High Risk Workers and High Risk Firms

Marlene Koffi — U of Toronto

Serdar Ozkan — U of Toronto

Sergio Salgado — Wharton

Marco Weissler — ECB

SED Annual Meeting, Minneapolis

July 1, 2021

Introduction

Motivation

Vast Heterogeneity in Labor Income Dynamics

Large heterogeneity in earnings dynamics individuals experience:

- In the average income profile (e.g., Baker (1997); Guvenen (2009))
- In the variance of income shocks (e.g., Browning *et al.* (2010); Meghir and Pistaferri (2004))
- In higher order moments of income shocks: skewness/kurtosis (e.g., Guvenen *et al.* (2018); Arellano *et al.* (2017))
- In unemployment risk and job finding rate (Karahan *et al.* (2019); Hall and Kudlyak (2019); Gregory *et al.* (2021))

These differences can be due to:

- Worker characteristics
  - Observed (e.g. age, income, etc., e.g., Guvenen *et al.* (2018)),
  - Unobserved (Browning *et al.* (2010); Bagger *et al.* (2014))
- Firm characteristics
  - Observed (e.g., size, average wage, etc.),
  - Unobserved (Jarosch (2015); Lentz *et al.* (2018))

This Paper

- Study heterogeneity in earnings dynamics by observable and unobservable worker and firm components
- Use matched employer-employee administrative data from Germany
- Observable het. Earnings risk conditional on jointly observable worker and firm characteristics
- Unobservable het. Use clustering algorithms to classify similar workers & firms by features of earnings dynamics
- Estimate an individual income process that allows for both worker and firm heterogeneity
    - To quantify firm's importance for workers' income dynamics and risk.

Data and Sample

Matched Employer-Employee Panel Dataset

- Integrated Employment Biographies (IEB) data file of the German social security system
  (see Oberschachtsiek *et al.* (2010) for details)
- Between 1975 and 2019 from West Germany, from 1990 to 2019 for the entire country.
- Includes all employees with the exception of civil servants and self-employed workers
- Sample: Male workers between ages 25 and 55

Include information on

- Personal characteristics of workers—e.g., gender, birth date, education level, and occupation,
- on employment—e.g., earnings, days worked, occupations, 5-digit job code, full-time or part-time
- On establishments (firms)—e.g., industry, location, average wages and number of employees

Earnings above social security limit is imputed. (a lá Card *et al.* (2013))

Earnings Dynamics By Observables

Methodology

Methodology

Residual log earnings growth: control for age and year effects

- Investigate mean, variance, skewness and kurtosis of 1-, 3- and 5-year earnings growth
  (today we focus on 3-year changes)

Condition workers w.r.t. their recent earnings and age (e.g., Guvenen *et al.* (2018))

- Rank workers into 50 quantiles based on past 3-year earnings average b/w $t-1$ and $t-3$ within their age groups
- Include those who have positive earnings in $t-1$ and and at least one more year to ensure labor market attachment

Condition firms w.r.t. their size, employment growth, and average wage

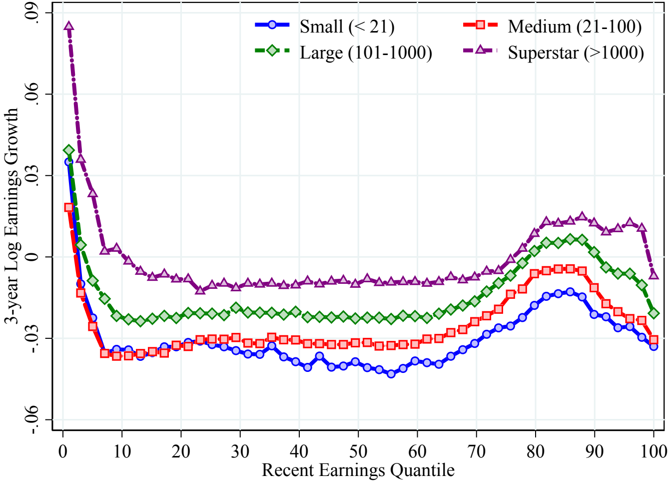
- Include firms with at least 10 workers in year t
- Workers are assigned to firms based on their main employment in year t.

After selection our sample consists of ~11M workers and ~5M firms for a total of 144M worker/year obs.

Earnings Dynamics By Observables

1. Average Earnings Growth

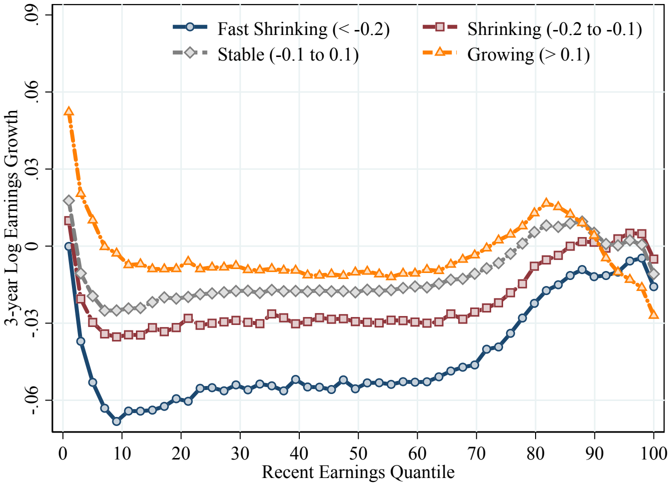
Average 3-Year Earnings Growth in Big and Small Firms



Mean reversion warning! Some mean reversion is to be expected at the lower and higher ends of the recent earnings distribution (x-axis)

- Rank firms by their size in year *t* into
  - small (<=20 employees),
  - medium (21-100 employees),
  - large (101-1,000 employees), and
  - superstars (>1,000 employees)
- Workers of super-star firms experience
  - steeper earnings growth relative to similar workers in smaller firms
  - the effect is particularly more pronounced for low-income workers

Average 3-Year Earnings Growth in Growing and Shrinking Firms



- Rank firms by their arc-percent employment growth between t and $t+3$:

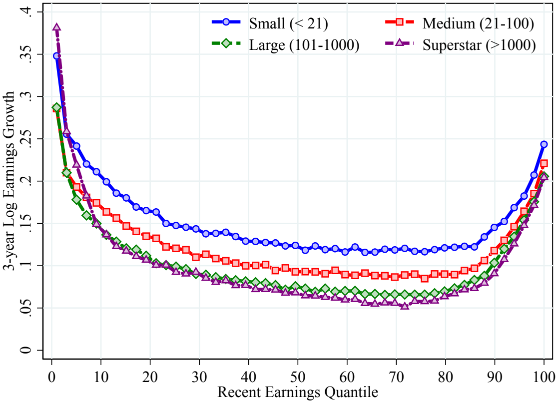
$$\Delta^{arc}_{j,3} = 2\frac{(emp_{t+3} - emp_t)}{(emp_{t+3} + emp_t)}$$

  - fast shrinking (<-0.2),
  - shrinking (-0.2 to -0.1),
  - stable (-0.1 to 0.1),
  - growing (0.1 or more)
- Workers in growing firms experience much steeper growth in their earnings.
  - Especially low-income workers.

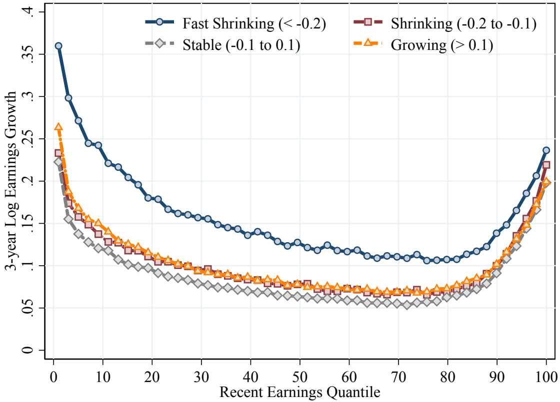
Earnings Dynamics By Observables

2. Dispersion of Earnings Growth

Dispersion of Earnings Growth by Firm Size and Firm Growth



(a) By Firm Size



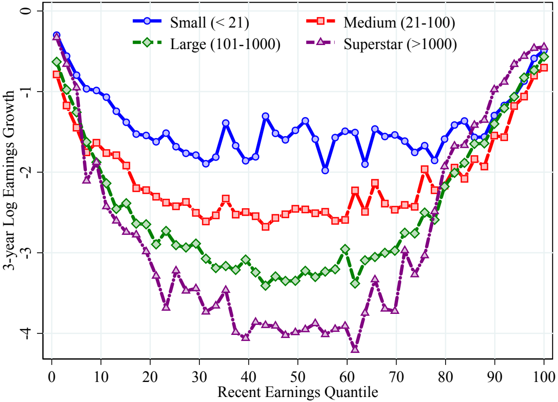
(b) By Firm Growth

- Low- and high-income workers experience larger dispersion of earnings growth (compare left and right ends)
- Workers at small (left panel) and fast shrinking firms (right panel) also experience more dispersion in earnings growth
- Variation across workers, however, is significantly larger than across firms

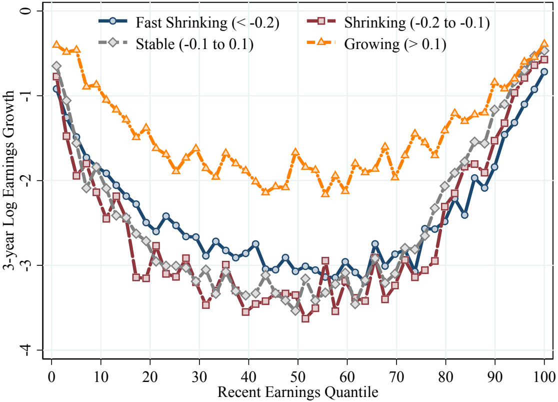
Earnings Dynamics By Observables

3. Skewness of Earnings Growth

Skewness of Earnings Growth by Firm Size and Firm Growth



(a) By Firm Size



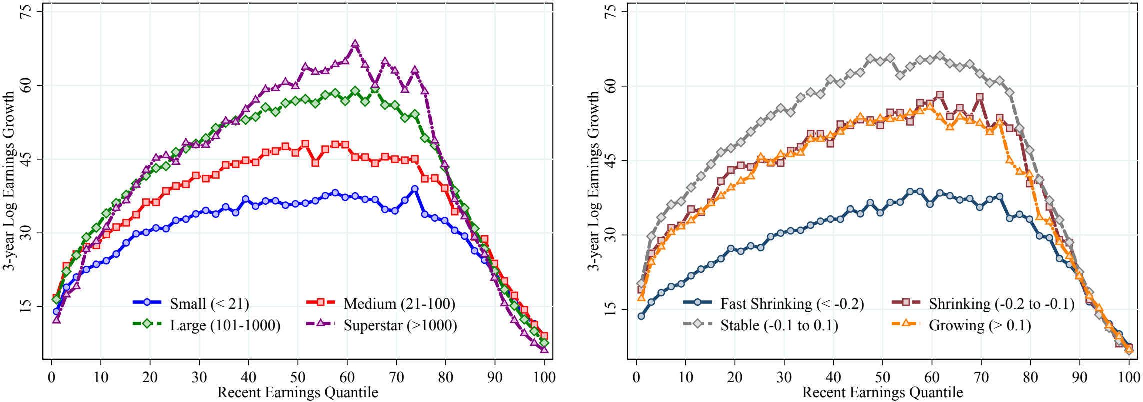
(b) By Firm Growth

- Low- and high-income workers experience larger skewness of earnings growth (compare left and right ends)
- For mid-wage workers, skewness is larger at small and fast growing firms relative to similar workers in employed in other firms types
- Firm variation is more important for middle income worker than low- and high-income workers

Earnings Dynamics By Observables

4. Kurtosis of Earnings Growth

Kurtosis of Earnings Growth by Firm Size and Firm Growth



(a) By Firm Size

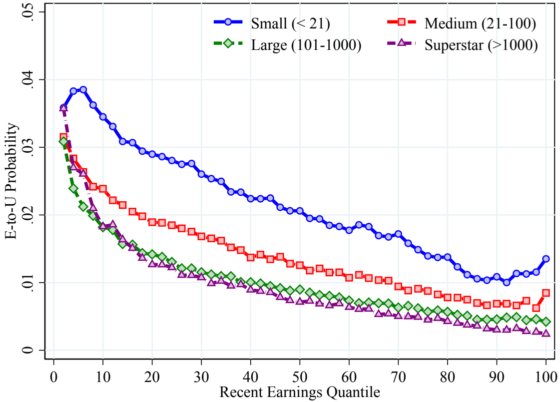
(b) By Firm Growth

- Low- and high-earnings workers experience lower kurtosis relative to middle-income workers
- Workers at smaller and fast shrinking firms experience lower kurtosis, specially for middle income workers
- Variation across the past income distribution is more pronounced compared to the firm variation

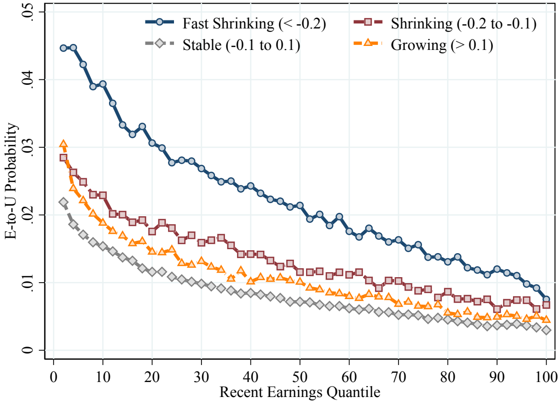
Earnings Dynamics By Observables

Unemployment Risk

Unemployment Risk by Firm Size and Firm Growth



(a) By Firm Size



(b) By Firm Growth

- The (quarterly) probability of moving into unemployment is higher for workers in small- and shrinking-firms
- Cross-firm differences is more significant for low- than for high-wage workers
- Again variation across the past income distribution is much more pronounced compared to the firm variation.

Earnings Risk: Unobservable Types

K-Means Clustering

Clustering Firms and Workers w.r.t Earnings Change distribution

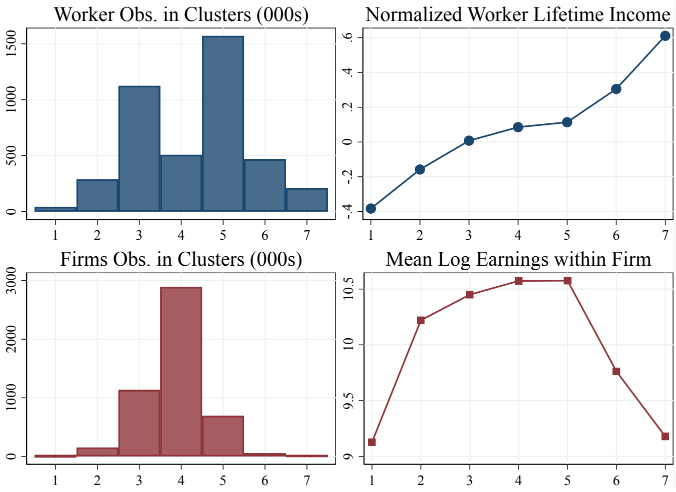
Previous results based on observable workers' and firms' characteristics

Empirical evidence shows significant variation of workers' wage *level* across *unobservable* characteristics (Abowd et al. 1999; Card et al. 2012; Bonhomme et al. 2018)

To account for unobservable heterogeneity we use *K-means clustering* algorithm grouping firms and workers of similar characteristics (as in Bonhomme-Lamadon-Manresa, 2018, 2020)

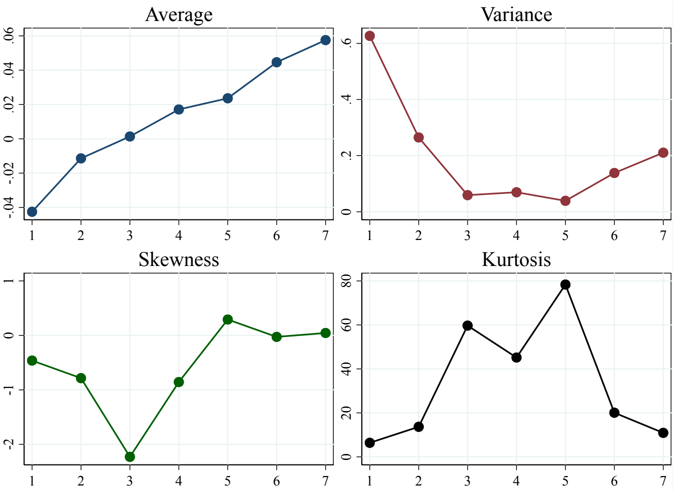
- We cluster firms using earnings growth distribution (percentiles) and unemployment risk
  - Require each firm to have at least 25 observations over a 5-year period
- We cluster workers using earnings growth distribution and unemployment risk
  - We require workers to have at least 20 observations over the life cycle ▸ K-means Details ▸ K-means Optimization

Number of Observations and Lifetime Incomes



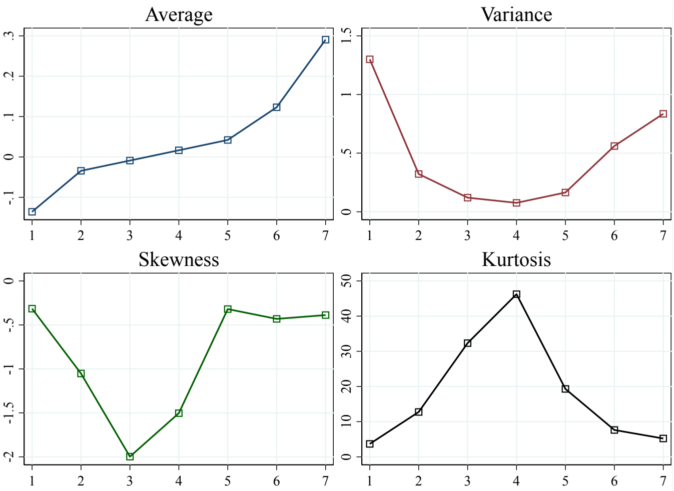
- Rank clusters by average earnings growth so #1 has the smallest growth.
- Workers are more equally distributed across clusters whereas most firm observations are in the middle 3 firm clusters.
- Workers in higher clusters have higher lifetime incomes.
- Workers in highest earnings growth does not have highest earnings.

Features of Income Dynamics by Person Clusters



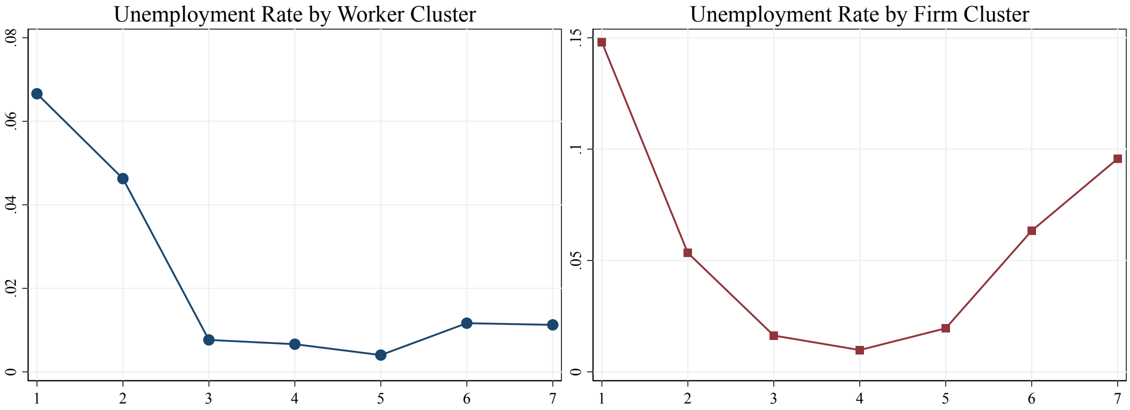
- We rank cluster of workers by average earnings growth
- A familiar pattern emerges:
  - Workers with higher earnings growth face less volatile earnings changes.
  - High- and low-growth workers face a more positively skewed and less leptokurtic distribution of earnings growth.

Features of Income Dynamics by Firm Clusters



- We rank cluster of firms by average earnings growth
- Similar patters seen for workers:
  - Workers in high- and low-growth firms experience higher wage dispersion,
  - a more positively skewed wage growth and lower kurtosis

Unemployment Risk by Worker and Firm Clusters



- Unemployment probability is significantly higher for workers in the first two clusters (similar to Karahan *et al.* (2019); Gregory *et al.* (2021)).
- The first and last firm cluster have a lot higher unemployment risk than other but they are small.

Two-sided Fixed Effect Regressions on Unemployment Risk

Model	1	2	3	4	5
N (millions)	3.0	3.0	3.0	3.0	3.0
Pseudo R2	0.005	0.093	0.15	0.17	0.18
Worker FE	No	No	Yes	Yes	Yes
Firm FE	No	Yes	No	Yes	Yes
Firm/Worker FE	No	No	No	No	Yes

We run AKM Probit regression on workers' employment-to-unemployment transition

$$Pr(EU_{ijt}) = \Phi\left(\alpha_i + \psi_{j(i,t)} + \gamma_{ij(i,t)} + \Gamma X_{it}\right),$$

where α, and ψ are worker and firm cluster fixed effects, and γ is their interaction, and X_{it} includes age, education and year dummies.

Worker fixed effects explains differences in unemployment risk better.

Sorting of Worker and Firm Across Clusters

- The figure shows the relative density of worker clusters (rows) in each firm cluster (column)
- The relatively darker colors in the diagonal cells point to some sorting in growth rates between firms and workers.

An Income Process with Worker and Firm Heterogeneity

$$
\begin{aligned}
\text{Log earnings:} \quad & y_t^i = g(X_{it}) + h_t^i + \psi^{k(j_t(i))} + z_t^i + \varepsilon_t^i \\
\text{Permanent productivity:} \quad & h_t^i = h_{t-1}^i + \beta^i + \gamma^{j_t(i)} \\
\text{Ex-ante heterogeneity:} \quad & h_0^i = \alpha^{l(i)}, \beta^i = \beta^{l(i)}, \gamma^{j_t(i)} = \gamma^{k(j_t(i))} \\
\text{Persistent component:} \quad & z_t^i = \rho z_{t-1}^i + \eta_t^i, \\
\text{Innovations to AR(1):} \quad & \eta_t^i \sim F^{\eta}_{l(i),k(j_t(i))} \\
\text{Initial condition of } z_0^i\text{:} \quad & z_0^i \sim F^{z_0}_{l(i)} \\
\text{Transitory shock:} \quad & \varepsilon_t^i \sim F^{\varepsilon}_{l(i),k(j_t(i))}
\end{aligned}
$$

- Workers are indexed by $i \in 1, \ldots, I$ and firms by $j \in 0, 1, \ldots, J$, where $j = 0$ reflects non-employment.
- Denote as $k = k(j)$ in $1, \ldots, K$ the class of firm j with $K \leq J$ and as $l = l(i)$ the class of firm $\in 1, \ldots, L \leq I$.
- Estimate this income process by targeting above moment.
- Use it to quantify the importance of worker and firm heterogeneity.

Conclusion

Summary of the Results

Summary of the Results

Workers at smaller and fast shrinking firms experience earnings changes with

- lower growth, higher dispersion, and lower kurtosis
- less negatively skewed in smaller firms, more negatively skewed in shrinking firms
- and higher unemployment risk

Clustering identify worker and firm types based on properties of income dynamics

- Workers with higher income growth face less volatile income changes
- High and low-growth workers face more positively skewed and less leptokurtic income growth
- These are similar to variation by recent earnings.

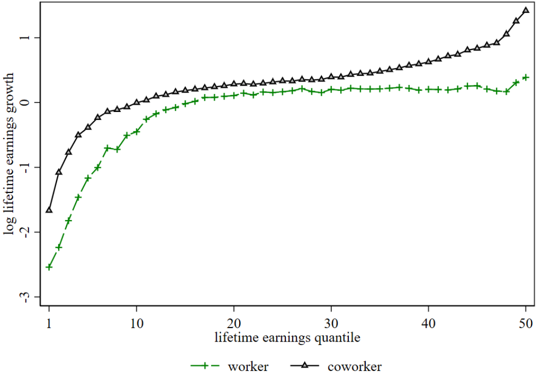
Main take away: firms are important for income dynamics

- but variation across workers characteristics seems to be larger than across firm characteristics

THANK YOU!

Lifetime Earnings Growth Heterogeneity

Earnings growth over the life cycle



- This figure shows lifetime earnings growth b/w 25-55 conditional on lifetime earnings quantile.
  - The baseline (worker's own earnings growth) vs using average firm earnings growth (using worker's colleagues earnings growth).
  -

K-means Algorithm: Details

Back

- K-means is a widely used and efficient clustering algorithm (Steinley (2006))
- K-means is used to partition dataset into K distinct groups/classes/ clusters: $C_1, C_2, ..., C_K$.
- The algorithm must satisfy the following conditions:
  - Specify K.
  - Each observation belongs to at least one of the K cluster: $C_1 \cup C_2 \cup ... \cup C_K = \{1, 2, ..., n\}$.
  - Non overlapping clusters: $C_k \cap C_{k'} = \oslash$ for all $k \neq k'$.
  - The squared Euclidean distance is used to measure the amount by which the observations within a cluster differ from each other (Dissimilarity metrics).
  - A good clustering aims to find homogeneous subgroups among the observations.

K-means Optimization

Back

- Let $W(C)$ measure the amount by which the observations within a cluster differ from each other.
- The discrete form analogue of our clustering procedure could be written as:

$$\min_{C_1,\ldots,C_K} \sum_{k=1}^{K} W(C_k) \tag{1}$$

- With $W(C_k) = \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{ij} - x_{i'j})^2$, $|C_k|$ is the number of observation in class k, x_{ij} is the value for individual i of the variable x_j used to construct the cluster.

Almost K^n ways to partition n observations into K clusters $\Rightarrow$ Difficult problem to solve.

Solution: Find a local minimum as follows:

1. Randomly put each observation in one of the cluster 1 to K.
2. For each of the K clusters, compute the cluster centroid (here the mean).
3. For each observation, compute the distance between each observation and each cluster centroid.
4. Assign each observation to the cluster whose centroid is closest.
5. Repeat step 2 to 4 until the cluster assignments stop changing.

Two-sided Fixed Effect Regressions on Wage Growth

Model	1	2	3	4	5
Worker FE	Yes	No	Yes	Yes	Yes
Worker trend	No	No	No	Yes	Yes
Firm FE	No	Yes	Yes	Yes	Yes
Firm trend	No	No	No	No	Yes
N (Millions)	2.9	2.9	2.9	2.9	2.9
adj. R2	0.028	0.028	0.032	0.033	0.037
Decomposition: Share of Total Variance					
var(pe)	0.006		0.005	0.055	0.053
var(fe)		0.006	0.004	0.004	0.010
var(xb)	0.022	0.022	0.022	0.064	0.072
var(res)	0.972	0.972	0.968	0.966	0.963
2*cov(pe,fe)			0.001	0.003	0.007
2*cov(pe,xb)	-0.000		-0.001	-0.091	-0.091
2*cov(xb,fe)		-0.000	-0.000	-0.003	-0.013

- We run standard fixed regression on workers' wage growth

$$\Delta w_{ijt} = \alpha_i + \beta_{it} + \psi_{ij} + \gamma_{jt} + \varepsilon_{ijt},$$

where β_{it} and γ_{it} are worker -and firm-specific trends

- Here i is a worker cluster and j is a firm cluster

References i

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

Bagger, J., Fontaine, F., Postel-Vinay, F. and Robin, J.-M. (2014). Tenure, experience, human capital and wages: A tractable equilibrium search model of wage dynamics. *American Economic Review*, **104** (6), 1551–1596.

Baker, M. (1997). Growth-rate heterogeneity and the covariance structure of life-cycle earnings. *Journal of Labor Economics*, **15** (2), 338–375.

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

Card, D., Heining, J. and Kline, P. (2013). Workplace heterogeneity and the rise of west german wage inequality. *The Quarterly journal of economics*, **128** (3), 967–1015.

Gregory, V., Menzio, G. and Wiczer, D. G. (2021). *The Alpha Beta Gamma of the Labor Market*. Tech. rep., National Bureau of Economic Research.

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

—, Karahan, F., Ozkan, S. and Song, J. (2018). *What do data on millions of U.S. workers reveal about life-cycle earnings risk?* Tech. rep., National Bureau of Economic Research.

Hall, R. E. and Kudlyak, M. (2019). *Job-finding and job-losing: A comprehensive model of heterogeneous individual labor-market dynamics*. Tech. rep., National Bureau of Economic Research.

Jarosch, G. (2015). *Searching for Job Security and the Consequences of Job Loss*. Working paper, Princeton University.

Karahan, F., Ozkan, S. and Song, J. (2019). Anatomy of lifetime earnings inequality: Heterogeneity in job ladder risk vs. human capital. *FRB of New York Staff Report*, (908).

Lentz, R., Piyapromdee, S. and Robin, J.-M. (2018). On worker and firm heterogeneity in wages and employment mobility: Evidence from anish register data. *PIER Discussion Papers*, **91**.

Meghir, C. and Pistaferri, L. (2004). Income variance dynamics and heterogeneity. *Econometrica*, **72, No.1**, 1–32.

Oberschachtsiek, D., Scioch, P., Seysen, C., Heining, J. *et al.* (2010). *Integrated employment biographies sample IEBS: handbook for the IEBS in the 2008 version*. Tech. rep., Institut für Arbeitsmarkt-und Berufsforschung (IAB), Nürnberg [Institute for ?

Steinley, D. (2006). K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, **59** (1), 1–34.