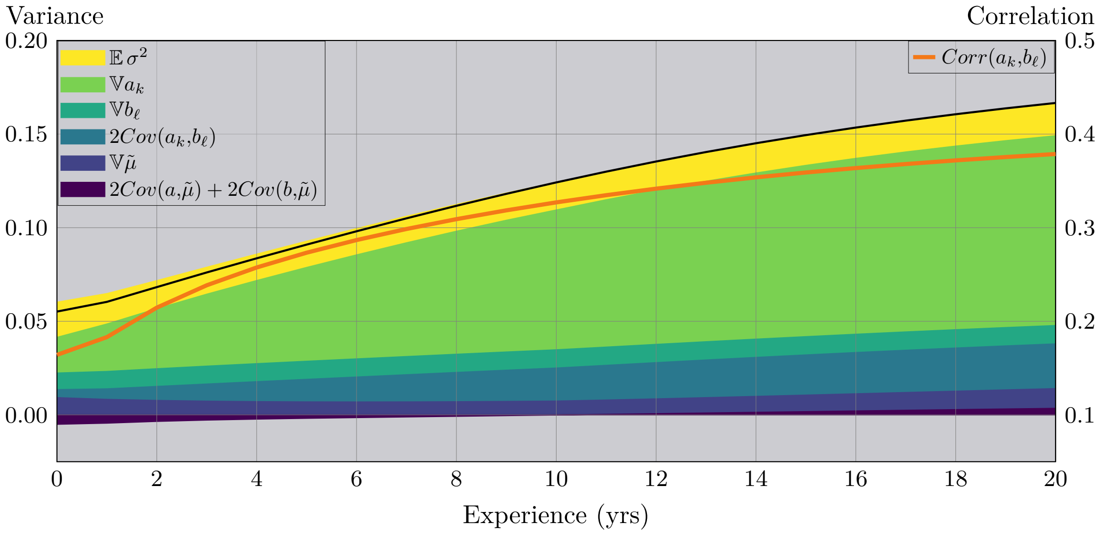
Card, Heining, and Kline (2013) that wage sorting is increasing over time in Denmark and Germany are also present in the Italian data using our estimator. Specifically, sorting is rapidly increasing during the 1980s from a correlation coefficient of 0.22 up to 0.38. Using the hard classification $p_t^V(k,\ell)$, wage sorting remains largely unchanged from 1990 and onward. The model smoothes this kink with a gradual reduction in the correlation coefficient increase over time. Overall, the importance of the covariance component is increasing over time in the wage variance decomposition. This aside, the overall qualitative conclusions from Table 1 remain the same.

Turning to the 3-year panel results, we see an almost perfect agreement between our hard classification and a KSS corrected two-way fixed effect decomposition for the 1999-2001 panel (shown by its year 2000 midpoint). But notably we also see a substantial disagreement in the early part of the panel, where the KSS corrected wage sorting estimate turns negative. Both approaches point to an increasing sorting over panel's longitudinal dimension, but the KSS corrected two-way fixed effect approach finds a considerably greater increase, given its initially negative sorting estimate.

7.3 Evolution by experience

We finally calculate the variance decomposition to individual wages at all experience levels τ, $w_\tau(\zeta)$. The distribution of (k,ℓ) at experience level τ is given by $p_{\tau,\zeta}$ =

Figure 8: Wage variance decomposition by experience, cohort $\zeta = 1982$



$p_{0,\zeta}Q^\tau$. We use the projection of $\mu_{k\ell}$ obtained for the *full sample*, holding constant the components $a_k, b_\ell, \widetilde{\mu}_{k\ell}$. Starting from $p_{0,\zeta}$, we simulate forward individual trajectories following the recursive equation $p_{\tau+1,\zeta} = p_{\tau,\zeta}Q$. As a result, since $p_{\tau,\zeta}(k,\ell)$ is generally different from $p(k,\ell)$, the residual $\tilde{\mu}_{k\ell}$ may not be perfectly orthogonal to a_k and b_ℓ and the variance decomposition may contain non-zero $\tilde{\mu}_{k\ell}$-covariances.

Specifically, omitting the cohort index, the whole-sample projection,

$$\begin{aligned} w_\tau &= \mu + a(k_\tau) + b(\ell_\tau) + \tilde{\mu}(k_\tau, \ell_\tau) + \sigma^2(k_\tau, \ell_\tau)N(0,1) \\ &:= \mu + a_\tau + b_\tau + \widetilde{\mu}_\tau + \sigma_\tau^2 N(0,1), \end{aligned}$$

where the normal error is independent of the whole sequence of states (k_τ, ℓ_τ), delivers the new variance decomposition by experience:

$$\mathbb{V}w_\tau = \mathbb{E}\sigma_\tau^2 + \mathbb{V}a_\tau + \mathbb{V}b_\tau + \mathbb{V}\tilde{\mu}_\tau + 2\mathrm{Cov}\left(a_\tau, b_\tau\right) + 2\mathrm{Cov}\left(a_\tau, \tilde{\mu}_\tau\right) + 2\mathrm{Cov}\left(b_\tau, \tilde{\mu}_\tau\right),$$

where the expectations are with respect to (k_τ, ℓ_τ).

Figure 8 shows this variance decomposition over 20 years for the $\zeta = 1982$ cohort. The black line is the overall wage variance, which almost triples over 20 years of experience. This is the well known result that individual wage trajectories tend to fan out with age.

The decomposition reveals that the increased wage variance is primarily driven by the worker effect. The increase in $\mathbb{V}a_\tau$ contributes about 75% of the overall 20-year increase in $\mathbb{V}w_\tau$. The worker effect contributes for 34% of the overall variance at the time of entry; 20 years later, this share is 60%. The within-group variance $\mathbb{E}\sigma_\tau^2$ decreases from 37% of the total variance to a modest 11%. The other main contributor to the wage variance is the covariance of worker and firm effects, which is also reflected in the correlation coefficient that increases continuously from .17 at entry to .39 after 20 years. Lastly, the firm effect sees its contribution declining from about 15% to 6%. These sharp results highlight the importance of worker type dynamics for the understanding of wage dynamics determinants.

8 Learning dynamics

We then study the dynamics of worker effects $a_\tau := a(k_\tau)$ as worker type k_τ changes with τ. We refer to the dynamics of a_τ as the learning process. The projection of $\mu_{k\ell}$ (equation (2)) on worker and firm types naturally gives the individual dynamics of a_k the interpretation of general learning in the sense that a_k, by construction, has the same impact on earnings regardless of which firm type the worker is paired with. Thus, in terms of the classic general vs job specific human capital literature, we consider this a learning process in general human capital.

8.1 The learning ladder

We first project the first difference of worker effects Δa_τ by lagged state (k, ℓ) on worker and firm types, separately by worker type block, $\kappa \in \{1, 2, 3\}$:

$$\overline{\Delta a}_{k\ell} := \mathbb{E}\left[\Delta a_\tau \mid (k_{\tau-1}, \ell_{\tau-1}) = (k, \ell), \ell > 0\right] = c + \alpha_k + \beta_\ell + \gamma_{k\ell}.$$

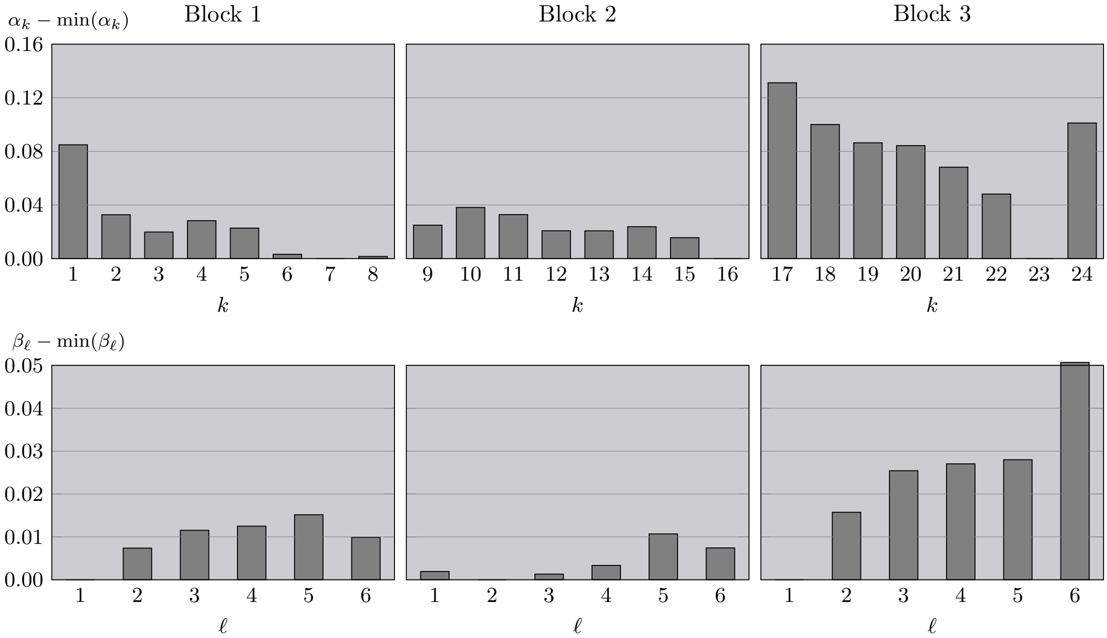
Figure 9 shows the estimates of the projection. First, we see that the contribution of the employer to worker's learning β_ℓ is substantially lower than the endogenous part α_k. Learning dynamics is primarily self-powered or endogenous, but this does not mean that firms do not contribute to learning dynamics. Second, there is a substantial difference between blocks. Block 3 workers clearly experience faster learning. Thirdly, the worker and firm effects vary in opposition. For all blocks, endogenous learning decreases as a function of the current type k, while the contribution of the firm type ℓ to learning increases with ℓ. This is particularly apparent for Block 3.

Overall, a linear approximation of the worker and firm effects ($\alpha_k = \rho_a a_k$ and $\beta_\ell = \rho_b b_\ell$) would be an accurate description, with a negative ρ_a and a positive ρ_b[13]

Summing up, we see a process with decreasing returns to general learning where high-type firms provide more learning. The learning environment in Lentz and Roys

[13]Specifically, $\rho_a = -0.106, -0.0446, -0.0845$ and $\rho_b = 0.0319, 0.0380, 0.1224$, highly significant.

Figure 9: Regression of $\Delta a_{k\ell}$ on worker and firm types



(2024) is an example of this. Here, in an otherwise standard Postel-Vinay and Robin (2002) economy, human capital is accumulated on the job according to a training intensity that is endogenously chosen specific to the firm and worker pair employment contract. In this setup, if the production function is supermodular, more productive firms pay on average higher wages and their workers also accumulate human capital more intensely, consistent with our findings.

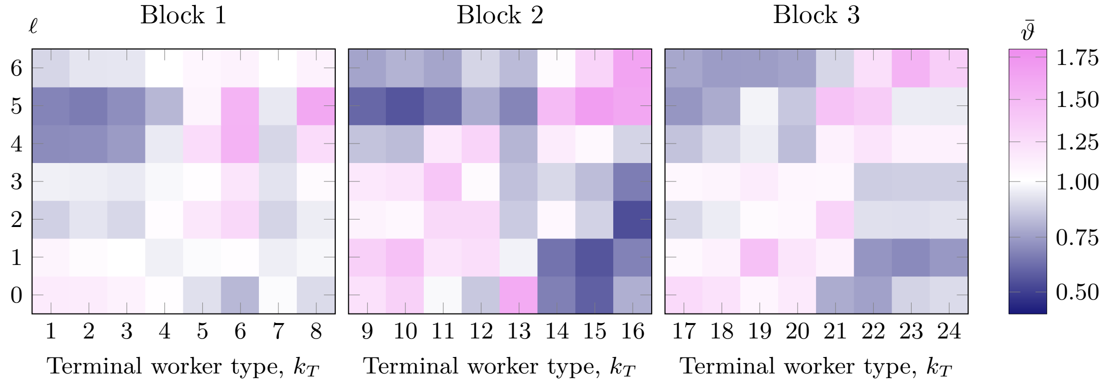
8.2 Long-term effects of learning firm effects

We next provide additional confirmation on the persistence of firm type effects in learning. We do so by quantifying the importance of the current firm type to the worker's future wage type outcomes. Specifically, conditional on k_τ, we ask how the terminal worker type realization, k_T, relates to the current firm type ℓ_τ. To do so, we calculate the normalized probability:

$$\vartheta_\zeta\left(\ell_\tau \mid k_\tau, k_T\right) = \frac{\Pr\left(\ell_\tau \mid k_\tau, k_T, \zeta\right)}{\frac{1}{K}\sum_{k'}\Pr\left(\ell_\tau \mid k_\tau, k_T = k', \zeta\right)},$$

where the normalization is done to facilitate comparison across (k_τ, ℓ_τ). This measure can be further simplified by averaging over the current worker type k_τ using the

Figure 10: Age 35 firm type probability conditional on age 45 worker type, ζ=1982.



Notes: Normalized probabilities $\Pr(\ell_\tau \mid k_T, \zeta)$.

marginal worker type distribution, $p_{\tau,\zeta}(k)$:

$$\bar{\vartheta}_\zeta(\ell_\tau \mid k_T) = \sum_{k=1}^{K} \vartheta_\zeta(\ell_\tau \mid k, k_T)\, p_{\tau,\zeta}(k).$$

The probabilities $\Pr(\ell_\tau \mid k_\tau, k_T, \zeta)$ are straightforwardly constructed from the transition matrix Q and probabilities $p_{\tau,\zeta}$ using Bayes' rule.

In Figure 10, we show $\bar{\vartheta}_\zeta(\ell_\tau \mid k_T)$ for $\zeta = 1982$, $\tau = 10$ and $T = 20$. That is, conditional on the 1982-cohort's worker type at age 45, the normalized average probability that the worker is with firm type ℓ at age 35. We see a clear ladder structure: the diagonal matrix elements are higher than the off-diagonal ones. A low (resp. high) terminal worker type in experience year 20 makes it more likely that the worker is with a low firm type or non-employed (resp. high) 10 years earlier. For example, in Block 3, going from $k_T = 17$ to $k_T = 24$ increases the probability that the worker is in firm type 6 ten years earlier almost fourfold, $1.79/0.47 = 3.8$. The same terminal outcome difference decreases the probability that the worker was non-employed ten years earlier by almost half, $1.39/0.72 = 1.9$. Thus, we are seeing very persistent impacts of firm type on worker type dynamics.

The conclusion is that the usual way of measuring sorting by the correlation between worker and firm fixed effects is hiding a firm-dependent learning mechanism. It is not just that workers of a given type match with firms of a given type, but this correlation is made internally via a firm-specific human capital accumulation process. It could be that workers and firms match randomly, but the lucky worker matching with a good firm moves up the learning ladder faster. Eventually, better firms get then more often matched with better worker. We will later measure how big is learning versus mobility as a sorting mechanism.

Figure 11: Non-employment impact on future expected a_k

Notes: Average effect of unemployment $\mathbb{E}\left[a_{\tau+n} \mid (k_\tau, \ell_\tau) = (k, 0)\right] - \mathbb{E}\left[a_{\tau+n} \mid (k_\tau, \ell_\tau) = (k, \ell)\right]$ after n years, weighted by $M_{k\ell 0}p(k, \ell)$.

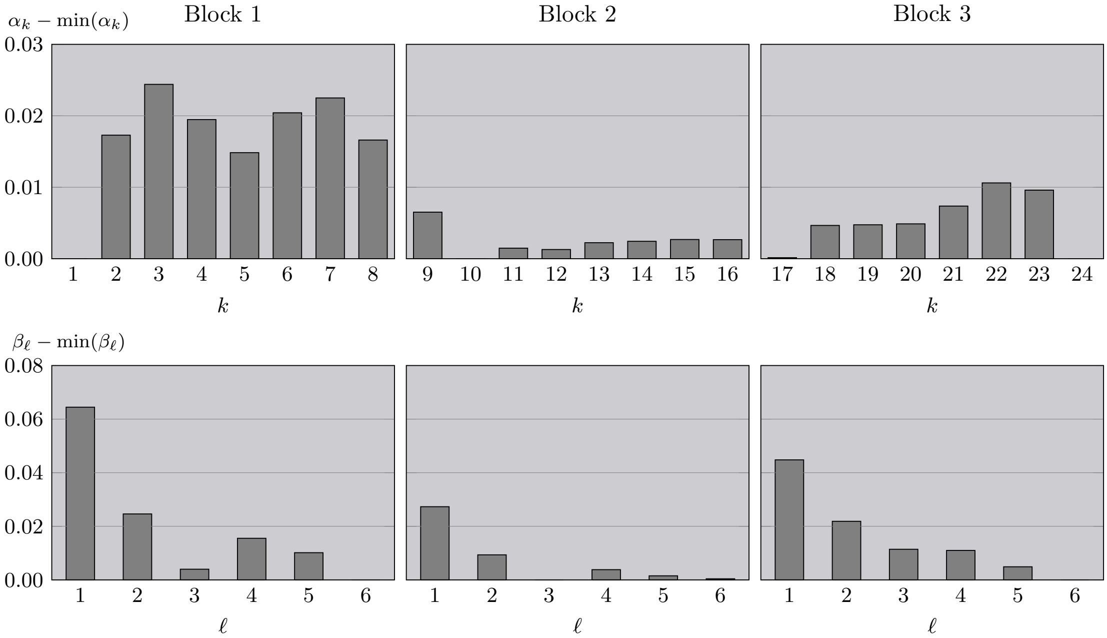
8.3 The impact of non-employment

Following up on the results in the previous section, we turn to a particular quantification of non-employment's impact on a worker's wage type dynamics. We quantify the impact of a non-employment realization on expected worker type n years after a non-employment spell,

$$\frac{\sum_{k,\ell>0} M_{k\ell 0}p(k, \ell)\left(\mathbb{E}\left[a_{\tau+n} \mid (k_\tau, \ell_\tau) = (k, 0)\right] - \mathbb{E}\left[a_{\tau+n} \mid (k_\tau, \ell_\tau) = (k, \ell)\right]\right)}{\sum_{k,\ell>0} M_{k\ell 0}p(k, \ell)}.$$

This uses the empirical match distribution $p(k, \ell)$ (any τ) and weighs by non-employment risk. The impact is shown in Figure 11. It gradually accumulates to take its maximum magnitude after about 7 years where the expected worker type is about 1.6% lower than if the non-employment shock had not happened. The impact is very persistent at about 1% even after 20 years. The long-lasting impact highlights that non-employment's negative impact on the worker's expected wage is not primarily a result of non-employment immediately depreciating the worker's wage type. Rather, non-employment takes the worker off an existing wage type growth path and on average puts the worker on a lower growth path. This results in an accumulation of losses that only begins to reverse after 7 years.

Figure 12: Regression of $\overline{\Delta b}_{k\ell}$ on worker and firm types



9 The dynamics of sorting

9.1 The firm ladder

As we did for worker types, we first project the first difference of firm effects Δb_τ by lagged state (k, ℓ) on worker and firm types:

$$\overline{\Delta b}_{k\ell} := \mathbb{E}\left[\Delta b_\tau \mid (k_{\tau-1}, \ell_{\tau-1}) = (k, \ell), \ell > 0\right] = c + \alpha_k + \beta_\ell + \gamma_{k\ell}.$$

Figure 12 shows the results of projection. As in the case of $\overline{\Delta a}_{k\ell}$, the bottom row demonstrates a negative relationship between the level and its change: The lower the level, the greater the growth. When low on the firm ladder, mobility up the ladder is stronger. The top row results indicate some tendency that upward firm mobility is positively related to worker wage type, but it is really only for block 3 workers where this pattern is somewhat robust. If one performs a linear approximation of the worker and firm fixed effects ($\alpha_k = \rho_a a_k$ and $\beta_\ell = \rho_b b_\ell$), one obtains a negative ρ_b and a positive ρ_a, but only in the case of Block 3 workers is ρ_a significant.[14] Thus, while the mobility patterns in the model provide some support for the model estimate's positive wage sorting, it is not immediately obvious that it can account for its magnitude. In the following section we discuss the interaction between learning and mobility.

[14]Specifically, by block, $\rho_a = 0.015, 0.004, 0.008$ and $\rho_b = -0.142, -0.050, -0.110$, where block 1 and 2 ρ_a not significant.

9.2 The learning and mobility contributions to sorting

Figure 8 highlights that sorting is increasing in experience. The model contains two main sorting channels. Sorting by labor mobility and sorting by worker type change — we call it learning. By the first channel, increased wage sorting is a result of high wage workers progressively moving toward high wage firms. The second channel stems from the fact that worker-type dynamics depends on the employer's type, via $A(k' \mid k, \ell)$. By the second channel, sorting increases if workers climb the worker wage type ladder faster in high wage firms than in low wage firms.

To measure the magnitude of the two channels, the change in employment conditional sorting can be decomposed as

$$\begin{aligned} \Delta \mathrm{Cov}(a_\tau, b_\tau) &:= \mathrm{Cov}\left(a_\tau, b_\tau\right) - \mathrm{Cov}\left(a_{\tau-1}, b_{\tau-1}\right) \\ &= \mathrm{Cov}\left(a_{\tau-1}, \Delta b_\tau\right) + \mathrm{Cov}\left(\Delta a_\tau, b_{\tau-1}\right) + \mathrm{Cov}\left(\Delta a_\tau, \Delta b_\tau\right). \end{aligned} \tag{4}$$

That is, the year-on-year change in sorting equals the sum of three covariances: (1) the covariance between the lagged worker effect and the change in the employer effect, (2) the covariance between the lagged employer effect and the change in the worker effect, and (3) the covariance of the cross-change. The first term, $\mathrm{Cov}\left(a_{\tau-1}, \Delta b_\tau\right)$, captures sorting by mobility, and the second, $\mathrm{Cov}\left(\Delta a_\tau, b_{\tau-1}\right)$, captures sorting by learning (worker type change). The cross term contains traces of both channels but turns out be a largely secondary contributor. These covariances can be calculated using the joint probability $p_{\tau,\zeta}(k, \ell)\, Q(k', \ell' \mid k, \ell)$ of $(k_{\tau-1}, \ell_{\tau-1}) = (k, \ell)$ and $(k_\tau, \ell_\tau) = (k', \ell')$ (see Section 4.5).

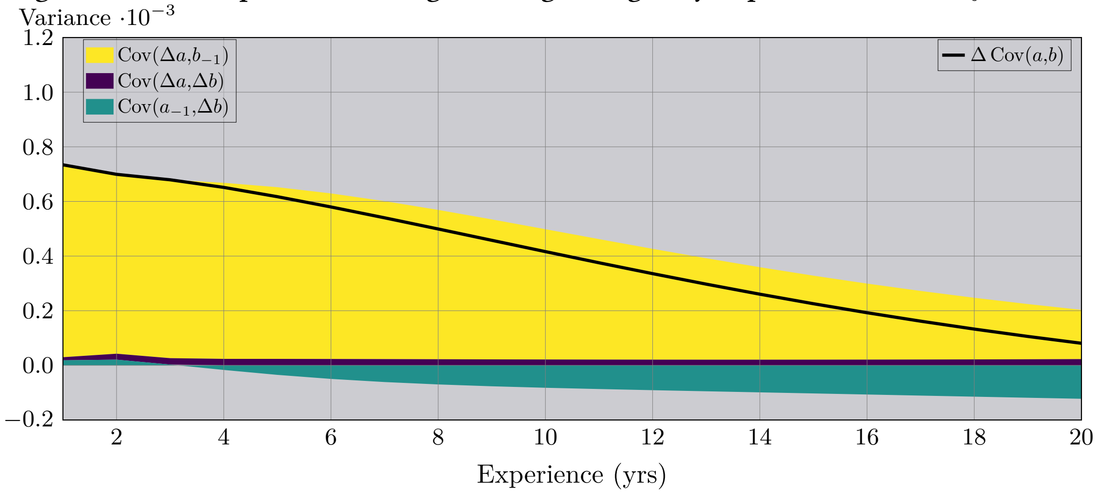
The black line in Figure 13 shows the change in sorting by experience-in-panel for the $\zeta = 1982$ cohort, the same cohort as in Figure 8. In terms of age, this reflects the cohort's experience from age 25 to age 45. Four main lessons can be drawn from this graph.

First, the sorting experience profile, as measured by the year-on-year change in the covariance of worker and firm effects, increases with experience ($\Delta\mathrm{Cov}(a_\tau, b_\tau) > 0$), albeit more slowly ($\Delta\mathrm{Cov}(a_\tau, b_\tau)$ decreasing).

Second, the decomposition of the change in learning versus mobility is illustrated by colored areas, with the mobility channel in green and the learning channel in yellow. The learning channel is a steady and a much bigger contributor at all experience levels. Yet, the absolute contribution of the mobility channel increases with experience.

Third, the learning channel is positive. One learns faster in better firms ($\mathrm{Cov}\left(\Delta a_\tau, b_{\tau-1}\right) > 0$), as already demonstrated in Subsection 8.1. The mobility channel is negative. More capable workers benefit less from mobility ($\mathrm{Cov}\left(a_{\tau-1}, \Delta b_\tau\right) < 0$). This is likely a

Figure 13: Decomposition of wage sorting changes by experience, cohort $\zeta = 1982$



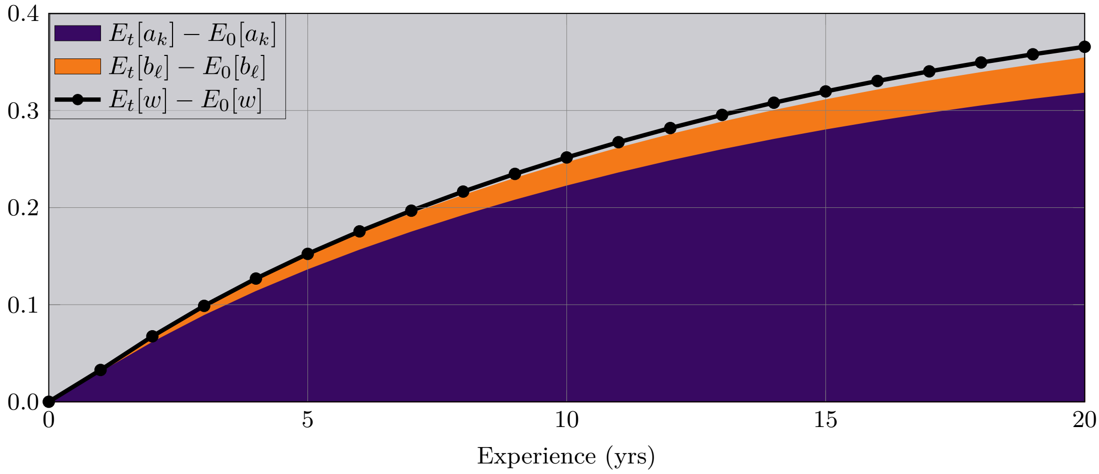
Notes: Decomposition of $\Delta\text{Cov}(a_\tau, b_\tau) := \text{Cov}\,(a_\tau, b_\tau) - \text{Cov}\,(a_{\tau-1}, b_{\tau-1})$ into $\text{Cov}\,(a_{\tau-1}, \Delta b_\tau)$, $\text{Cov}\,(\Delta a_\tau, b_{\tau-1})$ and $\text{Cov}\,(\Delta a_\tau, \Delta b_\tau)$.

consequence of the concavity of b_ℓ and of positive sorting.

Fourth, the model does not directly model an impact of experience. Its mechanism must be understood as an endogenous process, specifically a consequence of the sorting process itself. As workers grow older, they move up the learning and the firm ladders simultaneously. $\text{Cov}\,(\Delta a_\tau, b_{\tau-1})$ decreases with b, because by the time one reaches a high-wage firm, one has typically already accumulated much learning. $\text{Cov}\,(a_{\tau-1}, \Delta b_\tau)$ also decreases with a, both because of reduced mobility and of the reduced return to mobility.

These results highlight the important role of differential worker type dynamics across firm types: We find that the greater worker wage type growth in high wage type firms is the dominant contributor to sorting.[15]

Figure 14: Wage growth decomposition, cohort $\zeta = 1982$.



10 Wage growth

10.1 Cumulated wage growth

We first look at how wage level changes with experience. In Figure 2, the estimated model's average wages as a function of experience were shown in the middle column. As noted, cohort wages are increasing in experience. The younger cohorts at entry achieve about a 35% wage increase over 20 years of experience. This wage change can be decomposed into changes in the average worker wage type, the average firm wage type, and the average non-linear component. Specifically,

$$\mathbb{E}\Delta_0 w_\tau := \mathbb{E}\left(w_\tau - w_0\right) = \mathbb{E}\Delta_0 a_\tau + \mathbb{E}\Delta_0 b_\tau + \mathbb{E}\Delta_0 \tilde{\mu}_\tau.$$

Figure 14 shows the wage growth decomposition for the $\zeta = 1982$ cohort. The dotted line shows the change in average wages. The decomposition is shown as a stacked area plot. The purple area shows the growth in the average worker wage type, the orange area shows the growth in the average firm type. Any difference between the sum of the two areas and the total wage change line reflects a change in the level of worker-firm interactions. But as can be seen, there is little selection on the non-linear

[15] These are necessarily marginal considerations: On the margin, for the given level of sorting at a given experience level, we find that worker mobility is a drag on sorting. This is not to say that mobility is devoid of patterns that would induce positive sorting. We could in principle ask the counterfactual question of how much sorting would obtain if we force $A(k' \mid k, \ell)$ to be identity matrices and therefore eliminate worker type dynamics. In this case, the only source of sorting would be through mobility. It is entirely possible that such a counterfactual would demonstrate that the estimated mobility channel induces a positive sorting outcome (presumably it would be less than the one that obtains in our estimate). We are largely refraining from counterfactuals like these in the analysis, but mention it here for clarification.

component as the cohort ages.

There is growth in both the worker effect and the firm effect. Thus, workers are changing types in the direction of higher wage types and they are also moving jobs in the direction of higher wage firms. As can also be seen, the increase in the worker wage type is the dominant source of growth. Worker wage type growth accounts for 87% of overall wage growth.

10.2 Dependence to initial conditions

Digging further, we then decompose the wage variance into both the initial state (k_0, ℓ_0) and the current state (k_τ, ℓ_τ) as follows:

$$\begin{aligned}\mathbb{V}\, w_\tau &= \mathbb{E}\sigma_\tau^2 + \mathbb{V}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau\right) \\ &= \mathbb{E}\sigma_\tau^2 + \mathbb{V}\,\mathbb{E}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau \mid k_0, \ell_0\right) + \mathbb{E}\,\mathbb{V}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau \mid k_0, \ell_0\right).\end{aligned}$$

The between-initial state term is,

$$\begin{aligned}\mathbb{V}\,\mathbb{E}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau \mid k_0, \ell_0\right) = {} & \mathbb{V}\,\mathbb{E}_0 a_\tau + \mathbb{V}\,\mathbb{E}_0 b_\tau + \mathbb{V}\,\mathbb{E}_0 \widetilde{\mu}_\tau + 2\mathrm{Cov}\left(\mathbb{E}_0 a_\tau, \mathbb{E}_0 b_\tau\right) \\ & + 2\mathrm{Cov}\left(\mathbb{E}_0 a_\tau, \mathbb{E}_0 \widetilde{\mu}_\tau\right) + 2\mathrm{Cov}\left(\mathbb{E}_0 b_\tau, \mathbb{E}_0 \widetilde{\mu}_\tau\right),\end{aligned}$$

where $\mathbb{E}_0(\cdot) := \mathbb{E}\left(\cdot \mid k_0, \ell_0\right)$ is a short-cut for the expectation operator with respect to the distribution $\Pr(k_\tau, \ell_\tau \mid k_0, \ell_0) = Q^\tau(k_\tau, \ell_\tau \mid k_0, \ell_0)$, ie the element of matrix Q^τ in row (k_0, ℓ_0) and column (k_τ, ℓ_τ).

The within-initial state term is,

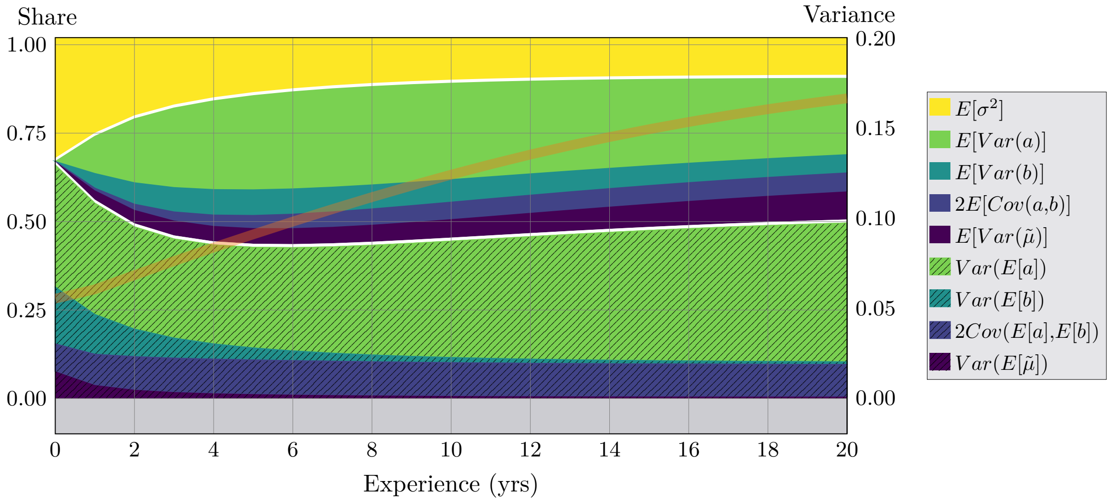
$$\begin{aligned}\mathbb{E}\,\mathbb{V}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau \mid k_0, \ell_0\right) = {} & \mathbb{E}\,\mathbb{V}_0\, a_\tau + \mathbb{E}\,\mathbb{V}_0\, b_\tau + \mathbb{E}\,\mathbb{V}_0\, \widetilde{\mu}_\tau \\ & + 2\mathbb{E}\mathrm{Cov}_0\left(a_\tau, b_\tau\right) + 2\mathrm{Cov}_0\left(a_\tau, \widetilde{\mu}_\tau\right) + 2\mathrm{Cov}_0\left(b_\tau, \widetilde{\mu}_\tau\right),\end{aligned}$$

where we use the same notational short-cut for the second-order moments of $\Pr(k_\tau, \ell_\tau \mid k_0, \ell_0)$.

Note that $\Pr(k_\tau, \ell_\tau \mid k_0, \ell_0) = Q^\tau(k_\tau, \ell_\tau \mid k_0, \ell_0)$ is independent of the cohort. The distribution of (k_0, ℓ_0) is $p_{0,\zeta}$, which depends on the cohort.

Figure 15 shows the variance decomposition between and within initial states for the $\zeta = 1982$ cohort by experience. It is a stacked area graph where each variance contribution is stated as a share of total variance. The shares are read on the left vertical axis. The total variance is read on the right vertical axis and is shown with the orange line. The decomposition is divided into three separate areas divided by narrow white bands. The top area (yellow) is the residual variation $\mathbb{E}\sigma_\tau^2$. The middle

Figure 15: Wage variance decomposition between and within initial state, $\zeta = 1982$



area represents the within initial-state variation, $\mathbb{E}\mathrm{Var}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau \mid k_0, \ell_0\right)$. Finally, the bottom area shown with cross-line pattern represents the between initial-state variation, $\mathrm{Var}\,\mathbb{E}\left(a_\tau + b_\tau + \widetilde{\mu}_\tau \mid k_0, \ell_0\right)$.

Variation across initial states (lower, shaded part) accounts for 63% of overall wage variation at the cohort's outset. By construction the within variation is equal to zero at $\tau = 0$. The other, independent source of initial variation is the residual variation, $\mathbb{E}\sigma_0^2$, which at the outset contributes the remaining 37% of wage variation. The between initial-match variation is primarily explained by the worker effect at 34 percentage points and the firm effect explains another 15 points. As experience grows, the initial conditions diminish somewhat in importance and come to account for about 50% of the overall wage variance at 20 years of experience. At that point, the residual variation $\mathbb{E}\sigma_\tau^2$ has come to account for 11%. A little more than 40% of the overall variation after 20 years is a result of average wage variation across workers and firms that is unrelated to the initial state. Again, the contribution from worker type variation, $\mathbb{E}\mathrm{Var}(a_\tau \mid k_0, \ell_0)$, dominates this source of variation at a little more than 22 percentage points. This is considerably smaller than the 38 percentage point worker type variance contribution from the between initial conditions, $\mathrm{Var}\,\mathbb{E}(a_\tau \mid k_0, \ell_0)$, at $\tau = 20$ years of experience. In other words, at 20 years of experience, about 63% of the worker's contribution to the overall wage variance is explained by the initial conditions at age 25.

This is not a surprising result. It is driven by the fixed part of worker heterogeneity. Drawing k_0 in the last group ($k_0 = 17, \ldots, 24$) puts the worker on a steeper ladder. Since we do not condition on education or other measurable skill, it was important to incorporate this block-diagonal structure in the type-transition matrix A. On the other hand, close to 40% of the worker type's contribution is not predetermined. This emphasizes that it was also important to incorporate a source of change in worker-specific types into the model.

Sorting explains an increasing part of wage variation. At the outset of the cohort, it accounts for about 7.5% of overall variation. After 20 years of experience, the sorting contribution has increased to 14.5%. The firm type contribution is decreasing in experience. Firm type variation contributes 14.6% at the outset and has declined to 5.6%, and at that point almost all of the firm type variation (86%) is unrelated to the initial conditions.

11 Conclusion

This paper develops a framework for analyzing the joint dynamics of wages, employment mobility, and worker–firm sorting in the presence of two-sided heterogeneity and evolving worker types. The model extends finite mixture approaches used with matched employer–employee data by allowing worker productivity types to follow a hidden Markov process. In particular, the worker type dynamics are allowed to depend on the latent firm type the worker is matched with. This structure makes it possible to study wage dynamics and sorting patterns that arise both from worker mobility across firms and from changes in worker productivity within matches. The estimation combines this dynamic structure with a variational expectation–maximization algorithm that jointly classifies worker and firm types using the full likelihood of wage and mobility outcomes. This approach substantially improves the classification of latent types relative to procedures that treat the worker and firm sides separately.

Applying the model to administrative matched employer–employee data from the Veneto region of Italy reveals new insights into the sources of wage dispersion and the evolution of worker–firm sorting. More broadly, the results suggest that career dynamics and wage inequality cannot be understood solely through worker mobility across firms. Changes in worker productivity within employment relationships play a central role in shaping wage growth, sorting patterns, and the evolution of wage dispersion. Incorporating worker type dynamics into models of matched employer–employee data therefore provides a richer perspective on labor market dynamics and opens new avenues for studying how firms contribute to worker productivity growth over the life cycle.

References

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High Wage Workers and High Wage Firms," *Econometrica*, 67(2), 251–334.

Abowd, J. M., K. L. McKinney, and I. M. Schmutte (2019): "Modeling Endogenous Mobility in Earnings Determination," *Journal of Business & Economic Statistics*, 37(3), 405–418.

Allman, E. S., C. Matias, and J. A. Rhodes (2009): "Identifiability of parameters in latent structure models with many observed variables," *Annals of Statistics*, 37, 3099–3132.

Altonji, J. G., A. A. Smith, and I. Vidangos (2013): "Modeling Earnings Dynamics," *Econometrica*, 81(4), 1395–1454.

Arellano, M., R. Blundell, and S. Bonhomme (2017): "Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework," *Econometrica*, 85(3), 693–734.

Arellano-Bover, J., and F. Saltiel (2026): "Differences in On-the-Job Learning across Firms," *Journal of Labor Economics*, 44(1), 149–188.

Bagger, J., F. Fontaine, F. Postel-Vinay, and J.-M. Robin (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104(6), 1551?96.

Bagger, J., and R. Lentz (2019): "An Empirical Model of Wage Dispersion with Sorting," *The Review of Economic Studies*, 86(1), 153–190.

Bagger, J., K. L. Sørensen, and R. Vejlin (2013): "Wage sorting trends," *Economics Letters*, 118(1), 63–67.

Balke, N., S. Bonhomme, and T. Lamadon (2026): "Indirect Variational Inference: Applications to Earnings Dynamics," .

Bickel, P., D. Choi, X. Chang, and H. Zhang (2013): "Asymptotic normality of maximum likelihood and its variational approximation for stochastic blockmodels," *The Annals of Statistics*, 41(4), 1922 – 1943.

Biernacki, C., G. Celeux, and G. Govaert (2000): "Assessing a mixture model for clustering with the integrated completed likelihood," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(7), 719–725, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Bonhomme, S., T. Lamadon, and E. Manresa (2019): "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 87(3), 699–739.

Borovičková, K., and C. Macaluso (2025): "Heterogeneous job ladders," *Journal of Monetary Economics*, 150(C), None.

Brault, V., C. Keribin, and M. Mariadassou (2020): "Consistency and asymptotic normality of Latent Block Model estimators," *Electronic Journal of Statistics*, 14(1), 1234 – 1268.

Browning, M., M. Ejrnæs, and J. Alvarez (2010): "Modelling Income Processes with Lots of Heterogeneity," *The Review of Economic Studies*, 77(4), 1353–1381.

Card, D., J. Heining, and P. Kline (2013): "Workplace Heterogeneity and the Rise of West German Wage Inequality," *The Quarterly Journal of Economics*, 128(3), 967–1015.

Celeux, G., and G. Govaert (1992): "A classification {EM} algorithm for clustering and two stochastic versions," *Computational Statistics & Data Analysis*, 14(3), 315 – 332.

Celisse, A., J.-J. Daudin, and L. Pierre (2012): "Consistency of maximum-likelihood and variational estimators in the stochastic block model," *Electronic Journal of Statistics*, 6(none), 1847 – 1899.

Dasgupta, A., J. Hopcroft, and F. McSherry (2004): "Spectral analysis of random graphs with skewed degree distributions," in *45th Annual IEEE Symposium on Foundations of Computer Science*, pp. 602–610.

Daudin, J. J., F. Picard, and S. Robin (2008): "A mixture model for random graphs," *Statistics and Computing*, 18(2), 173–183.

Govaert, G., and M. Nadif (2008): "Block clustering with Bernoulli mixture models: Comparison of different approaches," *Computational Statistics & Data Analysis*, 52(6), 3233–3245.

Gregory, V. (2026): "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search," *Working Paper*.

Gualdani, C., E. Pastorino, A. de Paula, and S. Salgado (2026): "Identifying Uncertainty, Learning about Productivity, and Human Capital Acquisition: A Reassessment of Labor Market Sorting and Firm Monopsony Power," *NBER Working Paper*, (34973).

Guvenen, F., F. Karahan, S. Ozkan, and J. Song (2021): "What Do Data on Millions of U.S. Workers Reveal About Lifecycle Earnings Dynamics?," *Econometrica*, 89(5), 2303–2339.

Hagedorn, M., T. H. Law, and I. Manovskii (2017): "Identifying Equilibrium Models of Labor Market Sorting," *Econometrica*, 85, 29–65.

Hoffmann, F. (2019): "HIP, RIP, and the robustness of empirical earnings processes," *Quantitative Economics*, 10(3), 1279–1315.

Jaakkola, T. S. (2001): "Tutorial on Variational Approximation Methods," in *Advanced Mean Field Methods: Theory and Practice*. The MIT Press.

Karrer, B., and M. E. J. Newman (2011): "Stochastic blockmodels and community structure in networks," *Physical Review E*, 83(1).

Kasahara, H., and K. Shimotsu (2014): "Nonparametric identification and estimation of the number of components in multivariate mixtures," *Journal of the Royal Statistical*

Society, Series B, 76, 97–111.

Kline, P., R. Saggio, and M. Sølvsten (2020): "Leave-out estimation of variance components," *Econometrica*, 88(5), 1859–1898.

Larremore, D. B., A. Clauset, and A. Z. Jacobs (2014): "Efficiently inferring community structure in bipartite networks," *Phys. Rev. E*, 90, 012805.

Lee, C., and D. J. Wilkinson (2019): "A review of stochastic block models and extensions for graph clustering," *Applied Network Science*, 4(1), 122.

Lentz, R., S. Piyapromdee, and J.-M. Robin (2023): "The Anatomy of Sorting - Evidence From Danish Data," *Econometrica*, 91(6), 2409–2455.

Lentz, R., and N. Roys (2024): "Training and Search On the Job," *Review of Economic Dynamics*, 53(None), 123–146.

Lise, J., C. Meghir, and J.-M. Robin (2016): "Matching, sorting and wages," *Review of Economic Dynamics*, 19, 63 – 87, Special Issue in Honor of Dale Mortensen.

Lise, J., and F. Postel-Vinay (2020): "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, 110(8), 2328–2376.

Lochner, L., and Y. Shin (2014): "Understanding Earnings Dynamics: Identifying and Estimating the Changing Roles of Unobserved Ability, Permanent and Transitory Shocks," NBER Working Papers 20068, National Bureau of Economic Research, Inc.

Longepierre, L., and C. Matias (2019): "Consistency of the maximum likelihood and variational estimators in a dynamic stochastic block model," *Electronic Journal of Statistics*, 13(2), 4157 – 4223.

Matias, C., and V. Miele (2017): "Statistical clustering of temporal networks through a dynamic stochastic block model," *Journal of the Royal Statistical Society Series B*, 79(4), 1119–1141.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

Moffitt, R. A., and P. Gottschalk (1995): "Trends in the Covariance Structure of Earnings in the U.S., 1969-1987," *Johns Hopkins University, Department of Economics.*

Postel-Vinay, F., and J.-M. Robin (2002): "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity," *Econometrica*, 70(6), 2295–2350.

Rabiner, L. (1989): "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, 77(2), 257–286.

Sorkin, I. (2018): "Ranking Firms Using Revealed Preference," *The Quarterly Journal of Economics*, 133(3), 1331–1393.

Taber, C., and R. Vejlin (2020): "Estimation of a Roy/Search/Compensating Differential Model of the Labor Market," *Econometrica*, 88(3), 1031–1069.

Zhao, Y., N. Hao, and J. Zhu (2024): "Variational Estimators of the Degree-corrected

Latent Block Model for Bipartite Networks," *Journal of Machine Learning Research*, 25(150), 1–42.

ONLINE APPENDIX

A Proof of the identification Theorem

A.1 Identifying matrices

We start by constructing a collection of identifying matrices. Fix one firm type $\ell = 1, ..., L$. We assume wages discrete to simplify the argument.

Within spell wage dynamics. The probability $\mathbb{P}(\ell, w, \neg, w')$ of two wages w and w' in periods 1 and 2 in the same firm of type ℓ (the $\neg$ sign means no firm change between the two periods) is, with obvious notations,

$$\begin{aligned}\mathbb{P}(\ell, w, \neg, w') &= \sum_{k,k'} m(k, \ell)\, f(w|k, \ell)\, M(\neg|k, \ell)\, A(k'|k, \ell)\, f(w'|k', \ell)\\ &= \sum_{k'} \left(\sum_{k} m(k, \ell)\, f(w|k, \ell)\, M(\neg|k, \ell)\, A(k'|k, \ell) \right) f(w'|k', \ell),\end{aligned}$$

where the sums are over $\{1, ..., K\}$. Let us stack the parameters into the following matrices,

$$\Pi_\ell = \text{diag}[m(k, \ell)]_k, \quad F_\ell = [f(w|k, \ell)]_{w\times k}, \quad A_\ell = [A(k'|k, \ell)]_{k\times k'}, \quad D_{\ell\neg} = \text{diag}[M(\neg|k, \ell)]_k.$$

We have $Q_{\ell\neg} \equiv [\mathbb{P}(\ell, w, \neg, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\neg} A_\ell F_\ell^\top$.

Between spell wage dynamics. The probability of wages w and w' in periods 1 and 2 with a job change from ℓ to ℓ' is

$$\mathbb{P}(\ell, w, \ell', w') = \sum_{k,k'} m(k, \ell)\, f(w|k, \ell)\, M(\ell'|k, \ell)\, A(k'|k, \ell)\, f(w'|k', \ell').$$

Denoting $D_{\ell\ell'} = \text{diag}[M(\ell'|k, \ell)]_k$, we have

$$Q_{\ell\ell'} \equiv [\mathbb{P}(\ell, w, \ell', w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\ell'} A_\ell F_{\ell'}^\top.$$

Transition to unemployment. The probability of wage w in period 1 in a firm of type $\ell = 1, ..., L$ followed by a transition to non-employment in period 2 is

$$\mathbb{P}(\ell, w, 0) = \sum_k m(k, \ell)\, f(w|k, \ell)\, M(0|k, \ell).$$

In matrix notations we have

$$Q_{\ell 0} = [\mathbb{P}(\ell, w, 0)]_w = F_\ell \Pi_\ell D_{\ell 0} e_K,$$

for $D_{\ell 0} = \mathrm{diag}[\delta_{k\ell}]_k$ and e_K is the K-vector of ones.

Wage dynamics with an unemployment interruption. Lastly, consider the probability of wages w and w' and employer types ℓ and ℓ' in periods 1 and 3 with an intermediate non-employment spell in period 2:

$$\mathbb{P}(\ell, w, 0, \ell', w') = \sum_{k,k',k''} m(k,\ell)\, f(w|k,\ell)\, M(0|k,\ell)\, A(k'|k,\ell)\, M(\ell'|k',0)\, A(k''|k',0)\, f(w'|k'',\ell').$$

In matrix terms, we have

$$Q_{\ell 0 \ell'} = [\mathbb{P}(\ell, w, 0, \ell', w')]_{w \times w'} = F_\ell \Pi_\ell D_{\ell 0} A_\ell D_{0\ell'} A_0 F_{\ell'}^\top,$$

with $A_0 = [A(k'|k,0)]_{k\times k'}$, $D_{\ell 0} = \mathrm{diag}[M(0|k,\ell)]_k$ and $D_{0\ell'} = \mathrm{diag}[M(\ell'|k,0)]_k$.

A.2 Proof of the Theorem

We prove identification through the following steps.

1. Consider the matrix of wage changes within the same firm of group ℓ,

$$Q_{\ell\neg} \equiv [\mathbb{P}(\ell, w, \neg, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\neg} A_\ell F_\ell^\top = F_\ell \Pi_\ell D_{\ell\neg} G_\ell^\top$$

   with $G_\ell = F_\ell A_\ell^\top$. Given the SVD $Q_{\ell\neg} = U_\ell \Lambda_\ell V_\ell^\top$,[16] write

$$U_\ell^\top Q_{\ell\neg} V_\ell \Lambda_\ell^{-1} = U_\ell^\top F_\ell \Pi_\ell D_{\ell\neg} G_\ell^\top V_\ell \Lambda_\ell^{-1} = I_K.$$

   Let $W_\ell = U_\ell^\top F_\ell \Pi_\ell D_{\ell\neg}$. Then $W_\ell^{-1} = G_\ell^\top V_\ell \Lambda_\ell^{-1}$.

2. Consider the matrix of wage changes between two spells in the same firm group ℓ,

$$Q_{\ell\ell} \equiv [\mathbb{P}(\ell, w, \ell, w')]_{w\times w'} = F_\ell \Pi_\ell D_{\ell\ell} A_\ell F_\ell^\top = F_\ell \Pi_\ell D_{\ell\ell} G_\ell^\top.$$

   It holds that

$$U_\ell^\top Q_{\ell\ell} V_\ell \Lambda_\ell^{-1} = U_\ell^\top F_\ell \Pi_\ell D_{\ell\ell} G_\ell^\top V_\ell \Lambda_\ell^{-1} = W_\ell D_{\ell\neg}^{-1} D_{\ell\ell} W_\ell^{-1}.$$

[16] For a matrix $Q \in \mathbb{R}^{m\times n}$ of rank $K \leq m, n$, the SVD has $Q = U\Lambda V^\top$ where U is $m \times K$ with orthonormal columns, Λ is diagonal $K \times K$ and V is $n \times K$ with orthonormal columns.

The diagonal matrix $D_{\ell\neg}^{-1}D_{\ell\ell}$ is identified as the eigenvectors (up to labeling). And knowing that the entries of $D_{\ell\neg}^{-1}D_{\ell\ell} = \mathrm{diag}[M(\ell|k,\ell)/M(\neg|k,\ell)]$ are all distinct (Condition 5), any matrix of eigenvectors is thus of the form

$$\widetilde{W}_\ell = W_\ell \Delta_\ell^{-1} = U_\ell^\top F_\ell \Pi_\ell D_{\ell\neg}\Delta_\ell^{-1}, \quad \widetilde{W}_\ell^{-1} = \Delta_\ell G_\ell^\top V_\ell \Lambda_\ell^{-1},$$

for some non singular diagonal matrix Δ_ℓ.

3. Consider $\widetilde{W}_\ell^{-1}$. The matrix $\widetilde{W}_\ell^{-1}\Lambda_\ell V_\ell^\top$ identifies $\Delta_\ell G_\ell^\top$. The rows of F_ℓ and the columns of A_ℓ sum to one. It follows that the rows of G_ℓ also sum to one, which then identifies he diagonal matrix Δ_ℓ from $\Delta_\ell G_\ell^\top$, and G_ℓ is also identified. The matrix $\widetilde{W}_\ell$ then identifies $F_\ell \Pi_\ell D_{\ell\neg}$. And since the rows of F_ℓ sums to one, the diagonal matrix $\Pi_\ell D_{\ell\neg}$ is identified and F_ℓ is identified. Since $G_\ell = F_\ell A_\ell^\top$ with G_ℓ and F_ℓ both identified, we finally also identify A_ℓ. It is quite remarkable that $Q_{\ell\neg}$ and $Q_{\ell\ell}$ (mobility within and between firms of type ℓ) suffice to identify F_ℓ, A_ℓ and the products of diagonal matrices $D_{\ell\neg}^{-1}D_{\ell\ell}$ and $\Pi_\ell D_{\ell\neg}$.

4. Having proceeded independently for each firm type ℓ, how do we know that the worker group k that we have thus labelled for firm type ℓ corresponds to the worker group k' that we have thus labelled for firm type ℓ'? Take matrix $Q_{\ell\ell'} = [\mathbb{P}(\ell, w, \ell', w')]_{w\times w'}$ corresponding to a job mobility from ℓ to ℓ'. We know that
$$Q_{\ell\ell'} = F_\ell \Pi_\ell D_{\ell\ell'} A_\ell F_{\ell'}^\top = (F_\ell \Pi_\ell D_{\ell\neg})(D_{\ell\neg}^{-1}D_{\ell\ell'})A_\ell F_{\ell'}^\top$$
with $D_{\ell\neg}^{-1}D_{\ell\ell'}$ diagonal. So given arbitrarily chosen worker-group labels for ℓ and ℓ' one should relabel worker groups for ℓ' (say) by reordering the columns of $F_{\ell'}$ so that $(F_\ell \Pi_\ell D_{\ell\neg})^+ Q_{\ell\ell'} (F_{\ell'}^+)^\top A_\ell^{-1} = D_{\ell\neg}^{-1}D_{\ell\ell'}$ is a diagonal matrix (denoting $B^+ = (B^\top B)^{-1}B^\top$ for any full column rank matrix B). This procedure also identifies $D_{\ell\neg}^{-1}D_{\ell\ell'}$.

5. Note that $D_{\ell\neg} + \sum_{\ell'} D_{\ell\ell'} = I_K$. Hence
$$D_{\ell\neg} = [I_K + \sum_{\ell'} D_{\ell\neg}^{-1}D_{\ell\ell'}]^{-1}$$
is identified, and so are all $D_{\ell\ell'}$. Hence, we also identify $D_{\ell\neg}$ and Π_ℓ from $D_{\ell\neg}^{-1}D_{\ell\ell}$ and $\Pi_\ell D_{\ell\neg}$.

6. Consider $Q_{\ell 0} = [\mathbb{P}(\ell, w, 0)]_w = F_\ell \Pi_\ell D_{\ell 0} e_K$, for $D_{\ell 0} = \mathrm{diag}[M(0|k,\ell)]_k$ and e_K is the K-vector of ones. Hence $D_{\ell 0} e_K = [M(0|k,\ell)]_k$ is identified by regressing $[\mathbb{P}(\ell, w, 0)]_w$ on the columns of $F_\ell \Pi_\ell$.

7. Consider $Q_{\ell 0 \ell'} = [\mathbb{P}(\ell, w, 0, \ell', w')]_{w \times w'} = F_\ell \Pi_\ell D_{\ell 0} A_\ell D_{0\ell'} A_0 F_{\ell'}^\top$, with $A_0 = [A(k'|k,0)]_{k \times k'}$ and $D_{0\ell'} = \mathrm{diag}[M(\ell'|k,0)]_k$. This identifies $D_{0\ell'} A_0 = [A(k'|k,0)M(\ell'|k,0)]$ for all $\ell' = 1, ..., L$. Consider diagonal terms $A(k|k,0)M(\ell'|k,0)$. One can eliminate $A(k|k,0)$(non zero by Condition 6) from these equations and identify $M(\ell'|k,0)/M(1|k,0)$ for all $\ell' = 2, ..., L$. Hence, all $M(\ell'|k,0)$, $\ell' = 1, ..., L$, are identified because $\sum_{\ell'=1}^{L} M(\ell'|k,0) = 1$. Then the whole matrix A_0 follows.

B MLE for the complete information model (online)

We start by rewriting the log-likelihood as follows:

$$\ln \mathcal{L}\left(Z^f\right) = \sum_{\ell=1}^{L} n_\ell^f \ln \pi_\ell^f, \quad \ln \mathcal{L}\left(Z^w \mid Z^f\right) = \sum_{k=1}^{K} n_k^w(1) \ln \pi_k^w + \sum_{\ell=0}^{L} \sum_{k=1}^{K} \sum_{k'=1}^{K} n_{k\ell k'}^w \ln A_{k\ell k'},$$

where $n_\ell^f = \sum_{j=1}^{J} Z_{j\ell}^f$ counts the number of firms of type ℓ, and where $n_k^w(1)$ counts the number of workers initially of type k and $n_{k\ell k'}^w$ the number of workers of type k in a state ℓ in one period, and switching to type k' in the next period:

$$n_k^w(1) = \sum_{i=1}^{I} Z_{ik}^w(1), \quad n_{k0k'}^w = \sum_{i=1}^{I} \sum_{s=1}^{S_i - 1} Z_{ik}^w(s) Y_{i0}(s) Z_{ik'}^w(s+1),$$
$$n_{k\ell k'}^w = \sum_{i=1}^{I} \sum_{s=1}^{S_i - 1} Z_{ik}^w(s) \left(\sum_{j=1}^{J} Y_{ij}(s) Z_{j\ell}^f \right) Z_{ik'}^w(s+1), \ \ell \geq 1.$$

The MLE of type probabilities are obtained by maximizing $\ln \mathcal{L}\left(Z^f\right)$ and $\ln \mathcal{L}\left(Z^w \mid Z^f\right)$ subject to the constraint that probabilities must add up to one, that is

$$\widehat{\pi}_\ell^f = \frac{n_\ell^f}{J}, \quad \widehat{\pi}_k^w = \frac{n_k^w(1)}{I}, \quad \widehat{A}_{k\ell k'} = \frac{n_{k\ell k'}^w}{\sum_{k'=1}^{K} n_{k\ell k'}^w},$$

where $\sum_{k'=1}^{K} n_{k\ell k'}^w$ is the number of matches (k, ℓ) in periods $1 : S_i - 1$.

We next turn to the log-likelihood of emissions given worker and firm types:

$$
\begin{aligned}
\ln \mathcal{L}\left(X \mid Z^f, Z^w\right) = &\sum_{j=1}^{L}\sum_{\ell=1}^{L} n^w_{j\ell} \ln \theta_{j\ell} + \sum_{k=1}^{K}\sum_{\ell=0}^{L} n^w_{k\ell}(1) \ln m_{k\ell} \\
&+ \sum_{k=1}^{K}\sum_{\ell'=1}^{L} n^w_{k0\ell'} \ln M_{k0\ell'} + \sum_{k=1}^{K} n^w_{k0\neg} \ln M_{k0\neg} \\
&+ \sum_{k=1}^{K}\sum_{\ell=1}^{L}\sum_{\ell'=0}^{L} n^w_{k\ell\ell'} \ln M_{k\ell\ell'} + \sum_{k=1}^{K}\sum_{\ell=1}^{L} n^w_{k\ell\neg} \ln M_{k\ell\neg} \\
&+ \sum_{k=1}^{K}\sum_{\ell=1}^{L}\sum_{i=1}^{I}\sum_{s=1}^{S_i} Z^w_{ik}(s) \left(\sum_{j=1}^{J} Y_{ij}(s) Z^f_{j\ell}\right) \ln \varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right),
\end{aligned}
$$

where

$$
n^w_{j\ell} = \sum_{i=1}^{I} Y_{ij}(1) Z^f_{j\ell} + \sum_{i=1}^{I}\sum_{s=2}^{S_i}\sum_{j_-=0}^{J} \mathbf{1}\{j_- \neq j\} Y_{ij_-}(s-1) Y_{ij}(s) Z^f_{j\ell},\ j, \ell \geq 1,
$$

is the number of workers initially employed by firm j/type ℓ or who later are observed to move to such a firm and type. The other worker counts are transparent:

$$
\begin{aligned}
n^w_{k0}(1) &= \sum_{i=1}^{I} Z^w_{ik}(1) Y_{i0}(1), \\
n^w_{k\ell}(1) &= \sum_{i=1}^{I}\sum_{s=1}^{S_i-1} Z^w_{ik}(1) \sum_{j=1}^{J} Y_{ij}(1) Z^f_{j\ell},\ \ell \geq 1, \\
n^w_{k0\ell'} &= \sum_{i=1}^{I}\sum_{s=1}^{S_i-1} Z^w_{ik}(s) Y_{i0}(s) \sum_{j=1}^{J} Y_{ij}(s+1) Z^f_{j\ell'},\ \ell' \geq 1, \\
n^w_{k0\neg} &= \sum_{i=1}^{I}\sum_{s=1}^{S_i-1} Z^w_{ik}(s) Y_{i0}(s) D_i(s), \\
n^w_{k\ell 0} &= \sum_{i=1}^{I}\sum_{s=1}^{S_i-1} Z^w_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s) Z^f_{j\ell} Y_{i0}(s+1),\ \ell \geq 1, \\
n^w_{k\ell\ell'} &= \sum_{i=1}^{I}\sum_{s=1}^{S_i-1} Z^w_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s) Z^f_{j\ell} \sum_{j'=1}^{J} Y_{ij'}(s+1) Z^f_{j'\ell'},\ \ell, \ell' \geq 1, \\
n^w_{k\ell\neg} &= \sum_{i=1}^{I}\sum_{s=1}^{S_i-1} Z^w_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s) Z^f_{j\ell} D_i(s),\ \ell \geq 1,
\end{aligned}
$$

We may also set $n^w_{k00} = 0$ (no transition from unemployment to unemployment as they are counted in $n^w_{k0\neg}$).

Maximizing $\ln \mathcal{L}\left(X \mid Z^f, Z^w\right)$ subject to appropriate constraints yields the following simple moment estimators:

$$\begin{aligned}
\widehat{\theta}_{j\ell} &= \frac{n^w_{j\ell}}{\sum_{j=1}^J n^w_{j\ell}}, \\
\widehat{m}_{k\ell} &= \frac{n^w_{k\ell}(1)}{\sum_{\ell=0}^L n^w_{k\ell}(1)},\ \ell \geq 0, \\
\widehat{M}_{k0\neg} &= \frac{n^w_{k0\neg}}{n^w_{k0\neg} + \sum_{\ell'=1}^L n^w_{k0\ell'}}, \quad \widehat{M}_{k0\ell'} = \frac{n^w_{k0\ell'}}{n^w_{k0\neg} + \sum_{\ell'=1}^L n^w_{k0\ell'}},\ \ell' \geq 1, \\
\widehat{M}_{k\ell\neg} &= \frac{n^w_{k\ell\neg}}{n^w_{k\ell\neg} + \sum_{\ell'=0}^L n^w_{k\ell\ell'}}, \quad \widehat{M}_{k\ell\ell'} = \frac{n^w_{k\ell\ell'}}{n^w_{k\ell\neg} + \sum_{\ell'=0}^L n^w_{k\ell\ell'}},\ \ell \geq 1, \ell' \geq 0,
\end{aligned}$$

and

$$\begin{aligned}
\psi_1'(\widehat{\eta}_{k\ell}) = \widehat{\mu}_{k\ell} = \widehat{W}_{k\ell} &= \frac{1}{n^w_{k\ell}} \sum_{i=1}^I \sum_{s=1}^{S_i} Z^w_{ik}(s) \sum_{j=1}^J Y_{ij}(s) Z^f_{j\ell} W_i(s), \\
\psi_2'(\widehat{\eta}_{k\ell}) = \widehat{\mu}^2_{k\ell} + \widehat{\sigma}^2{}_{k\ell} = \widehat{W^2}_{k\ell} &= \frac{1}{n^w_{k\ell}} \sum_{i=1}^I \sum_{s=1}^{S_i} Z^w_{ik}(s) \sum_{j=1}^J Y_{ij}(s) Z^f_{j\ell} W^2_i(s).
\end{aligned}$$

Finally, we deduce the following value of the log-likelihood of emissions given types:

$$\begin{aligned}
\ln \mathcal{L}\left(X \mid Z^f, Z^w\right) = \sum_{\ell=1}^L n^f_\ell \sum_{j=1}^J \widehat{\theta}_{j\ell} \ln \theta_{j\ell} + \sum_{k=1}^K n^w_k(1) \sum_{\ell=0}^L \widehat{m}_{k\ell} \ln m^w_{k\ell} \\
+ \sum_{k=1}^K \sum_{\ell=0}^L n^w_{k\ell} \left(\sum_{\ell'=0}^L \widehat{M}_{k\ell\ell'} \ln M_{k\ell\ell'} + \widehat{M}_{k\ell\neg} \ln M_{k\ell\neg} \right) \\
+ \sum_{k=1}^K \sum_{\ell=1}^L n^w_{k\ell} \ln \varphi\left(\widehat{W}_{k\ell}, \widehat{W^2}_{k\ell}, \eta_{k\ell}\right),
\end{aligned}$$

where

$$n^f_\ell = \sum_{j=1}^J n^w_{j\ell}, \quad n^w_k(1) = \sum_{\ell=0}^L n^w_{k\ell}(1), \quad n^w_{k\ell} = n^w_{k\ell\neg} + \sum_{\ell'=0}^L n^w_{k\ell\ell'}.$$

C Details on the VEM estimation algorithm (online)

C.1 Expected log-likelihood

The expected log-likelihood is

$$\mathbb{E}_R \ln \mathcal{L}(X, Z^f, Z^w) = \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^f, Z^w)$$

where

$$\begin{aligned}\ln \mathcal{L}(X, Z^f, Z^w) &= \ln \mathcal{L}(Z^f) + \ln \mathcal{L}(X, Z^w \mid Z^f) \\ &= \ln \mathcal{L}(Z^f) + \sum_{i=1}^{I} \ln \mathcal{L}_i(X, Z^w \mid Z^f)\end{aligned}$$

with

$$\mathbb{E}_R \ln \mathcal{L}(Z^f) = \sum_{j=1}^{J} \sum_{\ell=1}^{L} \tau_{j\ell} \ln \pi_\ell^f$$

and where the individual likelihoods $\ln \mathcal{L}_i(X, Z^w \mid Z^f)$ can be integrated iteratively exactly like the complete log-likelihood as we now explain.

The initial worker type probability is still

$$\alpha_k(1) = \pi_k^w,$$

but the worker type transition probabilities for $s = 2, ..., S_i$ replace the certain Z^f by the probability distribution $\tau = (\tau_{j\ell})$:

$$\alpha_{kk'}(s|\tau) = \exp\left[\sum_{j=1}^{J} Y_{ij}(s-1) \sum_{\ell=0}^{L} \tau_{j\ell} \ln A_{k\ell k'}\right].$$

The emission probabilities also change in the same way:

- for $X_i(s)$:

$$\begin{aligned}\beta_k(1|\tau) = \exp\Bigg[\sum_{j=0}^{J} Y_{ij}(1) \sum_{\ell=0}^{L} \tau_{j\ell} \Bigg(&\ln m_{k\ell} + \ln \theta_{j\ell} \\ &+ \mathbf{1}\{j \neq 0\} \ln \varphi\left(T\left[W_i(1)\right], \eta_{k\ell}\right) \\ + D_i(1) \ln M_{k\ell\neg} + &\sum_{j'=0}^{J} Y_{ij'}(2) \sum_{\ell'=0}^{L} \tau_{j'\ell'} \mathbf{1}\{j' \neq j\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right]\Bigg)\Bigg],\end{aligned}$$

- for $X_i(s),\ s = 2, ..., S_i - 1$:

$$\beta_k(s|\tau) = \exp\left[\sum_{j=0}^{J} Y_{ij}(s) \sum_{\ell=1}^{L} \tau_{j\ell} \left(\mathbf{1}\{j \neq 0\} \ln \varphi\left(T\left[W_i(s)\right], \eta_{k\ell}\right)\right.\right.$$
$$\left.\left. + D_i(s) \ln M_{k\ell}^{\neg} + \sum_{j'=0}^{J} Y_{ij'}(s+1) \sum_{\ell'=0}^{L} \tau_{j'\ell'} \mathbf{1}\{j' \neq j\} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'}\right]\right)\right],$$

- for $X_i(S_i)$:

$$\beta_k(S_i|\tau) = \exp\left[\sum_{j=0}^{J} Y_{ij}(S_i) \sum_{\ell=1}^{L} \tau_{j\ell} \left(\mathbf{1}\{j \neq 0\} \ln \varphi\left(T\left[W_i(S_i)\right], \eta_{k\ell}\right) + D_i(s) \ln M_{k\ell}^{\neg}\right)\right].$$

For any distribution $R^w(Z^w)$ with

$$\mathbb{P}_{R^w}\left\{Z_i^w(s) = k\right\} = \zeta_{ik}(s), \quad \mathbb{P}_{R^w}\left\{Z_i^w(s) = k, Z_i^w(s+1) = k'\right\} = \xi_{ikk'}(s+1),$$

we finally obtain

$$\mathbb{E}_R \ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right) = \sum_{k=1}^{K} \zeta_{ik}(1) \ln\left[\alpha_k(1)\beta_k(1|\tau)\right]$$
$$+ \sum_{s=2}^{S_i} \sum_{k=1}^{K} \sum_{k'=1}^{K} \xi_{ikk'}(s) \ln \alpha_{kk'}(s|\tau) + \sum_{s=2}^{S_i} \sum_{k=1}^{K} \zeta_{ik}(s) \ln \beta_k(s|\tau).$$

C.2 M-step

The expected log-likelihood has exactly the same structure as the complete log-likelihood. It suffices to replace $Z_{j\ell}^f$ by $\tau_{j\ell}$, $Z_{ik}^w(s)$ by $\zeta_{ik}(s)$ and $Z_{ik}^w Z_{ik'}^w$ by $\xi_{ikk'}$. The VEM formulas for the emission parameters are therefore

$$\widetilde{\pi}_\ell^f = \frac{\widetilde{n}_\ell^f}{J}, \quad \widetilde{\pi}_k^w = \frac{\widetilde{n}_k^w(1)}{I}, \quad \widetilde{A}_{k\ell k'} = \frac{\widetilde{n}_{k\ell k'}^w}{\sum_{k'=1}^{K} \widetilde{n}_{k\ell k'}^w},$$

and

$$
\begin{aligned}
\widetilde{\theta}_{j\ell} &= \frac{\widetilde{n}^w_{j\ell}}{\sum_{j=1}^J \widetilde{n}^w_{j\ell}}, \\
\widetilde{m}_{k\ell} &= \frac{\widetilde{n}^w_{k\ell}(1)}{\sum_{\ell=0}^L \widetilde{n}^w_{k\ell}(1)},\ \ell \geq 0, \\
\widetilde{M}_{k0\neg} &= \frac{\widetilde{n}^w_{k0\neg}}{\widetilde{n}^w_{k0\neg} + \sum_{\ell'=1}^L \widetilde{n}^w_{k0\ell'}}, \quad \widetilde{M}_{k0\ell'} = \frac{\widetilde{n}^w_{k0\ell'}}{\widetilde{n}^w_{k0\neg} + \sum_{\ell'=1}^L \widetilde{n}^w_{k0\ell'}},\ \ell' \geq 1, \\
\widehat{M}_{k\ell\neg} &= \frac{n^w_{k\ell\neg}}{\widetilde{n}^w_{k\ell\neg} + \sum_{\ell'=0}^L \widetilde{n}^w_{k\ell\ell'}}, \quad \widehat{M}_{k\ell\ell'} = \frac{n^w_{k\ell\ell'}}{\widetilde{n}^w_{k\ell\neg} + \sum_{\ell'=0}^L \widetilde{n}^w_{k\ell\ell'}},\ \ell \geq 1, \ell' \geq 0,
\end{aligned}
$$

and

$$
\begin{aligned}
\psi_1'(\widetilde{\eta}_{k\ell}) &= \widetilde{\mu}_{k\ell} = \widetilde{W}_{k\ell} = \frac{1}{\widetilde{n}^w_{k\ell}} \sum_{i=1}^I \sum_{s=1}^{S_i} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s) \tau_{j\ell} W_i(s), \\
\psi_2'(\widetilde{\eta}_{k\ell}) &= \widetilde{\mu}^2_{k\ell} + \widetilde{\sigma}^2{}_{k\ell} = \widetilde{W^2}_{k\ell} = \frac{1}{\widetilde{n}^w_{k\ell}} \sum_{i=1}^I \sum_{s=1}^{S_i} \zeta_{ik}(s) \sum_{j=1}^J Y_{ij}(s) \tau_{j\ell} W_i^2(s).
\end{aligned}
$$

with

$$
\begin{aligned}
\widetilde{n}^f_\ell &= \sum_{j=1}^{J} \tau_{j\ell} \\
\widetilde{n}^w_k(1) &= \sum_{i=1}^{I} \zeta_{ik}(1), \\
\widetilde{n}^w_{k0k'} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \sum_{i=1}^{I} Y_{i0}(s)\xi_{ikk'}(s+1), \\
\widetilde{n}^w_{k\ell k'} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \left(\sum_{j=1}^{J} Y_{ij}(s) Z^f_{j\ell} \right) \xi_{ikk'}(s+1),\ \ell \geq 1. \\
\widetilde{n}^w_{j\ell} &= \sum_{i=1}^{I} Y_{ij}(1)\tau_{j\ell} + \sum_{i=1}^{I} \sum_{s=2}^{S_i} \sum_{j'=0}^{J} \mathbf{1}\{j' \neq j\} Y_{ij'}(s-1) Y_{ij}(s)\tau_{j\ell},\ j,\ell \geq 1, \\
\widetilde{n}^w_{k0}(1) &= \sum_{i=1}^{I} \zeta_{ik}(s) Y_{i0}(s), \\
\widetilde{n}^w_{k\ell}(1) &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s)\tau_{j\ell},\ \ell \geq 1, \\
\widetilde{n}^w_{k0\ell'} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \zeta_{ik}(s) Y_{i0}(s) \sum_{j=1}^{J} Y_{ij}(s+1)\tau_{j\ell},\ \ell' \geq 1, \\
\widetilde{n}^w_{k0\neg} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \zeta_{ik}(s) Y_{i0}(s) D_i(s), \\
\widetilde{n}^w_{k\ell 0} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s)\tau_{j\ell} Y_{i0}(s+1),\ \ell \geq 1, \\
\widetilde{n}^w_{k\ell\ell'} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s)\tau_{j\ell} \sum_{j'=1}^{J} Y_{ij'}(s+1)\tau_{j'\ell'},\ \ell,\ell' \geq 1, \\
\widetilde{n}^w_{k\ell\neg} &= \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \zeta_{ik}(s) \sum_{j=1}^{J} Y_{ij}(s)\tau_{j\ell} D_i(s),\ \ell \geq 1,
\end{aligned}
$$

We may also set $\widetilde{n}^w_{k00} = 0$ (no transition from unemployment to unemployment as they are counted in $n^w_{k0\neg}$).

Note that we also obtain the following expression for

$$
\begin{aligned}
\mathbb{E}_R \ln \mathcal{L}\left(X, Z^f, Z^w\right) = J \sum_{\ell=1}^{L} \widetilde{\pi}_\ell^f \ln \pi_\ell^f + I \sum_{k=1}^{K} \widetilde{\pi}_k^w \ln \pi_k^w \\
+ \sum_{\ell=1}^{L} \widetilde{n}_\ell^f \sum_{j=1}^{J} \widetilde{\theta}_{j\ell} \ln \theta_{j\ell} + \sum_{k=1}^{K} \widetilde{n}_k^w(1) \sum_{\ell=0}^{L} \widetilde{m}_{k\ell} \ln m_{k\ell}^w \\
+ \sum_{k=1}^{K} \sum_{\ell=0}^{L} \widetilde{n}_{k\ell}^w \left(\sum_{\ell'=0}^{L} \widetilde{M}_{k\ell\ell'} \ln M_{k\ell\ell'} + \widetilde{M}_{k\ell\neg} \ln M_{k\ell\neg} \right) \\
+ \sum_{k=1}^{K} \sum_{\ell=1}^{L} \widetilde{n}_{k\ell}^w \ln \varphi\left(\widetilde{W}_{k\ell}, \widetilde{W^2}_{k\ell}, \eta_{k\ell}\right).
\end{aligned}
$$

C.3 Worker E-step

We update $\zeta_{ik}(s)$ and $\xi_{ikk'}(s)$ in the worker E-step as follows.

The worker E-step problem is

$$
\begin{aligned}
\max_{R^w} \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^w | Z^f) - \sum_{Z^w} R^w(Z^w) \ln R^w(Z^w) \\
= \max_{R^w} \sum_{Z^w} R^w(Z^w) \ln \left[\frac{\exp\left(\sum_{Z^f} R^f(Z^f) \ln \mathcal{L}(X, Z^w \mid Z^f)\right)}{R^w(Z^w)} \right]
\end{aligned}
$$

subject to the restriction $\sum_{Z^w} R^w(Z^w) = 1$. We recognize the Kuhlback-Leibler divergence, and therefore

$$
\begin{aligned}
R^w(Z^w) &\propto \exp\left(\sum_{Z^f} R^f(Z^f) \ln \mathcal{L}(X, Z^w \mid Z^f) \right) \\
&= \prod_{i=1}^{I} \frac{\exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i(X_i, Z_i^w \mid Z^f)}{\sum_{Z_i^w} \exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i(X_i, Z_i^w \mid Z^f)} \\
&= \prod_{i=1}^{I} R_i^w(Z_i^w).
\end{aligned}
$$

This means that the optimal worker posterior distribution $R^w(Z^w)$ is indeed independent across workers.

We then obtain the marginal posterior probabilities as:

$$\begin{aligned}\zeta_{ik}(s) &= \mathbb{P}_{R^w}\left(Z_i^w(s)=k\right)\\ &= \frac{\sum_{Z_i^w : Z_i^w(s)=k} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right)}{\sum_{Z_i^w} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right)}\end{aligned}$$

and

$$\begin{aligned}\xi_{ikk'}(s+1) &= \mathbb{P}_{R^w}\left(Z_{ik}^w(s)=1, Z_{ik'}^w(s+1)=1\right)\\ &= \frac{\sum_{Z_i^w : Z_i^w(s)=k, Z_i^w(s+1)=k'} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right)}{\sum_{Z_i^w} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right)},\end{aligned}$$

where

$$\begin{aligned}\ln \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w\right) &:= \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i\left(X_i, Z_i^w \mid Z^f\right)\\ &= \sum_{k=1}^{K} Z_{ik}^w(1) \ln\left[\alpha_k(1)\beta_k(1|\tau)\right]\\ &+ \sum_{s=2}^{S_i-1}\sum_{k=1}^{K} Z_{ik}^w(s-1)\sum_{k'=1}^{K} Z_{ik'}^w(s)\ln\alpha_{kk'}(s|\tau) + \sum_{s=2}^{S_i}\sum_{k=1}^{K} Z_{ik}^w(s)\ln\beta_k(s|\tau)\\ &= \ln\left[\alpha_{k_1}(1)\beta_{k_1}(1|\tau)\alpha_{k_1k_2}(2|\tau)\beta_{k_2}(2|\tau)\alpha_{k_2k_3}(3|\tau)\beta_{k_3}(3|\tau)...\right],\end{aligned}$$

with $Z_i^w = (k_1, k_2, k_3, ...)$. The fact that Z^f has been integrated out does not alter the structure of this likelihood. It is like the complete likelihood, with Z^f replaced by τ in the definition of the α, β state and emission probabilities.

The Forward-Backward algorithm can therefore applied to this non standard context, yielding the following simple formulas for the posterior worker type probabilities:

$$\zeta_{ik}(s) = \frac{F_{ik}(s)B_{ik}(s)}{L_i}, \quad \xi_{ikk'}(s+1) = \frac{F_{ik}(s)\alpha_{kk'}(s+1|\tau)\beta_{k'}(s+1|\tau)B_{ik'}(s+1)}{L_i},$$

where $F_{ik}(s)$ and $B_{ik}(s)$ are the Forward and Backward operators (defined just below) and with

$$L_i = \sum_{k=1}^{K} F_{ik}(s)B_{ik}(s).$$

The Forward algorithm. The forward step defines the variables,

$$
\begin{aligned}
F_{ik}(s) &= \sum_{Z_i^w(1:s-1)} \exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i \left(X_i(1:s), Z_i^w(1:s-1), Z_i^w(s) = k \mid Z^f \right) \\
&= \sum_{Z_i^w(1:s-1)} \overline{\mathcal{L}}_{i,\tau} \left(X_i(1:s), Z_i^w(1:s-1), Z_i^w(s) = k \right).
\end{aligned}
$$

Up to the averaged out firm types, this is the likelihood of the observed emissions for periods 1 to s, and of the worker being in state k at s. The forward variables can be calculated iteratively as

$$
\begin{aligned}
F_{ik}(1) &= \alpha_k(1)\beta_k(1), \\
F_{ik}(s) &= \sum_{\kappa=1}^{K} F_{i\kappa}(s-1)\alpha_{\kappa k}(s|\tau)\beta_\kappa(s|\tau), \quad s = 2, ..., S_i.
\end{aligned}
$$

Remark. For S_i large enough, the multiplication of many values lower than one can run into numerical underflow problems. To deal with this, we can use the "log-sum-exp" trick. First, we use a log coding:

$$
\ln F_{ik'}(s) = \ln \left(\sum_{k=1}^{K} \exp \left[\ln F_{ik}(s-1) + \ln \alpha_{kk'}(s) + \ln \beta_{k'}(s) \right] \right).
$$

Then, to avoid numerical underflows, we write the first log as

$$
\ln \sum_{k=1}^{K} e^{x_k} = c + \ln \sum_{k=1}^{K} e^{x_k - c}, \quad c = \max_k x_k.
$$

This way, we make sure that $e^{x_k - c} = 1$ for some $k = 1, ..., K$, and therefore that the log is positive. Alternatively, we can use the method in Rabiner (1989).

The Backward algorithm. The backward step defines, for $s = 1, ..., S_i - 1$,

$$
\begin{aligned}
B_{ik}(s) &= \sum_{Z_i^w(s+1:S_i)} \exp \sum_{Z^f} R^f(Z^f) \ln \mathcal{L}_i \left(X_i(s+1:S_i), Z_i^w(s+1:S_i) \mid Z^f, Z_i^w(s) = k \right) \\
&= \sum_{Z_i^w(s+1:S_i)} \overline{\mathcal{L}}_{i,\tau} \left(X_i(s+1:S_i), Z_i^w(s+1:S_i) \mid Z_i^w(s) = k \right).
\end{aligned}
$$

This is the likelihood of the emissions after $s+1$ given the worker type at s. Here again, the backward variables can be calculated iteratively. For

$$\begin{aligned}
B_{ik}(s) &= \sum_{k'=1}^{K} \alpha_{kk'}(s+1|\tau)\beta_{k'}(s+1|\tau)B_{ik'}(s+1),\ s = 1, ..., S_i - 1, \\
B_{ik}(S_i) &= 1.
\end{aligned}$$

We finally obtain, for $s = 1, ..., S_i$,

$$\sum_{Z_i^w(1:s-1), Z_i^w(s+1:S_i)} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w(1:s-1), Z_i^w(s)=k, Z_i^w(s+1:S_i)\right) = F_{ik}(s)B_{ik}(s),$$

and for $s = 1, ..., S_i - 1$,

$$\begin{aligned}
\sum_{Z_i^w(1:s-1), Z_i^w(s+2:S_i)} \overline{\mathcal{L}}_{i,\tau}\left(X_i, Z_i^w(1:s-1), Z_i^w(s)=k, Z_i^w(s+1)=k', Z_i^w(s+1:S_i)\right)& \\
= F_{ik}(s)\alpha_{kk'}(s+1|\tau)\beta_{k'}(s+1|\tau)B_{ik'}(s+1).&
\end{aligned}$$

C.4 Firm E-step

Here we directly solve the problem

$$\max_{R^f} \sum_{Z^f} R^f(Z^f) \sum_{Z^w} R^w(Z^w) \ln \mathcal{L}(X, Z^f, Z^w) - \sum_{Z^f} R^f(Z^f) \ln R^f(Z^f),$$

subject to

$$R^f(Z^f) = \prod_{j=1}^{J}\prod_{\ell=1}^{L} \tau_{j\ell}^{Z^f_{j\ell}}, \quad \sum_{\ell=1}^{L} \tau_{j\ell} = 1.$$

For workers, this direct approach was not as straightforward because $R_i^w(Z_i^w)$ did not factorize as the product of marginals for $Z_i^w(s)$. With the independence restriction

$$\sum_{Z^f} R^f(Z^f) \ln R^f(Z^f) = \sum_{j=1}^{J}\sum_{\ell=1}^{L} \tau_{j\ell} \ln \tau_{j\ell}.$$

The first-order condition for $\tau_{j\ell}$ ($j, \ell \neq 0$) is easily obtained from the expected log-likelihood as

$$g_{j\ell}(\tau) - 1 - \ln \tau_{j\ell} + \ln \lambda_j = 0,$$

where λ_j is determined by the constraint $\sum_{\ell=1}^{L} \tau_{j\ell} = 1$ and where

$$
\begin{aligned}
g_{j\ell}(\tau) = \ln \pi_\ell^f &+ \sum_{i=1}^{I} \sum_{k=1}^{K} \zeta_{ik}(1) Y_{ij}(1) \left[\ln m_{k\ell} + \ln \theta_{j\ell} \right] \\
&+ \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \sum_{k=1}^{K} \sum_{k'=1}^{K} \xi_{ikk'}(s) Y_{ij}(s) \ln A_{k\ell k'} \\
&+ \sum_{i=1}^{I} \sum_{s=1}^{S_i} \sum_{k=1}^{K} \zeta_{ik}(s) Y_{ij}(s) \ln \left[\varphi \left(T\left[W_i(s) \right], \eta_{k\ell} \right) + D_i(s) \ln M_{k\ell}^{-} \right] \\
&+ \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \sum_{k=1}^{K} \zeta_{ik}(s) Y_{ij}(s) \sum_{j'=0}^{J} \mathbf{1}\{j' \neq j\} Y_{ij'}(s+1) \sum_{\ell'=0}^{L} \tau_{j'\ell'} \left[\ln M_{k\ell\ell'} + \ln \theta_{j'\ell'} \right] \\
&+ \sum_{i=1}^{I} \sum_{s=1}^{S_i-1} \sum_{k=1}^{K} \zeta_{ik}(s) \sum_{j'=0}^{L} \mathbf{1}\{j' \neq j\} Y_{ij'}(s) \sum_{\ell'=0}^{L} \tau_{j'\ell'} Y_{ij}(s+1) \left[\ln M_{k\ell'\ell} + \ln \theta_{j\ell} \right].
\end{aligned}
$$

We shall calculate

$$
\tau_{j\ell} = \frac{e^{g_{j\ell}(\tau)}}{\sum_{\ell'} e^{g_{j\ell'}(\tau)}}.
$$

This is a sparse non-linear equation system in $\tau_{j\ell}$. While sparse, it is nevertheless a sizeable problem. In the practical implementation, we employ a speed improvement by not solving the system fully in the early iterations of the VEM algorithm. Specifically, for a given guess of τ^m, it is straightforward to determine $g_{j\ell}(\tau^m)$ and with it an updated firm prior τ^{m+1}. In practice, iteration produces a convergence towards a fixed point, and as such this simple iteration is an alternative to a more targeted non-linear sparse root solver. As a full solver, it is likely the case that the iterative scheme is dominated by other methods, but the iterative scheme offers a simple way of obtaining fast partial improvements in the firm prior. Thus, in early steps of the VEM algorithm, we do only a few iterations on the τ iterative solver, and only insist on a full solution as the VEM algorithm is getting close to meeting its stopping criterion.

D Further results (online)

D.1 Fit of wage growth

We can y calculate the cohort by age unconditional change in wage levels, $\mathbb{E}\left[\Delta w_{\tau_a,\zeta}\right]$, and the mean wage growth given a job change during the year, $\mathbb{E}_{JJ} \Delta w_{\tau_a,\zeta}$. Empirical variances, $\mathbb{V}\, \Delta w_{\tau_a,\zeta}$ and $\mathbb{V}_{JJ}\, \Delta w_{\tau_a,\zeta}$ can be similarly calculated.

The distribution of $\Delta w_{\tau_a,\zeta} := w_{\tau_a,\zeta} - w_{\tau_a-1,\zeta}$ is theoretically induced by the distribution of $(k_{\tau_a-1}, \ell_{\tau_a-1})$ and $(k_{\tau_a}, \ell_{\tau_a})$. We obtain the unconditional distribution of $(k_{\tau_a-1}, \ell_{\tau_a-1}, k_{\tau_a}, \ell_{\tau_a})$ as $\operatorname{diag}(p_{\tau_s(\tau_a+\zeta,\zeta)-1,\zeta})Q$, and that conditional on a job change ($JJ = 1$) as $\operatorname{diag}(p_{\tau_s(\tau_a+\zeta,\zeta)-1,\zeta})\overline{Q}_1$. Then, for example,

$$\mathbb{E}_{JJ}\Delta w_{\tau_a,\zeta} = \sum_{k,\ell} p_{\tau_s(\tau_a+\zeta,\zeta),\zeta}(k,\ell) \sum_{k',\ell'} Q_1(k',\ell' \mid k,\ell)\,(\mu_{k'\ell'} - \mu_{k\ell})$$

and

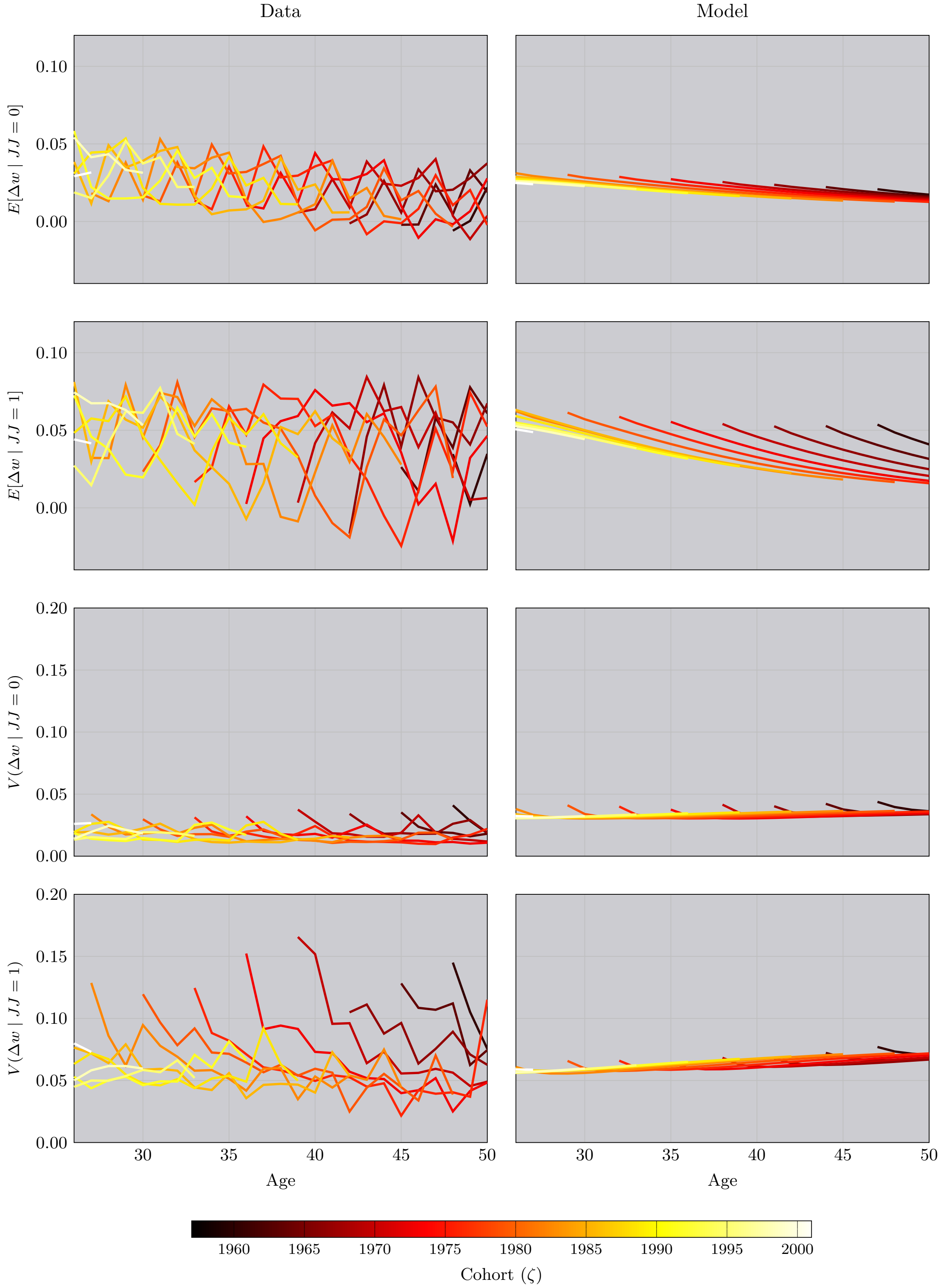
$$\mathbb{E}_{JJ}\Delta w^2_{\tau_a,\zeta} = \sum_{k,\ell} p_{\tau_s(\tau_a+\zeta,\zeta),\zeta}(k,\ell) \sum_{k',\ell'} Q_1(k',\ell' \mid k,\ell)\left[(\mu_{k'\ell'} - \mu_{k\ell})^2 + \sigma^2_{k'\ell'} + \sigma^2_{k\ell}\right].$$

Figure D.1 shows the fit of wage change averages and variances by cohort and age. Data are in the left column and model estimates in the right column. The variation in the data is noisy and to ease exposition, we show only every third cohort.

The first and second rows show the average wage change uncondtionally and conditional on a job-to-job change, respectively. Wage growth is generally positive and declining with age. Furthermore, wage growth conditional on a job-to-job change is greater than the unconditional growth. The model estimate fit the wage growth levels well as well as the decline with age. There is evidence of aggregate calendar time effects in the wage growth rates in the data which when viewed by age manifest in different cohorts at different ages corresponding to the same calendar time. The model only allows calendar time variation in its parameters through the cohort initial conditions. Consequently, the noise in the data age profiles is absent in the model estimates. To the extent that the calendar time shocks are temporary, this may be considered a virtue.

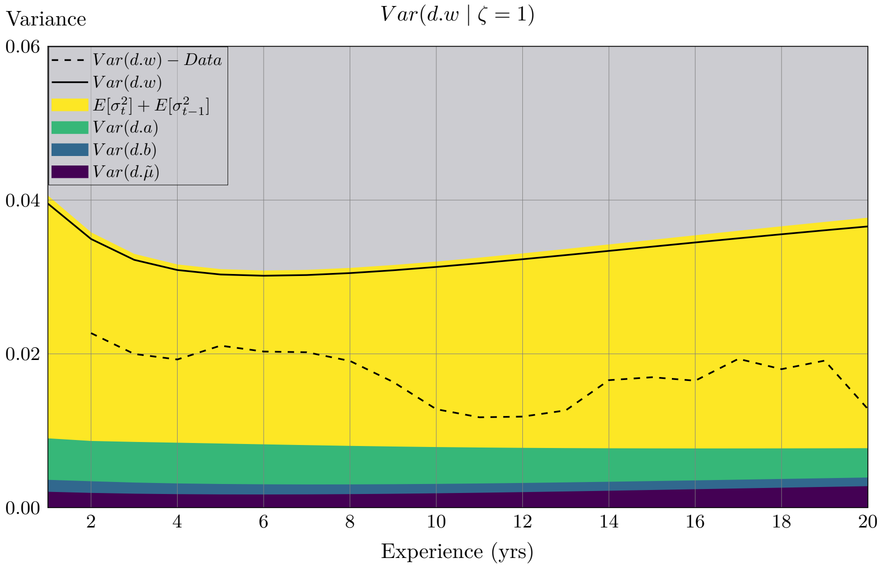
The third and fourth rows show the wage change variance unconditionally and job-to-job change conditionally. The wage change variance is decreasing in experience for the $\zeta < 1982$ cohorts, and is broadly greater if conditional on a job change. The model reproduces these patterns, but generally over-estimates the unconditional variance while being more or less at level with the empirical job change conditional wage change variance.

Figure D.1: Cohort age profile model fit: Wage growth.



Notes: Only every third cohort is drawn.

Figure D.2: Var(Δw_τ) decomposition for $\zeta = 1982$.



D.2 Yearly wage growth

Denote by Δ the one year change operator. The wage change variance can be decomposed as

$$\begin{aligned}
\mathbb{V}\Delta w_\tau &= \mathbb{E}\sigma_\tau^2 + \mathbb{E}\sigma_{\tau-1}^2 + \mathbb{V}(\Delta a_\tau + \Delta b_\tau + \Delta\widetilde{\mu}_\tau) \\
&= \mathbb{E}\sigma_\tau^2 + \mathbb{E}\sigma_{\tau-1}^2 + \mathbb{V}\Delta a_\tau + \mathbb{V}\Delta b_\tau + \mathbb{V}\Delta\widetilde{\mu}_\tau \\
&+ 2\mathrm{Cov}\left(\Delta a_\tau, \Delta b_\tau\right) + 2\mathrm{Cov}\left(\Delta a_\tau, \Delta\widetilde{\mu}_\tau\right) + 2\mathrm{Cov}\left(\Delta b_\tau, \Delta\widetilde{\mu}_\tau\right).
\end{aligned}$$

The decomposition allows us to compare the relative strength of changes to the wage projection components.

Figure D.2 shows the decomposition for the $\zeta = 1982$ cohort. The figure does not include the covariance terms. They turn out to be negligible, which can be seen by $\mathbb{V}\,\Delta w_\tau$ being fully explained by the remaining components. The decomposition reveals that $\mathbb{V}\,\Delta a_\tau$ is about 3-4 times as large as $\mathbb{V}\,\Delta b_\tau$. This is another demonstration of the importance of worker type dynamics to the overall wage process.

Note also that the predicted $\mathbb{V}\,\Delta w_\tau$ is about twice as large as the one in the data. This suggests that a better model would allow for autoregressive residuals. Suppose for example that the residual be

$$u_\tau = \epsilon_\tau + \phi\epsilon_{\tau-1},$$

with $\mathbb{V}(\epsilon_\tau | k, \ell) = \sigma_{k\ell}^2$ and $0 < \phi < 1$. Then there would be a negative covariance between u_τ and $u_{\tau-1}$ in $\mathbb{V}\,\Delta u_\tau$. We leave the exploration of a moving average specification for the error term (or any other autoregressive specification) to further studies, as the identification of the HMM with autoregressive residuals is not proven, and as the Baum-Welch algorithm would need to be adapted.

Table D.1: First-difference variance decomposition (shares of $\mathrm{Var}(\Delta w)$)

	HMM simulation		CD via μ	CD plugin	AKM	KSS
	level	share				
$\mathrm{Var}(\alpha)$	0.004	0.129	0.046	0.087	0.209	-0.023
$\mathrm{Var}(\psi)$	0.001	0.037	0.001	0.023	0.209	0.097
$2\mathrm{Cov}(\alpha,\psi)$	0.000	0.001	0.000	-0.028	-0.285	-0.115
$\mathrm{Var}(\tilde{\mu})$	0.002	0.069	0.023			
$2\mathrm{Cov}(\alpha,\tilde{\mu})$	-0.001	-0.019	0.000			
$2\mathrm{Cov}(\psi,\tilde{\mu})$	-0.000	-0.006	0.000			
$\mathbb{E}[\sigma_\tau^2]+\mathbb{E}[\sigma_{\tau-1}^2]$	0.027	0.788	0.931			
Residual				0.918	0.867	1.041
$\mathrm{Var}(\Delta w)$	0.034		0.023	0.023	0.023	0.023
$\mathrm{Corr}(\alpha,\psi)$	0.009		0.012	-0.309	-0.681	

Note: In columns 1–3 (HMM and CD via μ), α and ψ denote the year-to-year change components Δa_τ and Δb_τ of the $\mathrm{Var}(\Delta w_\tau)$ identity (Fig. 13). In columns 4–6 (CD plugin, AKM, KSS), α and ψ denote the FD-AKM worker fixed effect (worker-specific average wage growth) and current-firm contribution to growth. A negative KSS variance estimate arises when bias correction exceeds the plug-in estimate.

Table D.1 reports the seven-component decomposition of $\mathbb{V}\,\Delta w_\tau$ from Figure D.2 pooled over the full sample. The HMM column reveals two facts. First, residual within-cell noise dominates wage changes: $\mathbb{E}[\sigma_\tau^2]+\mathbb{E}[\sigma_{\tau-1}^2]$ alone accounts for 78.7% of $\mathbb{V}\,\Delta w_\tau$. Compared to the levels decomposition in Table 1, where the worker-type effect $\mathbb{V}\,a$ drove 60% of $\mathbb{V}\,w$, at the year-to-year change frequency types are persistent enough that almost all variation arises from within-cell noise rather than type transitions. Second, worker-type dynamics dominate firm-type dynamics: $\mathbb{V}\,\Delta a_\tau$ accounts for 13.0% of $\mathbb{V}\,\Delta w_\tau$ versus $\mathbb{V}\,\Delta b_\tau$ at 3.7%, mirroring the levels finding. The completed-data column via μ projection shows the same qualitative pattern with the residual share pushed even higher (93.1%) because hard Viterbi classification removes type-trajectory stochasticity.

If we turn to the conventional first-difference wage fixed effect regressions the residual contributions come to dominate the variance decomposition at well over .85 regardless of the approach, be it whether it is done on the completed data (column 3) or by AKM and KSS. AKM is biased upwards in its variance contribution assignment to both worker and firm types. Generally, the first-differences wage fixed effect regressions do not provide strong insights and we include them here primarily as a negative result, that the non-linearities in the first differences overwhelm insights from this approach. In the following sections we turn attention to worker wage type dynamics as a more productive focus for the study of growth dynamics.