Wage Decomposition

Paolo Zacchia

Microeconometrics

Lecture 17

Wages and labor market outcomes

- The career of workers over their lifetimes can be "measured" over multiple dimensions of interest: earnings, participation, unemployment status conditional on participation, *et cetera*.

- Economists typically called these dimensions "labor market outcomes." Many research questions are about the empirical search for the determinants of such outcomes.

- Labor market outcomes follow from individual choices and as such they typically depend on one another. For example, participation, labor supply and wages are intertwined in the classical selection model by Heckman (1979, see Lecture 11).

- Most of this lecture centers on econometric methods for the **"decomposition"** of **wages** between different contributing factors. Later, this lecture overviews the connection between these methods and the study of other economic outcomes.

Wage decomposition: introduction

- Economic theory predicts that wages are the result of a labor market equilibrium where labor demand and supply meet.

- As per the labor demand function, wages equal the marginal **productivity** of workers in their employing firms.

- Yet wages are notoriously **unequal** and wage inequality has been soaring in many countries over time. Why?

- Econometric methods for the decomposition of wages aim to divide the grand **variance** of (logarithmic) wages into some meaningful components: mainly the one that depends on the **workers** and the one that depends on **firms** – and more.

- This literature is known through the acronym "AKM" from the seminal paper by Abowd, Kramarz and Margolis (1999) and it entails methodological challenges of various sort.

A model for the decomposition of wages (1/10)

The regression equation introduced by AKM is as follows:

$$\log W_{it} = \alpha_i + \psi_{j(i,t)} + \boldsymbol{x}_{it}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{it}$$

where:

- $\log W_{it}$ is the logarithm of worker i's **wage** at time t;
- α_i is a fixed effect specific to **worker** i (for $i = 1, \ldots, N$);
- $\psi_{j(i,t)}$ is a fixed effect specific to **firm** j (for $j = 1, \ldots, J$);
- $j(i,t)$ is a **linkage function** that identifies the employer of worker i at time t;
- $\boldsymbol{x}_{it}$ are some **characteristics** of worker i that are observed at time t, with associated parameters $\boldsymbol{\beta}$;
- ε_{it} is a residual error term.

A model for the decomposition of wages (2/10)

What is the objective of an "AKM" analysis? Consider a simpler Mincer-like wage equation without any fixed effects:

$$\log W_{it} = \boldsymbol{x}_{it}^{\mathrm{T}} \boldsymbol{\beta}^{**} + \varepsilon_{it}$$

this model is known to typically explain about 30% of the grand variance of $\log W_{it}$. What about the remaining 70%?

Suppose that $\boldsymbol{x}_{it} = \mathbf{0}$ and $\varepsilon_{it} = 0$ for all i, t. Then:

$$\mathbb{V}\text{ar}\left[\log W_{it}\right] = \mathbb{V}\text{ar}\left[\alpha_i\right] + \mathbb{V}\text{ar}\left[\psi_{j(i,t)}\right] + 2\,\mathbb{C}\text{ov}\left[\alpha_i, \psi_{j(i,t)}\right]$$

where the first two terms represent the residual contributions of **workers** and **firms**, respectively, to $\mathbb{V}\text{ar}\left[\log W_{it}\right]$.

The third term instead – the covariance – is typically interpreted as the contribution of **sorting** ("high-wage workers at high-wage firms," and vice versa) occurring in the labor market.

A model for the decomposition of wages (3/10)

Rewrite the AKM model in matrix notation:

$$\mathbf{y} = \mathbf{D}\boldsymbol{\alpha} + \mathbf{F}\boldsymbol{\psi} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

where:

- $\mathbf{y}$ is the vector that stacks the observations of $\log W_{it}$;
- $\mathbf{D}$ is a matrix of **worker** dummy variables;
- $\mathbf{F}$ is a matrix of **firm** dummy variables, obtained via $j(i,t)$;
- $\mathbf{X}$ is the matrix that stacks the observations of $\boldsymbol{x}_{it}^{\mathrm{T}}$;
- $\boldsymbol{\varepsilon}$ is the vector that stacks the error terms ε_{it};
- $\boldsymbol{\alpha}$ is the collection of worker fixed effects α_i;
- $\boldsymbol{\psi}$ is the collection of firm fixed effects $\psi_{j(i,t)}$.

A model for the decomposition of wages (4/10)

Write $\boldsymbol{\theta} \equiv (\boldsymbol{\alpha}, \boldsymbol{\psi}, \boldsymbol{\beta})$. Interest typically falls on the estimation of the variance-covariance matrix of the model's parameters.

$$\mathbb{V}\text{ar}\left[\boldsymbol{\theta}\right] = \begin{bmatrix} \mathbb{V}\text{ar}\left[\boldsymbol{\alpha}\right] & \mathbb{C}\text{ov}\left[\boldsymbol{\alpha}, \boldsymbol{\psi}\right] & \mathbb{C}\text{ov}\left[\boldsymbol{\alpha}, \boldsymbol{\beta}\right] \\ \mathbb{C}\text{ov}\left[\boldsymbol{\psi}, \boldsymbol{\alpha}\right] & \mathbb{V}\text{ar}\left[\boldsymbol{\psi}\right] & \mathbb{C}\text{ov}\left[\boldsymbol{\psi}, \boldsymbol{\beta}\right] \\ \mathbb{C}\text{ov}\left[\boldsymbol{\beta}, \boldsymbol{\alpha}\right] & \mathbb{C}\text{ov}\left[\boldsymbol{\beta}, \boldsymbol{\psi}\right] & \mathbb{V}\text{ar}\left[\boldsymbol{\beta}\right] \end{bmatrix}$$

This can be manipulated so as to derive economically meaningful quantities. Occasionally, a secondary goal is to explain $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ as a function of time-invariant characteristics of workers and firms (this is performed e.g. via regressions for estimates $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\psi}}$).

Estimation is performed with **"matched employer-employee"** (MEE) data, typically obtained from administrative sources, that allow to construct the linkage function $j(i, t)$ and matrix $\mathbf{F}$. MEE datasets are usually very large, and so is the resulting dimension of $\boldsymbol{\theta}$: this leads to computational challenges.

A model for the decomposition of wages (5/10)

One reason that makes the AKM model popular is its generality. Suppose that the term $\mathbf{F}\boldsymbol{\psi}$ is omitted from the model. Then, the model becomes a "traditional" wage equation for panel data.

$$\log W_{it} = \alpha_i^* + \boldsymbol{x}_{it}^{\mathrm{T}} \boldsymbol{\beta}^* + \varepsilon_{it}$$

Let $\boldsymbol{\alpha}^*$ be the collection of worker fixed effects α_i^*. It follows that by the algebra of the linear model, *conditional on* $\mathbf{X}$, it is:

$$\boldsymbol{\alpha}^* = \boldsymbol{\alpha} + \left(\mathbf{D}^{\mathrm{T}} \mathbf{M}_{\mathbf{X}} \mathbf{D}\right)^{-1} \mathbf{D}^{\mathrm{T}} \mathbf{M}_{\mathbf{X}} \mathbf{F} \boldsymbol{\psi}$$

where $\mathbf{M}_{\mathbf{X}} = \mathbf{I} - \mathbf{X}\left(\mathbf{X}^{\mathrm{T}}\mathbf{X}\right)^{-1}\mathbf{X}^{\mathrm{T}}$. If $\mathbf{X}$ is orthogonal to both $\mathbf{D}$ and $\mathbf{F}$ this is the expression of an omitted variable bias, with:

$$\alpha_i^* = \alpha_i + \frac{1}{T} \sum_{t=1}^{T} \psi_{j(i,t)}$$

and straightforward interpretation. This kind of analysis is easily extended to $\boldsymbol{\beta}^*$, to the case where $\mathbf{D}\boldsymbol{\alpha}$ is omitted, *et cetera.*

A model for the decomposition of wages (6/10)

Identification of the model rests on two key **assumptions**: one is **statistical**, and the other one is a typical **rank** condition.

1. The statistical assumption is a seemingly standard condition about mean-independence of the error term.

   $$\mathbb{E}\left[\boldsymbol{\varepsilon}\middle|\,\mathbf{D},\mathbf{F},\mathbf{X}\right]=\mathbf{0}$$

   This is called the **"exogenous mobility"** assumption due to a key implication about conditioning on $\mathbf{F}$: the error term must be uncorrelated to changes in $j\left(i,t\right)$ (omitted variables have no bearing on the movement of workers across firms). The implications are elaborated by Card, Heining and Kline (2013) as it is discussed later.

Note that in the original paper by AKM, the error term ε_{it} was assumed for simplicity to be uncorrelated across workers and over time. However, more general structures can be allowed.

A model for the decomposition of wages (7/10)

2. The rank condition demands as usual that the cross-product of the full "design matrix" has full rank. Such a matrix is:

$$\mathbf{W} = \begin{bmatrix} \mathbf{D}^{\mathrm{T}}\mathbf{D} & \mathbf{D}^{\mathrm{T}}\mathbf{F} & \mathbf{D}^{\mathrm{T}}\mathbf{X} \\ \mathbf{F}^{\mathrm{T}}\mathbf{D} & \mathbf{F}^{\mathrm{T}}\mathbf{F} & \mathbf{F}^{\mathrm{T}}\mathbf{X} \\ \mathbf{X}^{\mathrm{T}}\mathbf{D} & \mathbf{X}^{\mathrm{T}}\mathbf{F} & \mathbf{X}^{\mathrm{T}}\mathbf{X} \end{bmatrix}$$

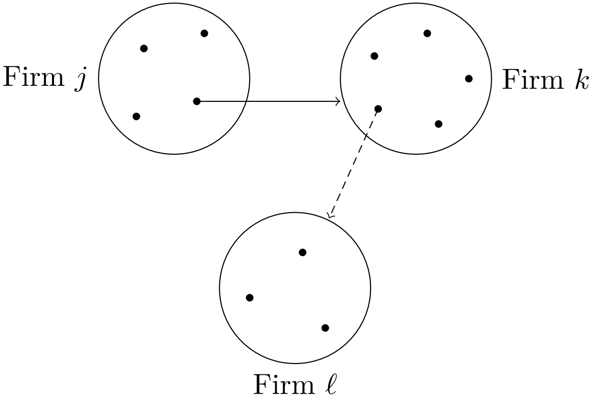
   and the critical condition is that $\mathbf{F}^{\mathrm{T}}\mathbf{D}$ (or $\mathbf{D}^{\mathrm{T}}\mathbf{F}$) has full row (or column) rank. In actual MEE datasets this is inevitably violated for some workers and firms, which must be therefore expunged from the estimation sample.

   Intuitively, if all the workers i of a firm j never leave it their own effects α_i are perfectly collinear with the firm effect ψ_j. Thus, identification is based on workers who are **movers**.

In AKM parlance, the estimation sample is called the **connected set**. More precisely, it is the *connected component of a "bipartite" network* defined by the linkage function $j(i,t)$.

A model for the decomposition of wages (8/10)

The intuition behind the movers-based algebraic identification is best illustrated graphically. In the figure, circles represent firms, dots are workers at $t = 0$, and arrows denote later movements.



Adding movements delivers a connected set that includes firms j and k (straight arrow), and/or k and ℓ (dashed arrow).

A model for the decomposition of wages (9/10)

The dimension of $\mathbf{W}$ makes inverting it generally **impractical**. How to estimate the model, then? The usual approach (although alternatives exist) proceeds in **two steps**.

1. First, estimate a **first-differenced** model for **movers only**:

$$\log W_{it} - \log W_{i(t-1)} =$$
$$= \psi_{j(i,t)} - \psi_{j(i,t-1)} + \left(\boldsymbol{x}_{it} - \boldsymbol{x}_{i(t-1)}\right)^{\mathrm{T}} \boldsymbol{\beta} + \left(\varepsilon_{it} - \varepsilon_{(t-1)}\right)$$

   for suitable observation pairs such that $j(i,t) \neq j(i,t-1)$. In some MEE datasets, this can still be quite expensive.

2. Then, recover worker effects using the **full connected set**:

$$\widehat{\alpha}_i = \frac{1}{T_i} \sum_{t=t_{0i}}^{T_i} \left(\log W_{it} - \widehat{\psi}_{j(i,t-1)} - \boldsymbol{x}_{it}^{\mathrm{T}} \widehat{\boldsymbol{\beta}}\right)$$

   where T_i is the total number of time periods (starting from t_{0i}) that worker i appears in the connected set.

A model for the decomposition of wages (10/10)

- The actual estimation performed in the original AKM paper also involved some additional "auxiliary" variables $\mathbf{Z}$ aimed at proxying for the correlation between $\mathbf{D}$ or $\mathbf{X}$, and $\mathbf{F}$.

- This allows AKM to simplify the estimation of $\boldsymbol{\theta}$ and $\mathbb{V}\text{ar}\left[\boldsymbol{\theta}\right]$ via two alternative approaches (the "order-dependent" and the "order-independent" ones) motivated on the assumptions $\mathbf{X}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{F} = \mathbf{0}$ and $\mathbf{D}^{\mathrm{T}}\mathbf{M}_{\mathbf{Z}}\mathbf{F} = \mathbf{0}$ ($\mathbf{M}_{\mathbf{Z}}$ is the residual-maker matrix for $\mathbf{Z}$). This solution is less common nowadays.

- AKM estimated their model on the French MEE data from the *Déclarations Annuelles des Salaires* from 1976 to 1987, featuring millions of observations (many of which unusable).

- They reported a rich set of results demonstrating the power of their empirical framework to answer many questions about wages and the labor market.

Questioning the assumptions (1/2)

- Following AKM, a large literature blossomed, as researchers began to apply their framework on different MEE datasets.

- For about a decade, some consistent results emerged:

$$\frac{\mathbb{V}\text{ar}\left[\psi_{j(i,t)}\right]}{\mathbb{V}\text{ar}\left[\alpha_i + \psi_{j(i,t)}\right]} \approx 0.25$$

  that is the variance explained by firm effects is about 20-30% of the total variance explained by fixed effects, and:

$$\mathbb{C}\text{ov}\left[\alpha_i, \psi_{j(i,t)}\right] \leq 0$$

  that is, the covariance between the two types of fixed effects (*sorting*) is zero if not negative.

- These results, especially the non-positive covariance/sorting, were often considered surprising if not unrealistic.

Questioning the assumptions (2/2)

The initial "AKM" literature thus led researchers to question the key identifying assumptions, or their implications.

- The article by Card, Heining and Kline (2013; CHK) took up the task of **testing exogenous mobility**, finding that the violation of it is *not* a concern in typical MEE datasets.

- In two "critical" papers, Andrews, Gill, Schank and Upward (2008, 2012; AGSU) showed that the rank condition is **not** sufficient to guarantee consistent estimation of $\mathbb{V}\mathrm{ar}[\boldsymbol{\theta}]$, even if $\boldsymbol{\theta}$ is consistently estimated, when there are *few movers*.

- AGSU empirically showed that this **limited mobility bias** is sizable: this might explain the puzzling results.

- The subsequent methodological literature, which is reviewed later, addressed this problem directly.

Testing exogenous mobility (1/5)

- The paper by CHK is notable mainly for its test of exogenous mobility. It has revitalized the "AKM" literature by proving its potential to answer important real-world questions.

- It is useful to understand the setting studied by CHK: (West) Germany, where wage inequality has risen dramatically since the nineties, as documented by CHK themselves.

- A pivotal time in the history of German industrial relations was 2003-2005, when the "Hartz" reforms were introduced; these gave incentives to firms for mostly part-time, low-wage job contracts and in addition cut unemployment benefits.

- CHK set out to study changes over time in the composition of the variance of log-wages under the AKM framework. To this end, they used MEE data from 1985 to 2009 maintained by the *Institut für Arbeitsmarkt- und Berufsforschung*.

Testing exogenous mobility (2/5)

CHK provide a comprehensive discussion about the mechanisms that can lead to endogenous mobility, and in particular to:

$$\mathbb{E}\left[\mathbf{F}\boldsymbol{\varepsilon}\right] \neq \mathbf{0}$$

that is, a correlation between firm indicators and the unobserved component of log-wages. They are listed as follows.

1. **Matching effects** (the most important threat): over time, workers might move to firms that are more "suited" to them (and vice versa). In short, sorting occurs on the unobserved transitory shock ε_{it}, and is not limited to α_i and $\psi_{j(i,t)}$.
2. **Drift** (a unit-root component of ε_{it}): if ε_{it} is persistent (e.g. because of human capital accumulation of some kind) this might predict the type of future employers.
3. **Fluctuations** (a minor problem): seasonal variations of ε_{it} might correlate with mobility patterns for some reason.

Testing exogenous mobility (3/5)

CHK suggest that if mobility is endogenous, the effect of moving between different types of firms (say, having effects ψ_1 and ψ_2) should be **asymmetric**. If a worker moves in one direction:

$$\mathbb{E}\left[\log W_{it} - \log W_{i(t-1)} \middle| j\left(i,t\right) = 2, j\left(i,t-1\right) = 1\right] = \\ = \psi_2 - \psi_1 + \mathbb{E}\left[\varepsilon_{it} - \varepsilon_{i(t-1)} \middle| j\left(i,t\right) = 2, j\left(i,t-1\right) = 1\right]$$

while if the same worker moves in the opposite direction:

$$\mathbb{E}\left[\log W_{it} - \log W_{i(t-1)} \middle| j\left(i,t\right) = 1, j\left(i,t-1\right) = 2\right] = \\ = \psi_1 - \psi_2 + \mathbb{E}\left[\varepsilon_{it} - \varepsilon_{i(t-1)} \middle| j\left(i,t\right) = 1, j\left(i,t-1\right) = 2\right]$$

where the “bias” terms on the right hand side are expected to be **both positive** (for example, because of matching effects).

Without the bias terms, the effect of moving is **symmetric**: here, it would be $\psi_2 - \psi_1$ and $\psi_1 - \psi_2$, respectively.

Testing exogenous mobility (4/5)

- CHK suggest to perform **event studies** about the effect on log-wages of the **movement** of workers between firms that belong to different percentiles of the distribution of the firm effects ψ_j (as estimated by the AKM model).

- Specifically, CHK worked with **quartiles** of ψ_j, examining movements from the top (fourth) quartile to the first, second and third, from the first to the other three, and so on.

- Movements between any pair of quartiles, but in the opposite direction, are strikingly **symmetric** (this is especially true for the two extreme quartiles).

- This suggests that exogenous mobility is a good assumption to maintain, and is not the driver of the unrealistic results.

- This finding has been later replicated in other MEE datasets.

Testing exogenous mobility (5/5)

- CHK also provide another useful test: if endogenous mobility occurs, the AKM **residuals** should be on average non-zero for selected combinations of workers (e.g. with high α_i) and firms (e.g. with high ψ_j). This can be tested!

- CHK calculated the average residuals over a grid defined by **deciles** of the estimated effects of the two types. They found large deviations from zero only for cells where either workers or firms from the respective bottom decile appear.

- Then CHK perform the variance decomposition separately for different intervals of the recent German history.

- They found that in 1985-1991, the results are "traditional," but in 2002-2009, the contribution of $\mathbb{V}\text{ar}\left[\psi_j\right]$ rises and that of $2\,\mathbb{C}\text{ov}\left[\alpha_i, \psi_j\right]$ is positive and close to it! Overall, "sorting" effects explain about 34% of the increase in wage inequality.

The limited mobility bias (1/5)

How are the AKM variance components of interest traditionally estimated? Note that since matrices $\mathbf{D}$ and $\mathbf{F}$ are non-stochastic, in the population (writing j for $j(i,t)$ as a shorthand):

$$\mathbb{V}\text{ar}[\boldsymbol{\alpha}] = \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\alpha}}{N^*} = \sum_{i=1}^{N}\sum_{t=t_{0i}}^{T_i}\frac{(\alpha_i - \mathbb{E}[\alpha_i])^2}{N^*}$$

$$\mathbb{V}\text{ar}[\boldsymbol{\psi}] = \frac{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*} = \sum_{i=1}^{N}\sum_{t=t_{0i}}^{T_i}\frac{(\psi_j - \mathbb{E}[\psi_j])^2}{N^*}$$

$$\mathbb{C}\text{ov}[\boldsymbol{\alpha},\boldsymbol{\psi}] = \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*} = \sum_{i=1}^{N}\sum_{t=t_{0i}}^{T_i}\frac{(\alpha_i - \mathbb{E}[\alpha_i])(\psi_j - \mathbb{E}[\psi_j])}{N^*}$$

where N^* is the total size of the data ($N^* = NT$ if the sample is balanced, $N^* = \sum_{i=1}^{N} T_i$ if the sample is unbalanced) and:

$$\mathbf{A} = \mathbf{I} - \frac{1}{N^*}\boldsymbol{\iota}^{\mathrm{T}}\boldsymbol{\iota}$$

is the “demeaning” matrix of conformable $N^* \times N^*$ size.

The limited mobility bias (2/5)

By the analogy principle, the sample analogues of these moments:

$$\widehat{\mathbb{V}\mathrm{ar}}\left[\boldsymbol{\alpha}\right] = \frac{\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\widehat{\boldsymbol{\alpha}}}{N^*} = \sum_{i=1}^{N}\sum_{t=t_{0i}}^{T_i}\frac{\left(\widehat{\alpha}_i - \overline{\widehat{\alpha}}\right)^2}{N^*}$$

$$\widehat{\mathbb{V}\mathrm{ar}}\left[\boldsymbol{\psi}\right] = \frac{\widehat{\boldsymbol{\psi}}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\widehat{\boldsymbol{\psi}}}{N^*} = \sum_{i=1}^{N}\sum_{t=t_{0i}}^{T_i}\frac{\left(\widehat{\psi}_j - \overline{\widehat{\psi}}\right)^2}{N^*}$$

$$\widehat{\mathbb{C}\mathrm{ov}}\left[\boldsymbol{\alpha},\boldsymbol{\psi}\right] = \frac{\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\widehat{\boldsymbol{\psi}}}{N^*} = \sum_{i=1}^{N}\sum_{t=t_{0i}}^{T_i}\frac{\left(\widehat{\alpha}_i - \overline{\widehat{\alpha}}\right)\left(\widehat{\psi}_j - \overline{\widehat{\psi}}\right)}{N^*}$$

where $\overline{\widehat{\alpha}}$ and $\overline{\widehat{\psi}}$ are the empirical averages of the estimated effects of the two kinds, deliver intuitive estimators of the quantities of interest. These formulae were used by both AKM and CHK.

These estimators are **biased**. This is shown under the simplifying AKM assumption of homoscedasticity: $\mathbb{V}\mathrm{ar}\left[\boldsymbol{\varepsilon}\middle|\,\mathbf{W}\right] = \sigma^2\mathbf{I}$.

The limited mobility bias (3/5)

First, rewrite the model by partialing out $\mathbf{X}$.

$$\mathbf{M_X y} = \mathbf{M_X D}\boldsymbol{\alpha} + \mathbf{M_X F}\boldsymbol{\psi} + \mathbf{M_X}\boldsymbol{\varepsilon}$$

By repeated applications of the Frisch-Waugh-Lovell theorem:

$$\widehat{\boldsymbol{\alpha}} = \left(\mathbf{D}^{\mathrm{T}}\mathbf{Q_F D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{Q_F y}$$

$$\widehat{\boldsymbol{\psi}} = \left(\mathbf{F}^{\mathrm{T}}\mathbf{Q_D F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{Q_D y}$$

where for $\mathbf{G} = \mathbf{D}, \mathbf{F}$, it is as follows.

$$\mathbf{Q_G} \equiv \mathbf{M_X}\left[\mathbf{I} - \mathbf{M_X G}\left(\mathbf{G}^{\mathrm{T}}\mathbf{M_X G}\right)^{-1}\mathbf{G}^{\mathrm{T}}\mathbf{M_X}\right]\mathbf{M_X}$$

By applying the typical decomposition of OLS, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\psi}}$ can be expressed as functions of the true parameters.

$$\widehat{\boldsymbol{\alpha}} = \boldsymbol{\alpha} + \left(\mathbf{D}^{\mathrm{T}}\mathbf{Q_F D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{Q_F}\boldsymbol{\varepsilon}$$

$$\widehat{\boldsymbol{\psi}} = \boldsymbol{\psi} + \left(\mathbf{F}^{\mathrm{T}}\mathbf{Q_D F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{Q_D}\boldsymbol{\varepsilon}$$

The limited mobility bias (4/5)

The expectations of the key estimators are calculated as follows.

$$\mathbb{E}\left[\frac{\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\widehat{\boldsymbol{\alpha}}}{N^*}\right] = \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\boldsymbol{\alpha}}{N^*} + \frac{\sigma^2}{N^*}\mathrm{Tr}\left[\left(\mathbf{D}^{\mathrm{T}}\mathbf{Q}_{\mathbf{F}}\mathbf{D}\right)^{-1}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{D}\right]$$

$$\mathbb{E}\left[\frac{\widehat{\boldsymbol{\psi}}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\widehat{\boldsymbol{\psi}}}{N^*}\right] = \frac{\boldsymbol{\psi}^{\mathrm{T}}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*} + \frac{\sigma^2}{N^*}\mathrm{Tr}\left[\left(\mathbf{F}^{\mathrm{T}}\mathbf{Q}_{\mathbf{D}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\right]$$

$$\mathbb{E}\left[\frac{\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\widehat{\boldsymbol{\psi}}}{N^*}\right] = \frac{\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{D}^{\mathrm{T}}\mathbf{A}\mathbf{F}\boldsymbol{\psi}}{N^*} + \\ + \frac{\sigma^2}{N^*}\mathrm{Tr}\left[\mathbf{Q}_{\mathbf{D}}\mathbf{D}\left(\mathbf{F}^{\mathrm{T}}\mathbf{Q}_{\mathbf{D}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{A}\mathbf{F}\left(\mathbf{F}^{\mathrm{T}}\mathbf{Q}_{\mathbf{D}}\mathbf{F}\right)^{-1}\mathbf{F}^{\mathrm{T}}\mathbf{Q}_{\mathbf{D}}\right]$$

The expression of the covariance bias is formidable. In their 2008 paper, AGSU show that this term is **negative** and it **decreases** in magnitude (towards zero) as the number of **movers** increases.

Instead, the two variance terms are shown to be always positive: they are proportional to traces of semi-definite positive matrices.

The limited mobility bias (5/5)

- Intuitively, the two variance estimators are positively biased because they are themselves based on *uncertain* $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\psi}}$, and the estimation errors sum up quadratically.

- Conversely, the covariance estimator is negatively biased to compensate, as all terms must sum up to the grand variance of the logarithmic wages. Moreover, the lack of many movers introduces noise in the estimation of the covariance: in this case, estimation errors add up negatively.

- As identification of different effects depends on the existence of movers, so does the precision of the estimates. This is akin to the failure of the rank condition in linear models being an extreme, "perfect" case of multicollinearity.

- In their 2012 paper, AGSU empirically show the bias using samples of MEE data with varying proportions of movers.

Correcting the limited mobility bias

The work by AGSU has been key to understand the problem in technically rigorous terms, and spurred the search for **solutions**. Research in this regard has taken three major directions.

1. The statistically "orthodox" approach would be to **correct** for the bias analytically. This is not so straightforward here: the current approach relies on the "leave-out" estimator by Kline, Saggio and Sølvsten (2020; KSS).

2. Alternatively, one can model $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ as **random effects** that are **not** estimated: instead, researchers would estimate their variance components directly. This approach demands restrictive assumptions as in all random effects models.

3. Bonhomme, Lamadon and Manresa (2019; BLM) propose a latent variables framework that allows for both **non-linear** and **dynamic** effects of worker and firm "discrete" types.

Leave-out estimation (1/8)

What follows is an overview of the approach developed by KSS, the one most in line with the original AKM framework.

Their results, however, have more general implications: they aim at consistent estimation of any quadratic form of the kind:

$$\theta = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{C}\boldsymbol{\beta}$$

where $\mathbf{C}$ is some given full-rank, non-stochastic matrix and $\boldsymbol{\beta}$ is the parameter vector of a linear model $Y_i = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_i$. Clearly, $\mathbf{C}$ may be chosen so that θ represents an AKM variance-covariance component, yet the result by KSS transcend this particular case.

In particular:

- if $\mathbf{C}$ is positive semi-definite, θ is a *variance* component;
- if $\mathbf{C}$ is non-definite, θ may be called a *covariance* component.

Leave-out estimation (2/8)

A naïve "plug-in" estimator of θ: $\widehat{\theta}_{PI} = \widehat{\boldsymbol{\beta}}_{OLS}^{\mathrm{T}} \mathbf{C} \widehat{\boldsymbol{\beta}}_{OLS}$, is *biased*:

$$\mathbb{E}\left[\widehat{\theta}_{PI}\right] - \theta = \mathrm{Tr}\left(\mathbf{C}\,\mathbb{V}\mathrm{ar}\left[\widehat{\boldsymbol{\beta}}_{OLS}\right]\right) = \sum_{i=1}^{N} B_{ii}\sigma_i^2$$

where, for $\mathbf{S}_{\boldsymbol{xx}} \equiv \sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}$ and under *heteroscedasticity*:

- $B_{ii} \equiv \mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{C} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i$ is a weight specific to observation i;
- $\sigma_i^2 \equiv \mathbb{E}\left[\varepsilon_i^2 \middle| \boldsymbol{x}_i\right]$ is the variance of observation i's error term.

If the model were *homoscedastic* ($\sigma_i^2 = \sigma^2$ for $i = 1, \ldots, N$), this is easy to *correct* for. A "homoscedasticity only" estimate of θ is:

$$\widehat{\theta}_{HO} = \widehat{\boldsymbol{\beta}}_{OLS}^{\mathrm{T}} \mathbf{C} \widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} \frac{B_{ii}}{N-K} \sum_{i=1}^{N} \left(y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS}\right)^2$$

exploiting the standard unbiased estimator of σ^2. KSS devise an analogous approach for the general, heteroscedastic case.

Leave-out estimation (3/8)

The bias correction devised by KSS is based on a **"leave-*i*-out"** OLS estimator of $\boldsymbol{\beta}$, which is calculated by removing observation i from the standard OLS formula.

$$\widehat{\boldsymbol{\beta}}_{-i} = \left(\mathbf{S}_{\boldsymbol{xx}} - \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}}\right)^{-1} \sum_{\ell \neq i} \mathbf{x}_\ell y_\ell$$

This delivers an *unbiased* estimator of σ_i^2.

$$\widehat{\sigma}_i^2 = y_i \left(y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i}\right)$$

Thus, an **unbiased** estimator for θ can be obtained as follows.

$$\widehat{\theta}_{KSS} = \widehat{\boldsymbol{\beta}}_{OLS}^{\mathrm{T}} \mathbf{C} \widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} B_{ii} \widehat{\sigma}_i^2$$

This also leads to an *unbiased* estimator of $\mathbb{V}\mathrm{ar}\left[\widehat{\boldsymbol{\beta}}_{OLS}\right]$.

$$\widehat{\mathbb{V}\mathrm{ar}}\left[\widehat{\boldsymbol{\beta}}_{OLS}\right] = \mathbf{S}_{\boldsymbol{xx}}^{-1} \left(\sum_{i=1}^{N} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \widehat{\sigma}_i^2\right) \mathbf{S}_{\boldsymbol{xx}}^{-1}$$

Leave-out estimation (4/8)

It may be complicated to calculate N leave-i-out estimators every time. KSS suggest to exploit the easier-to-compute N quantities:

$$P_{ii} \equiv \mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i$$

since they deliver a simpler expression for the $y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i}$ terms.

$$\begin{aligned} y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i} &= y_i - \mathbf{x}_i^{\mathrm{T}} \left(\mathbf{S}_{\boldsymbol{xx}} - \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \right)^{-1} \sum_{\ell \neq i} \mathbf{x}_\ell y_\ell \\ &= y_i - \mathbf{x}_i^{\mathrm{T}} \left(\mathbf{S}_{\boldsymbol{xx}}^{-1} + \frac{\mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1}}{1 - P_{ii}} \right) \left(\sum_{\ell=1}^{N} \mathbf{x}_\ell y_\ell - \mathbf{x}_i y_i \right) \\ &= \left(1 + P_{ii} + \frac{P_{ii}^2}{1 - P_{ii}} \right) y_i - \left(1 + \frac{P_{ii}}{1 - P_{ii}} \right) \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS} \\ &= \frac{y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS}}{1 - P_{ii}} \end{aligned}$$

The second line uses the Sherman-Morrison-Woodbury formula.

Leave-out estimation (5/8)

The KSS estimator can also be motivated via a *change in variable* $\widetilde{\mathbf{x}}_i \equiv \mathbf{C}\mathbf{S}_{\boldsymbol{xx}}^{-1}\mathbf{x}_i$. Observe that:

$$\theta = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{C}\boldsymbol{\beta} = \boldsymbol{\beta}^{\mathrm{T}}\mathbf{S}_{\boldsymbol{xx}}\mathbf{S}_{\boldsymbol{xx}}^{-1}\mathbf{C}\boldsymbol{\beta} = \sum_{i=1}^{N}\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i\widetilde{\mathbf{x}}_i^{\mathrm{T}}\boldsymbol{\beta} = \sum_{i=1}^{N}\mathbb{E}_{Y|\mathbf{x}}\left[Y_i\widetilde{\mathbf{x}}_i^{\mathrm{T}}\boldsymbol{\beta}\right]$$

and thanks again to the Sherman-Morrison-Woodbury formula, the KSS estimator can also be written as follows.

$$\begin{aligned}
\widehat{\theta}_{KSS} = \sum_{i=1}^{N} y_i\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{-i} &= \sum_{i=1}^{N} y_i\widetilde{\mathbf{x}}_i^{\mathrm{T}}\left(\mathbf{S}_{\boldsymbol{xx}}^{-1} + \frac{\mathbf{S}_{\boldsymbol{xx}}^{-1}\mathbf{x}_i\mathbf{x}_i^{\mathrm{T}}\mathbf{S}_{\boldsymbol{xx}}^{-1}}{1-P_{ii}}\right)\sum_{\ell\neq i}\mathbf{x}_\ell y_\ell \\
&= \sum_{i=1}^{N}\left[y_i\widetilde{\mathbf{x}}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{OLS} - B_{ii}y_i^2 + B_{ii}y_i\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{-i}\right] \\
&= \widehat{\boldsymbol{\beta}}_{OLS}^{\mathrm{T}}\mathbf{C}\widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} B_{ii}\widehat{\sigma}_i^2
\end{aligned}$$

Leave-out estimation (6/8)

This is a technical digression. The previous derivation exploited the following relationship:

$$\begin{aligned}
\sum_{i=1}^{N} y_i \widetilde{\mathbf{x}}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \sum_{\ell \neq i} \mathbf{x}_\ell y_\ell &= \sum_{i=1}^{N} y_i \widetilde{\mathbf{x}}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} \widetilde{\mathbf{x}}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i y_i^2 \\
&= \sum_{i=1}^{N} y_i \widetilde{\mathbf{x}}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} \underbrace{\mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{C} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i}_{=B_{ii}} y_i^2 \\
&= \sum_{i=1}^{N} y_i \mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{C} \widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} B_{ii} y_i^2 \\
&= \widehat{\boldsymbol{\beta}}_{OLS}^{\mathrm{T}} \mathbf{C} \widehat{\boldsymbol{\beta}}_{OLS} - \sum_{i=1}^{N} B_{ii} y_i^2
\end{aligned}$$

…and…

$$B_{ii} y_i^2 - B_{ii} y_i \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i} = B_{ii} y_i \left(1 - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i}\right) = B_{ii} \widehat{\sigma}_i^2$$

Leave-out estimation (7/8)

…and…

$$\sum_{i=1}^{N} y_i \widetilde{\mathbf{x}}_i^{\mathrm{T}} \frac{\mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i \mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1}}{1 - P_{ii}} \sum_{\ell \neq i} \mathbf{x}_\ell y_\ell = \sum_{i=1}^{N} y_i \underbrace{\mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{C} \mathbf{S}_{\boldsymbol{xx}}^{-1} \mathbf{x}_i}_{=B_{ii}} \underbrace{\mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i}}_{=\widehat{y}_i}$$

where $\widehat{y}_i$ above is derived as follows.

$$\begin{aligned}
\widehat{y}_i &= \frac{\mathbf{x}_i^{\mathrm{T}} \mathbf{S}_{\boldsymbol{xx}}^{-1}}{1 - P_{ii}} \left(\sum_{\ell=1}^{N} \mathbf{x}_\ell y_\ell - \mathbf{x}_i y_i \right) \\
&= \frac{\mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS}}{1 - P_{ii}} - \frac{P_{ii} y_i}{1 - P_{ii}} + \frac{y_i - y_i}{1 - P_{ii}} \\
&= y_i - \frac{y_i - \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{OLS}}{1 - P_{ii}} \\
&= y_i - y_i + \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i} \\
&= \mathbf{x}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}_{-i}
\end{aligned}$$

Leave-out estimation (8/8)

- After introducing their estimator, in their paper KSS show its potential across different settings, although the primary motivation is the AKM model.

- Then, KSS derive its asymptotic properties: a daunting task in this case (the key estimator is a quadratic form).

- They also provide routines for the quick estimation of the N key weights P_{ii} using standard computing languages.

- KSS also test their estimator on MEE data from the Italian region of Veneto, comparing it with both the "plug-in" (PI) and the "homoscedasticity only" (HO) estimators.

- Their estimates of the *correlation* between $\boldsymbol{\alpha}$ and $\boldsymbol{\psi}$ are zero or negative using the PI, in the range 5-15% with the HO, and in the range 20-30% with the KSS estimator.

A random effects approach (1/3)

- In a review of the literature, Bonhomme et al. (2020) argue that a random effects approach is a viable alternative.

- In this environment the worker fixed effects $\boldsymbol{\alpha}$ are assumed to have some distribution and moments that depends on $\mathbf{D}$.

- Identification requires some **covariance restrictions**:

$$\begin{aligned} \mathbb{C}\text{ov}\left(\alpha_i, \psi_j\right) &= 0 \quad \text{for } j \notin \mathcal{J}\left(i\right) \\ \mathbb{C}\text{ov}\left(\psi_j, \psi_{j'}\right) &= 0 \quad \text{for } j \neq j' \\ \mathbb{C}\text{ov}\left(\varepsilon_{it}, \psi_{i't'}\right) &= 0 \quad \text{for } i \neq i',\ t \neq t' \end{aligned}$$

  where $\mathcal{J}\left(i\right) = \left\{j : j \neq j\left(i,t\right)\ \forall t = t_{0i}, \ldots, T_i\right\}$ is the set of firms where worker i *never* works.

- This assumptions may not be very desirable, but they could be somewhat relaxed (the ones reported are illustrative).

A random effects approach (2/3)

To understand how identification works, consider any two workers i and i' who initially work at the same firm, and subsequently move to two different firms: at time t and t' respectively. Assume that $j(i,t) \notin \mathcal{J}(i')$ and $j(i',t') \notin \mathcal{J}(i)$.

The analysis of **movers** in a **first differences** model illustrates identification of key moments. Let (for simplicity) $\boldsymbol{x}_{it} = \mathbf{0}$ for all workers i and time periods t. One can show that:

$$\mathbb{C}\text{ov}\left(\log W_{it'} - \log W_{it}, \log W_{i't'} - \log W_{i't}\right) = \mathbb{V}\text{ar}\left(\psi_{j(i,t)}\right)$$

as well as the following (both results are shown next).

$$\mathbb{C}\text{ov}\left(\log W_{it'} - \log W_{it}, \log W_{i't'}\right) = -\,\mathbb{C}\text{ov}\left(\psi_{j(i',t)}, \alpha_{i'}\right)$$

These moments are estimable via minimum distance; Bonhomme et al. show that this approach delivers very precise estimates.

A random effects approach (3/3)

The variance of the firm effects is identified as follows.

$$\begin{aligned}\mathbb{C}\text{ov}\,(\log W_{it'} - \log W_{it}\,,\, \log W_{i't'} - \log W_{i't}) = \\ &= \mathbb{C}\text{ov}\left(\psi_{j(i,t')} - \psi_{j(i,t)} + \varepsilon_{it'} - \varepsilon_{it},\right.\\ &\qquad \left.\psi_{j(i',t')} - \psi_{j(i',t)} + \varepsilon_{i't'} - \varepsilon_{i't}\right)\\ &= \mathbb{C}\text{ov}\left(\psi_{j(i,t')} - \psi_{j(i,t)}, \psi_{j(i',t')} - \psi_{j(i',t)}\right)\\ &= \mathbb{C}\text{ov}\left(\psi_{j(i,t)}, \psi_{j(i',t)}\right)\\ &= \mathbb{V}\text{ar}\left(\psi_{j(i,t)}\right)\end{aligned}$$

The covariance between both effect types is identified as follows.

$$\begin{aligned}\mathbb{C}\text{ov}\,(\log W_{it'} - \log W_{it}, \log W_{i't'}) = \\ &= \mathbb{C}\text{ov}\left(\psi_{j(i,t')} - \psi_{j(i,t)} + \varepsilon_{it'} - \varepsilon_{it}, \alpha_{i'} + \psi_{j(i',t')} + \varepsilon_{i't'}\right)\\ &= \mathbb{C}\text{ov}\left(\psi_{j(i,t')} - \psi_{j(i,t)}, \alpha_{i'} + \psi_{j(i',t')}\right)\\ &= \mathbb{C}\text{ov}\left(\psi_{j(i,t')} - \psi_{j(i,t)}, \alpha_{i'}\right)\\ &= -\,\mathbb{C}\text{ov}\left(\psi_{j(i',t)}, \alpha_{i'}\right)\end{aligned}$$

Discrete types of workers and firms (1/10)

- Random effects have appealing features, but they come with costly assumptions. At the same time, they afford the option to estimate more general models.

- The random effects framework proposed by BLM features a major innovation: worker and firm effects can possibly enter *non-linearly* and *dynamically* into the log-wage equation.

- This allows to estimate economically meaningful effects: the *interaction* between worker and firm effects (e.g. "matching" effects) and the dependence of today's wage on the *past* wage and a worker's *past* employers.

- These options come at a cost: the *types* of both workers and firms are assumed to belong to a *discrete* set.

- The estimation approach is original and deserves discussion.

Discrete types of workers and firms (2/10)

Consider the following model, called **"static"** by BLM.

$$\log W_{it} = a_t\left(c_{it}\right) + b_t\left(c_{it}\right)\alpha_i + \boldsymbol{x}_{it}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{it}$$

where, contrasting with AKM:

- α_i is a **random** effect specific to worker i; it is assumed that the support of α_i is a **discrete** set of dimension L;
- $c_{it} = c\left(j\left(i,t\right)\right)$ is the **class** (type) of worker i's employer at time t; it assumed that the image of $c\left(\cdot\right)$ – and the support of c_{it} – is a **discrete** set of dimension C;
- $a_t\left(\cdot\right)$ and $b_t\left(\cdot\right)$ are two arbitrary, unknown functions of c_{it}; they are possibly (though not necessarily) time-dependent.

Notably, the AKM model is a **special case** of this model, where $L = N$, $C = J$, $a_t\left(c_{it}\right) = \psi_{j(i,t)}$ and $b_t\left(c_{it}\right) = 1$ for all (i,t) pairs.

Discrete types of workers and firms (3/10)

The aim of BLM is to estimate the full distributions of log-wages across different types of workers and firms, so as to learn insights on the labor market *and* on the interaction effects $b_t(\cdot)$.

To this end, BLM introduce two **assumptions** for their "static" model, which resemble and extend the AKM assumptions.

Let $m_{it} = 1$ denote that worker i **moves** to a new firm in period $t+1$. Also let $Y_{it} = \log W_{it}$. The assumptions are as follows.

1. **Exogenous mobility**: m_{it}, $c_{i(t+1)}$, $\boldsymbol{x}_{i(t+1)}$ are independent of $(Y_{it_{0i}}, \dots, Y_{it})$ conditional on all previous realizations of m_{is}, $c_{i(s+1)}$ and $\boldsymbol{x}_{i(s+1)}$ for all $s < t$, and on α_i.

2. **Serial independence**: $Y_{i(t+1)}$ is independent of all previous realizations of $Y_{i(s+1)}$, m_{is}, $c_{i(s+1)}$ and $\boldsymbol{x}_{i(s+1)}$ for all $s < t$, conditional on α_i, $c_{i(t+1)}$, $\boldsymbol{x}_{i(t+1)}$ and $m_{it} = 1$.

Discrete types of workers and firms (4/10)

Under these assumptions, the joint c.d.f. of **movers'** log-wages before and after the move (occurred between $t = 1$ and $t = 2$) is:

$$\mathbb{P}\left[Y_{i1} \leq y_1, Y_{i2} \leq y_2 \middle| c_{i1} = c, c_{i2} = c', m_{i1} = 1\right] = \\ = \sum_{\alpha=1}^{L} F_{c\alpha}\left(y_1\right) F_{c'\alpha}^{m}\left(y_2\right) p_{cc'}\left(\alpha\right)$$

where:

- $F_{c\alpha}\left(y_1\right)$ is the c.d.f. of log-wages for workers of type α who are employed in firms of class c;
- $F_{c'\alpha}^{m}\left(y_2\right)$ is the c.d.f. of log-wages for **movers** of type α who are employed in firms of class c';
- $p_{cc'}\left(\alpha\right)$ is the proportion of workers of type α who move from a firm of class c to a firm of class c' between $t = 1$ and $t = 2$.

Discrete types of workers and firms (5/10)

In their Theorem 1, BLM prove that the components of this joint c.d.f. – as well as the unconditional initial type proportions $q_c(\alpha)$ at $t = 1$ – are non-parametrically identified in a connected set.

This identifies "ratios" $b_t(c')/b_t(c)$ (for $c \neq c'$) of the interaction effects across firm types. An example illustrates the intuition.

- Let $C = 2$: there are only two firm types.
- Also let $\boldsymbol{x}_{it} = \mathbf{0}$ for all pairs (i, t), for simplicity.
- Consider workers who move between $t = 1$ and $t = 2$.
- Some move from $c = 1$ to $c = 2$, the others make the reverse step from $c = 2$ to $c = 1$ (as in the mover analysis by CHK).
- In both subgroups, there are workers of **different type** $\boldsymbol{\alpha_i}$. The "average type" is $\mathbb{E}_{12}[\alpha_i]$ in the former subgroup, and $\mathbb{E}_{21}[\alpha_i]$ in the latter.

Discrete types of workers and firms (6/10)

Let $b_t(\cdot) = b(\cdot)$ be constant in time. By the two assumptions:

$$\mathbb{E}_{\mathbb{W}}[\log W_{it}] = a(c_{it}) + b(c_{it})\,\mathbb{E}_{\mathbb{W}}[\alpha_i]$$

where both expectations are taken with respect to a subset of the worker population $\mathbb{W}$. Let these be the '12' and '21' subgroups (to be used as subscripts instead of $\mathbb{W}$) from the example; then:

$$\begin{aligned}
\mathbb{E}_{12}[\log W_{i2}] - \mathbb{E}_{21}[\log W_{i1}] &= b(2)\left\{\mathbb{E}_{12}[\alpha_i] - \mathbb{E}_{21}[\alpha_i]\right\} \\
\mathbb{E}_{12}[\log W_{i1}] - \mathbb{E}_{21}[\log W_{i2}] &= b(1)\left\{\mathbb{E}_{12}[\alpha_i] - \mathbb{E}_{21}[\alpha_i]\right\}
\end{aligned}$$

and if $\mathbb{E}_{12}[\alpha_i] \neq \mathbb{E}_{21}[\alpha_i]$ (a "rank condition"):

$$\frac{b(2)}{b(1)} = \frac{\mathbb{E}_{12}[\log W_{i2}] - \mathbb{E}_{21}[\log W_{i1}]}{\mathbb{E}_{12}[\log W_{i1}] - \mathbb{E}_{21}[\log W_{i2}]}$$

hence, the **ratio** $b(2)/b(1)$ is identified. This can be generalized to $C > 2$, time effects on $b(\cdot)$, and covariates $\boldsymbol{x}_{it}$.

Discrete types of workers and firms (7/10)

The BLM estimation approach proceeds in two steps, plus some "post-estimation" analysis.

1. The **first step** identifies firm types non-parametrically. This adapts a **"*k*-means"** classification algorithm borrowed from statistics and machine learning, by solving the problem

$$\min_{c(1),\dots,c(J);H_1,\dots,H_C} \sum_{j=1}^{N} N_j \int_{\mathbb{R}} \left(\widehat{F}_j \left(y \right) - H_{c(j)} \right)^2 d\mu \left(y \right)$$

   where:

   - N_j is the number of workers at firm j;
   - $\widehat{F}_j \left(y \right)$ is the empirical c.d.f. of log-wages in firm j;
   - H_c is some c.d.f. for $c = 1, \dots, C$;
   - $\mu \left(y \right)$ is a discrete or continuous measure.

Discrete types of workers and firms (8/10)

2. Having recovered estimates of firm classes $\widehat{c}(j)$ for all firms $j = 1, \ldots, J$, in the **second step** one can estimate all key parameters of the log-wages distributions. Using the sample of all job **movers** N_m, by the two assumptions one can write a log-likelihood function for two time periods $t = 1, 2$.

$$\log \mathcal{L}_{12}^{m}\left(\boldsymbol{\theta}_p, \boldsymbol{\theta}_f, \boldsymbol{\theta}_{f^m}\right) = \sum_{i=1}^{N_m} \sum_{c=1}^{C} \sum_{c'=1}^{C} \mathbb{1}\left[\widehat{c}_{i1} = c\right] \mathbb{1}\left[\widehat{c}_{i2} = c'\right] \times$$

$$\times \log \left(\sum_{\alpha=1}^{L} p_{cc'}\left(\alpha; \boldsymbol{\theta}_p\right) f_{c\alpha}\left(y_{i1}; \boldsymbol{\theta}_f\right) f_{c'\alpha}^{m}\left(y_{i2}; \boldsymbol{\theta}_{f^m}\right) \right)$$

- The worker type weights $p_{cc'}$ are treated as probabilities of a **mixture model**. This demands $\widehat{\boldsymbol{\theta}}_p$, $\widehat{\boldsymbol{\theta}}_f$ and $\widehat{\boldsymbol{\theta}}_{f^m}$ to be estimated via an "EM" algorithm (overviewed later).
- The estimates $\widehat{\boldsymbol{\theta}}_f$ are exploited to recover the proportion of types $q_c(\alpha)$ in the full worker population at $t = 1$.

Discrete types of workers and firms (9/10)

BLM also propose a more extended **"dynamic"** model:

$$\log W_{it} = \rho_t \log W_{i(t-1)} + a_{1t}(c_{it}) + a_{2t}\left(c_{i(t-1)}\right) + \\ + b_t(c_{it})\,\alpha_i + \boldsymbol{x}_{it}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{it}$$

where, contrasting with the "static" case:

- parameter ρ_t measures the **time dependence** of log-wages on its own past realization;
- $a_{1t}(\cdot)$ and $a_{2t}(\cdot)$ are "fixed effects" that depend on the class of the current and **previous employer**, respectively.

To achieve identification, BLM propose "Markovian" versions of their assumptions: future key variables are independent of their past realizations **conditional** on the current ones. The resulting Theorem 2 (identification) and extended estimation framework are based on four time periods rather than two.

Discrete types of workers and firms (10/10)

- BLM estimate their static model on Swedish MEE data for $L = 6$ and $C = 10$ (the results seem robust to other choices). In the second step of their estimation approach, they assume $f_{c\alpha}(\cdot)$ and $f^{m}_{c'\alpha}(\cdot)$ to be log-normal.

- They provide nice "visual" representations of their estimates about discrete types, which show that sorting indeed occurs and that lower-type workers especially benefit from the most higher-type employers.

- These results are confirmed by an AKM exercise in variance decomposition using projected data based on the estimates of the static model. There is evidence for complementarities $b_t(\cdot)$, but these appear small in magnitude.

- The dynamic model yields statistically significant estimates of ρ_t as well as large "previous employer" effects $a_{2t}(\cdot)$.

Mixture models and the EM algorithm (1/4)

To understand estimation in BLM's second step, it is quite useful to overview the statistical approach to **mixture models**: models for data that are presumed originated from multiple populations.

A random variable X follows a (finite) **mixture distribution** if its c.d.f. can be written as:

$$F_X(x; \boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{c=1}^{C} \pi_c F_c(x; \boldsymbol{\theta}_c)$$

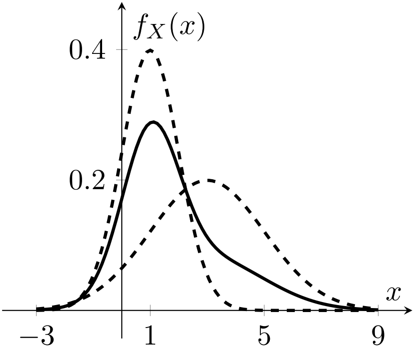
where for $c = 1, \ldots, C$, $F_c(x; \boldsymbol{\theta}_c)$ is a c.d.f. with parameters $\boldsymbol{\theta}_c$, $\pi_c \in [0, 1]$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_C)$, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_C)$, and $\sum_{c=1}^{C} \pi_c = 1$.

- An analogous expression applies to the p.m.f. or p.d.f. of X.
- This is a model for populations that originate from "mixing" C subpopulations with associated **latent variables**, whose c.d.f.s are $F_c(x; \boldsymbol{\theta}_c)$ and whose weights $\boldsymbol{\pi}$ are **unknown**.

Mixture models and the EM algorithm (2/4)

A prominent mixture model is the **Gaussian mixture**, where all the $F_c(x; \theta_c)$ distributions are normal and where $\theta_c = (\mu_c, \sigma_c^2)$ are the corresponding location and scale parameter.

The figure below represents a Gaussian mixture distribution for $C = 2$ (solid line) with $\pi_1 = 1 - \pi_2 = 0.6$. The latent distributions (dashed lines) have $(\mu_1, \sigma_1^2) = (1, 1)$ and $(\mu_2, \sigma_2^2) = (3, 4)$.



Mixture models and the EM algorithm (3/4)

How to estimate the parameters $(\boldsymbol{\theta}, \boldsymbol{\pi})$? The log-likelihood reads:

$$\log \mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\pi} \middle| x_1, \ldots, x_N\right) = \sum_{i=1}^{N} \log \left(\sum_{c=1}^{C} \pi_c f_c\left(x_i; \boldsymbol{\theta}_c\right) \right)$$

given a random sample $\{x_i\}_{i=1}^{N}$. This is not well suited to MLE. The First Order Conditions for one parameter subset $\boldsymbol{\theta}_c$ give:

$$\frac{\partial \log \mathcal{L}\left(\boldsymbol{\theta}, \boldsymbol{\pi} \middle| x_1, \ldots, x_N\right)}{\partial \boldsymbol{\theta}_c} = \sum_{i=1}^{N} \varphi_c\left(x_i; \boldsymbol{\theta}, \boldsymbol{\pi}\right) \frac{\partial \log f_c\left(x_i; \boldsymbol{\theta}_c\right)}{\partial \boldsymbol{\theta}_c}$$

where:

$$\varphi_c\left(x_i; \boldsymbol{\theta}, \boldsymbol{\pi}\right) = \frac{\pi_c f_c\left(x_i; \boldsymbol{\theta}_c\right)}{\sum_{c'=1}^{C} \pi_{c'} f_{c'}\left(x_i; \boldsymbol{\theta}_{c'}\right)}$$

is the posterior probability that observation i belongs to group c. Note (e.g. from the Gaussian mixture's case) that if the N values of $\varphi_c\left(x_i; \boldsymbol{\theta}, \boldsymbol{\pi}\right)$ were known, it would be possible to solve for the $\boldsymbol{\theta}$ parameters, and thus the $\boldsymbol{\pi}$ weights could also be retrieved.

Mixture models and the EM algorithm (4/4)

A popular solution is the **expectation-maximization** (E-M or "EM") iterative algorithm. Given some initial values $\left(\boldsymbol{\theta}^{(h)}, \boldsymbol{\pi}^{(h)}\right)$, the algorithm proceeds in two steps:

1. **E-step**: calculate the N posteriors $\varphi_c^{(h)}\left(x_i; \boldsymbol{\theta}^{(h)}, \boldsymbol{\pi}^{(h)}\right)$;
2. **M-step**: update the values $\left(\boldsymbol{\theta}^{(h+1)}, \boldsymbol{\pi}^{(h+1)}\right)$ from the FOCs;

and iterate until convergence.

The EM algorithm can be generalized to a **multivariate** mixture model where the subpopulation weights are functions of the data:

$$F_{\boldsymbol{x}}\left(\mathbf{x}; \boldsymbol{\theta}_p, \boldsymbol{\theta}_f\right) = \sum_{c=1}^{C} p_c\left(\mathbf{x}; \boldsymbol{\theta}_{pc}\right) F_c\left(\mathbf{x}; \boldsymbol{\theta}_{fc}\right)$$

where $\boldsymbol{\theta}_p = \left(\boldsymbol{\theta}_{p1}, \dots, \boldsymbol{\theta}_{pC}\right)$, $\boldsymbol{\theta}_f = \left(\boldsymbol{\theta}_{f1}, \dots, \boldsymbol{\theta}_{fC}\right)$ and $p_c\left(\mathbf{x}; \boldsymbol{\theta}_{pc}\right)$ has e.g. a multinomial logit form. The BLM model is a particular case of this, where $F_c\left(\mathbf{x}; \boldsymbol{\theta}_{fc}\right)$ is the joint c.d.f. of mover wages.

Origin effects and wage dynamics (1/5)

- The idea to allow for the effect of a worker's *past* employers has been recently introduced in a linear wage model as well. In particular, Di Addario, Kline, Saggio and Sølvsten (2021; DAKSS) estimate the following augmented AKM model:

$$\log W_{it} = \alpha_i + \psi_{j(i,t)} + \lambda_{j(i,t-1)} + \boldsymbol{x}_{it}^{\mathrm{T}}\boldsymbol{\beta} + \varepsilon_{it}$$

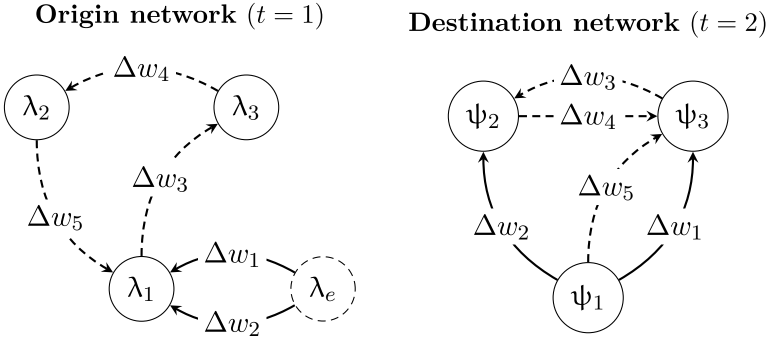
  where $\lambda_{j(i,t-1)}$ is the effect of worker i's employer at $t-1$ (it is also allowed to capture unemployment or first time jobs).

- Identification of this model is based on a stricter definition of "connected set:" any three firms in the set must be linked via a "closed walk" determined by worker movements.

- DAKSS adapt a leave-out approach to estimate the variance components of this model using Italian MEE data from the *Istituto Nazionale di Previdenza Sociale*: the "origin effects" $\lambda_{j(i,t-1)}$ seem to explain little of the grand variance of wages.

Origin effects and wage dynamics (2/5)

The intuition about identification is best illustrated graphically.



Let $w_i = \log W_i$. The figures represent a network of five workers and three firms; edges represent worker movements over two time periods. Two workers move first from an external firm e to firm 1 and later to firm 2 or 3 (continuous edges). Labels superimposed on edges indicate what wage changes identify what parameters.

Origin effects and wage dynamics (3/5)

A very attractive feature of this model is its connection with the literature about "random" *job search*. This is best illustrated via a stylized version of the model by Postel-Vinay and Robin (2002).

- Workers have productivity ϵ, derive utility $\mathcal{U}(w)$ from wage w if employed, and utility ϵb if unemployed.
- Firms have productivity p, which leads to a marginal product of ϵp when they hire a worker of type ϵ.
- Workers "randomly search" firms even if employed, leading them to meet firm types drawn from a distribution $F(\cdot)$.
- Random search is governed by a survival function $\overline{F}(\cdot)$.
- Any worker employed at a firm of type q will only move to a randomly met firm of type $p > q$ which offers a "poaching wage" that fully compensates the worker for the movement.

Origin effects and wage dynamics (4/5)

Postel-Vinay and Robin show that such a poaching wage $\phi(\epsilon, p, q)$ must satisfy, given a constant $\kappa \geq 0$ that captures other features of the model (arrival, discount and separation rates):

$$\mathcal{U}(\phi(\epsilon, p, q)) = \mathcal{U}(\epsilon q) - \kappa \int_q^p \overline{F}(x)\,\mathcal{U}'(\epsilon x)\,\epsilon dx$$

that is, $\phi(\epsilon, p, q)$ yields a utility superior to the best counteroffer that is affordable by the incumbent employer ($w = \epsilon q$), *minus* a *compensating differential* for the *change in expected future utility* that derives from searching while employed (the integral). In this model, unemployment is treated as a firm of type b.

- The parameters of this model govern the **dynamics** of wages in a worker's employment history.
- The literature on *"random"* search provides a framework for estimating these parameters based on **"indirect" inference** (an extension of the Method of Simulated Moments).

Origin effects and wage dynamics (5/5)

Assuming a *logarithmic utility* $\mathcal{U}(w) = \log(w)$, by this model:

$$\underbrace{\log(\phi(\epsilon, p, q))}_{=W_{it}} = \underbrace{\log(\epsilon)}_{=\alpha_i} + \underbrace{I(p)}_{=\psi_{j(i,t)}} + \underbrace{\log(q) - I(q)}_{=\lambda_{j(i,t-1)}}$$

for a worker i who moves between $t = 1$ and t. By the definition:

$$I(z) \equiv \kappa \int_z^\infty \frac{\overline{F}(x)}{x} dx$$

$I(p)$ and $I(q)$ follow from the Fundamental Theorem of Calculus.

- The connection with the DAKSS model is straightforward.
- It also holds for some more general random search models.
- This loads the DAKSS parameters with interpretation: e.g. for a firm j with productivity p, $\psi_j + \lambda_j = \log(p)$ holds; it also implies negative correlation between ψ_j and λ_j.