

Misallocation and Inequality*

Nezih Guner[†]

UAB, ICREA, BSE

Alessandro Ruggieri[‡]

University of Nottingham

September 30, 2022

Abstract

For a large set of countries, we document how labor earnings inequality varies with GDP per capita. As countries get richer, the mean-to-median ratio and the Gini coefficient decline. Yet, this decline masks divergent patterns: while inequality at the top of the earnings distribution falls, inequality at the bottom increases. We interpret these facts within a model economy with heterogeneous workers and firms, featuring industry dynamics, search frictions, and skill accumulation of workers through on-the-job learning and training. The benchmark economy is calibrated to the UK. We then study how the earnings distribution changes with distortions that penalize high-productivity firms and frictions that reduce match formation. Distortions and frictions reduce employment, average firm size, and GDP per capita. They also affect how much firms are willing to pay workers, how well high-skill workers are matched with high-productivity firms, and how much training workers receive. The model generates the observed cross-country relation between GDP per capita and earnings inequality, as well as a host of cross-country facts on firm size distribution, firms' training decisions, and workers' life-cycle and job tenure earnings profiles.

Keywords: earnings inequality, labor market frictions, correlated distortions, human capital, on-the-job training, productivity, firm size, life-cycle earning profiles

JEL Classification: E23, E24, J24, O11

*We would like to thank Jaime Arellano-Bover, Niklas Engbom, David Lagakos, Markus Poschke, Todd Schoellman, and seminar and workshop participants at the NBER Summer Institute, BSE Summer Forum, CEMFI, Georgetown University, LSE, St.Louis FED, University of Alicante, University of Nottingham and University of Kent for helpful comments. We also thank Juan Carlos Ruiz Garcia for his help with the data. Siqi Wei provided excellent research assistance. Guner acknowledges financial support from the Spanish Ministry of Science and Innovation Grant PID2019-111694GB-I00. The usual disclaimers apply.

[†]Universitat Autònoma de Barcelona, Cerdanyola del Valles, 08193, Spain, gunermail@gmail.com

[‡]University of Nottingham, Nottingham, NG72RD, United Kingdom, aruggierimail@gmail.com

1 Introduction

How does the distribution of labor earnings change with development? We answer this question using household surveys from around the world. As a country gets richer, the distribution of earnings shifts to the right, and the mean increases. Yet, the median increases even more, and the mean-median ratio falls. The Gini coefficient declines as well. However, not all inter-percentile ratios decline; while the p90-p50 ratio drops, the p50-p10 ratio increases. Hence, as countries get richer, the inequality at the bottom of the distribution increases, whereas it declines at the top.

We interpret these facts through the lens of a model economy with heterogeneous workers and firms. The model economy has three key ingredients. First, different firms pay different wages to workers with similar skills (Abowd et al., 1999; Card et al., 2013; Song et al., 2019). Identical workers receive higher wages in larger and more productive firms. Second, due to labor market frictions, matching between high-skilled workers and high-productivity firms is not instantaneous (Lise et al., 2016). Finally, firms differ in how much on-the-job training they provide.

In the benchmark economy, workers who differ by their initial (or pre-market) human capital levels search for firms in a frictional labor market. Some match with firms, while others remain unemployed and keep looking for a job. Firms are heterogeneous; they differ in their productivity and training costs. Hence, a worker who matches a high-productivity firm with low training costs enjoys high earnings and high earnings growth. Workers accumulate skills with job tenure and on-the-job training, and they lose them during non-employment spells. Through workers' and firms' dynamics, and workers' human capital accumulation, the model economy generates a host of facts that can be confronted with the data. The parameters of the model are estimated using firm- and worker-level data from the UK. The model replicates the observed firm size distribution, worker's earnings profiles, and training provision across different firms. It also produces a positive and large firm-size wage premium.

We then turn to cross-country differences. We assume that countries differ along two dimensions. First, following recent literature on misallocation (Guner et al., 2008; Restuccia and Rogerson, 2008; Hsieh and Klenow, 2009), we introduce distortions, taking the form of output wedges, that are correlated with firm size. These distortions are more extensive in some countries than others. The existing literature has focused on how misallocation affects cross-country differences in firm-size distribution and aggregate productivity. We focus on how misallocation affects earnings inequality. Firms that face distortions shrink and pay lower wages. Furthermore, size-correlated distortions make high and low-productivity firms more similar, compressing the overall earnings distribution.

We interpret these distortions broadly as regulations and market imperfections that disproportionately affect larger firms and hinder firm growth (see [Hopenhayn \(2014\)](#) and [Restuccia and Rogerson \(2017\)](#) for reviews). They might capture existing size-dependent policies, such as labor market regulations, or result from discretionary interventions by the government in financial markets. Finally, they can reflect the lack of well-defined property rights, as in [Akcigit et al. \(2021\)](#).¹

Second, we assume that countries also differ in the extent of labor market frictions. Some have a more efficient labor market, and workers and firms match easily, while in others, it takes longer to fill a vacancy or find a job. Higher frictions result in lower employment and higher non-employment duration. Search frictions also affect the equilibrium earnings distribution. Longer times to fill a vacancy prevent workers from accumulating human capital. It also makes workers less willing to wait for the right firms, reducing positive assortative matching between firms and workers. The link between search frictions and misallocation of labor has been recently emphasized by [Martellini and Menzio \(2021\)](#). [Poschke \(2019\)](#) shows that search frictions can account for cross-country differences in unemployment, wage employment, and self-employment. [Heise and Porzio \(2021\)](#) estimate a model of frictional labor market and show that spatial frictions generate misallocation across and within regions and affect the wage distribution. [Donovan et al. \(2020\)](#) highlights the role of labor market frictions and endogenous separations to explain how labor market flows and wage-tenure profiles vary with development. Finally, [Gulyas et al. \(2018\)](#) shows that misallocation of jobs and workers across firms due to search frictions generate significant output losses.

Distortions and frictions also affect on-the-job training provision in the model. We document that the share of establishments providing training and the share of workers receiving training in a firm increases with GDP per capita. Furthermore, the probability that a firm provides on-the-job training and the share of trained workers increases with establishment size. In the model, distortions and frictions directly reduce the gain from training in a given match, making firms less willing to incur training costs. Empirically, the importance of training for life-cycle inequality has been emphasized recently, among others, by [Gregory \(2021\)](#) and [Arellano-Bover and Saltiel \(2021\)](#). [Lentz and Roys \(2015\)](#) study human capital accumulation in a search model where firms can offer long-term contracts and commit to provide a certain level of expected utility to their employees. They show that increased labor market frictions reduce training in equilibrium. [Flinn et al. \(2017\)](#) estimate a search model

¹Beyond size-dependent distortions, financial frictions constitute another candidate for cross-country differences in firm size distribution and aggregate productivity ([Buera et al. \(2011\)](#), [Midrigan and Xu \(2014\)](#), [Moll \(2014\)](#), and [Gopinath et al. \(2017\)](#)). [David and Venkateswaran \(2019\)](#) try to disentangle different sources of misallocation. [David et al. \(2016\)](#) focus on the role of imperfect information.

of investment in general and match-specific human capital and use it to examine the impact of minimum wage policy on training provision.

We calibrate the correlated distortions to target average firm size and the matching efficiency to target the share of wage and salary workers for different countries. Lower distortions or frictions alone would increase firm size, employment, and GDP per capita in the model economy. But to match the observed cross-country differences, we need both: while distortions help us generate observations for the poorer countries, frictions help us generate observations for the richer ones.

Then we zoom into these economies and calculate inequality measures. Cross-country differences in earnings inequality that emerge from this exercise are remarkably close to what we observe in the data. The mean-to-median ratio and the Gini index declines as countries get richer. In the data, the p50-p10 earnings ratio increases by GDP per capita. In the model economy, higher distortions push p50-p10 down in poor countries and lower frictions push it up in rich ones. In the data, the p90-p50 earnings ratio falls by GDP per capita. In the model, this is generated by lower matching frictions in richer countries.

The model also fits a large set of cross-country facts on firm-size distributions, life-cycle earning profiles, and training outcomes. In particular: i) Together with average firm size, the dispersion and skewness of the firm-size distribution increase with GDP per capita ([Hopenhayn, 2016](#); [Bento and Restuccia, 2017](#); [Poschke, 2018](#)). ii) On-the-job training provision increases with development. iii) Earnings-experience profile becomes steeper with GDP per capita ([Lagakos et al., 2018](#)). iv) On the other hand, the earnings-tenure profile becomes flatter with GDP per capita ([Donovan et al., 2020](#)).

What are the mechanisms behind these patterns? We identify three forces. First, lower distortions make the relation between firm productivity and revenue steeper, widening the entire wage distribution and increasing p50-p10 and p90-p50 in high-income countries. Second, lower frictions reduce non-employment duration, increasing human capital accumulation and wages. But the impact is more substantial for relatively lower-skilled workers, leading to lower earnings inequality. Third, lower frictions also improve sorting between firms and workers. The sorting-effect increases wages for relatively high-skilled workers, who would work for a low-productivity firm in a less efficient labor market, increasing inequality. Quantitatively, the impact of second and third forces results in a higher p50-p10 and a lower p90-p50 in richer countries, as workers in the middle of the skill distribution benefit the most.

On-the-job training amplifies these patterns as it helps the workers in the middle of the skill distribution. In particular, in a richer country with lower distortions and frictions, the training of the workers in the middle of the skill distribution increases the most. Training

low-skilled workers might not cover the cost of training. On the other hand, training high-skilled workers improves their outside options and makes them likely to leave. As a result, the relation between workers' skills and training decisions becomes hump-shaped when distortions and frictions are lower. We find that training accounts up to 35% of the decline in the mean-to-median earnings ratio across countries. Finally, following [Alfonsi et al. \(2020\)](#), we evaluate a fully-subsidized training program for unemployed workers in a low-income country. The program increases employment by about 12% points, reduces earnings inequality, and generates enough increase in output to cover its cost.

While our focus on the interaction between misallocation and earnings inequality is novel, different elements of the model have been emphasized by the existing literature. [Bento and Restuccia \(2017\)](#) introduce correlated distortions into a competitive model of industry dynamics to account for cross-country differences in average firm size. [Guner et al. \(2018\)](#) document that, for a group of high-income countries, earnings of managers grow faster than for non-managers, and the earnings growth of managers relative to non-managers corresponds to output per worker. They interpret this finding within a span-of-control model where managers can invest in their skills. Hence, distortions not only affect average firm size, but also the accumulation of managerial skills. [Jovanovic \(2014\)](#) develops a model of growth with human capital accumulation where incomplete information on workers' ability generates worker-firm mismatch. He shows that better signals lead to a more efficient worker-firm assignment which, in turn, leads to higher human capital accumulation, faster long-run growth, and more income inequality. Finally, [Hsieh et al. \(2019\)](#) focus on misallocation of talent by gender and race in the US and find that the improved allocation of human capital across jobs can account for between 20% and 40% of income per capita growth in the last 50 years.

The link between labor market frictions and incentives of workers to invest in their skills has been studied by [Engbom \(2020\)](#). He shows that wages grow more over the life-cycle in countries where job-to-job mobility is more common. He then builds a life-cycle model of on-the-job training and job-to-job transitions where a more fluid labor market allocates workers to firms more efficiently and provides larger incentives for skill accumulation. [Karahan et al. \(2022\)](#) estimate a job ladder model with on-the-job learning and show that differences in life-time wage growth in the U.S. can be attributed to heterogeneity in job loss, job finding, and contact rates. Along similar lines, [Ma et al. \(2021\)](#) explore the role of firm-provided training in explaining why workers in richer countries have faster rates of wage growth over their lifetimes than workers in poorer countries. They find on-the-job training can explain between 10% and 15% of the income differences across countries

2 Cross-Country Facts

2.1 Earnings Distribution

This section documents how the distribution of earnings varies with GDP per capita across countries. The results are based on household surveys from 57 countries between 1974 and 2016, for a total of 502 country-year observations. The primary data sources are IPUMS International, European Union Survey on Income and Living Conditions (EU-SILC), and Luxembourg Income Study Database (LIS). The poorest country in our dataset is India in 1993, with a GDP per capita of 1,845 in 2011 USD, while the richest one is Luxembourg in 2007, with a GDP per capita of 97,864 in 2011 USD.²

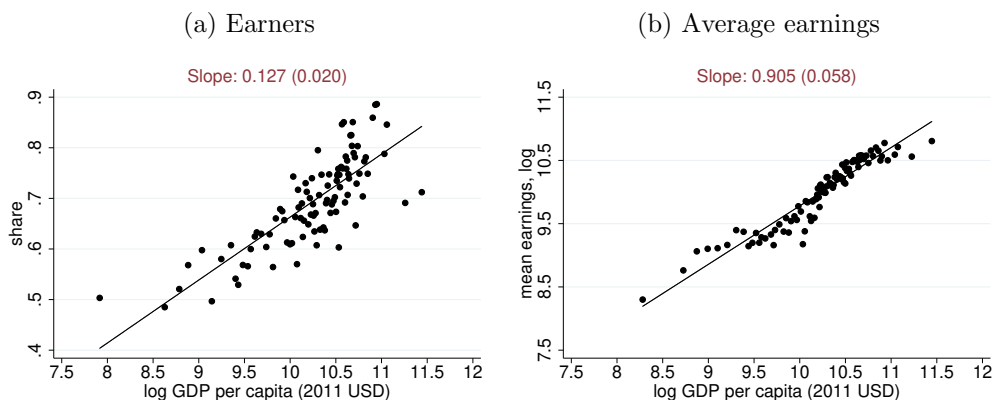
We restrict the sample to all individuals between 18 and 64 who are not students and have non-missing information on their wage and salary income. For each individual, we then calculate total gross wage and salary income (referred as *earnings* below), which includes any payment received as an employee. For each country-year, wage and salary earners (referred as *earners* below) consist of those with strictly positive wage and salary income. Hence, any earner in the sample has a labor relationship with an employee and receives payments from this relationship. The employees can be private or public; they can also be formal or informal. We label everyone with a zero earnings as a no-earner, including those who are out of the labor force, unemployed, unpaid family workers, or self-employed. We then study how the share of earners and the distribution of earnings change by GDP per capita.

Figure 1 shows how the share of earners (panel a) and the average log earnings (panel b) change with (log) GDP per capita across countries. To construct Figure 1, we first regress the outcome of interest for country i and year t on a set of year dummies. Each dot corresponds to average residuals from this regression for countries in a specific GDP-per-capita bin. We report 100 bins corresponding to the percentiles of the GDP-per-capita distribution. As we move from poorer to richer countries, workers become significantly more likely to work as employees and report positive earnings. Since our focus is on earnings inequality, we do not include very poor countries in the sample where a majority of workers do not receive any wage and salary income. The share of earners ranges from around 50 percent to almost 90 percent in the sample. Not surprisingly, the average log earnings increase almost one-to-one with log GDP per capita (panel b).

Figure 2 documents earnings inequality. As in Figure 1, each dot corresponds to the average values of the dependent variable (different inequality measures) for countries in a

²Details on these three datasets (sample restrictions and variable definitions) are provided in Appendix A.1.

Figure 1: Wage and salary earners and earnings across countries



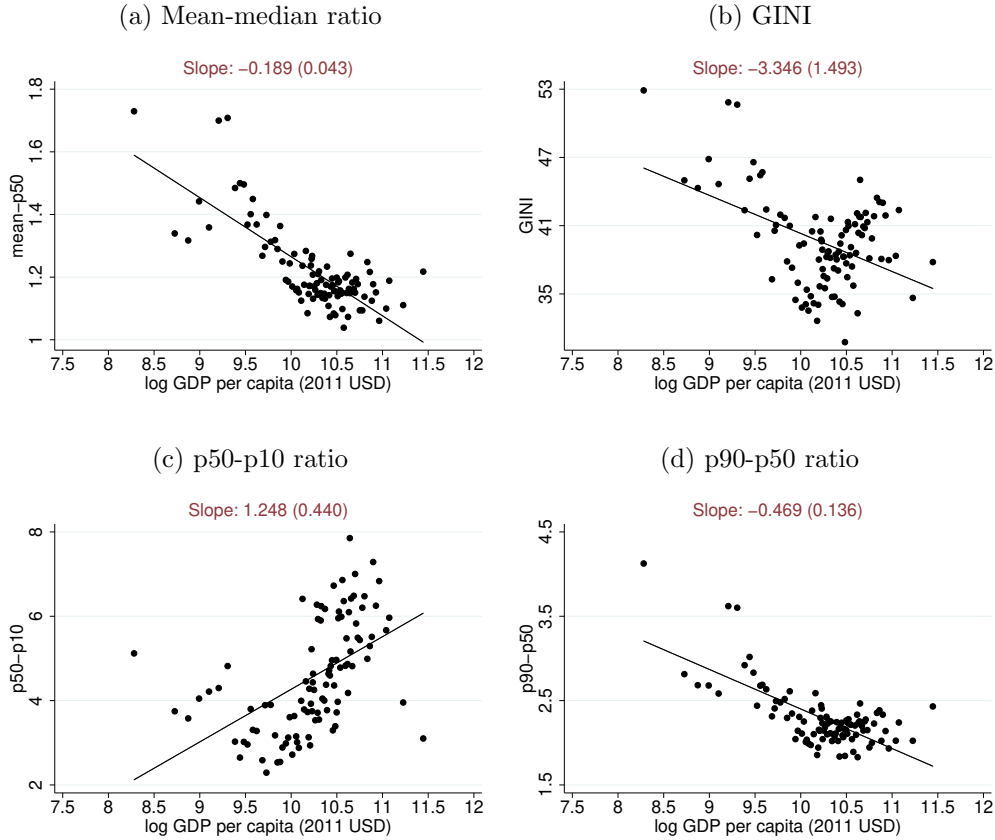
Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year-fixed effects. In red we report the estimated slope in the regression. Standard errors (in parenthesis) are robust and clustered at country level. Source: IPUMS, EU-SILC, LIS and author’s calculations.

specific bin of GDP per capita, after removing year-fixed effects. Panel (a) shows that the mean-to-median ratio declines significantly as countries get richer. It drops from around 1.6 for the poorest countries in the sample to about 1.1 for the richest ones. Hence, as countries get richer and the mean earnings increase, the median workers gain even more. The gains of the workers at the center of the earnings distribution also lower the Gini coefficient, as illustrated in panel (b). As we move from the sample’s poorest to the richest country, the Gini declines by around 15 percentage points, from 50% to 35%.

However, the decline in these two inequality measures masks significant heterogeneity in how the earnings distribution changes with development. Panel (c) of Figure 2 shows that the lower tail of the earnings distribution does not catch up with the median, and the gap between the bottom and the median opens up: the 50-to-10 ratio increases from around 3 in poor countries to about 8 in the richest ones. Yet, the opposite happens in the upper tail, as shown in panel (d). Labor income for workers in the 90th percentile does not grow as fast as the median, and the 90-to-50 ratio declines as countries get richer.

In Appendix A.2, we show that these findings are robust. They hold when we control not just for year dummies but also for cross-country differences in average years of schooling, women’s labor force participation rate, share of self-employment, shares of agricultural and industrial employment, the average number of hours worked per year, and value of trade over GDP. They also hold when we restrict the sample to workers employed in specific sectors (non-agriculture or industry) or workers with or without a college degree. We also find the same patterns when the sample is restricted to males, household heads, or prime-age workers

Figure 2: Earnings inequality across countries



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year-fixed effects. In red we report the estimated slope in the regression. Standard errors (in parenthesis) are robust and clustered at country level. Source: IPUMS, EU-SILC, LIS and author's calculations.

(25 to 55). Finally, we show that independent of the particular cut-off we use (40-to-10 versus 80-to-50 ratios, or 50-to-20 versus 90-to-60 ratios), earnings in lower and upper tails grow much slower than those in the center of the distribution.³

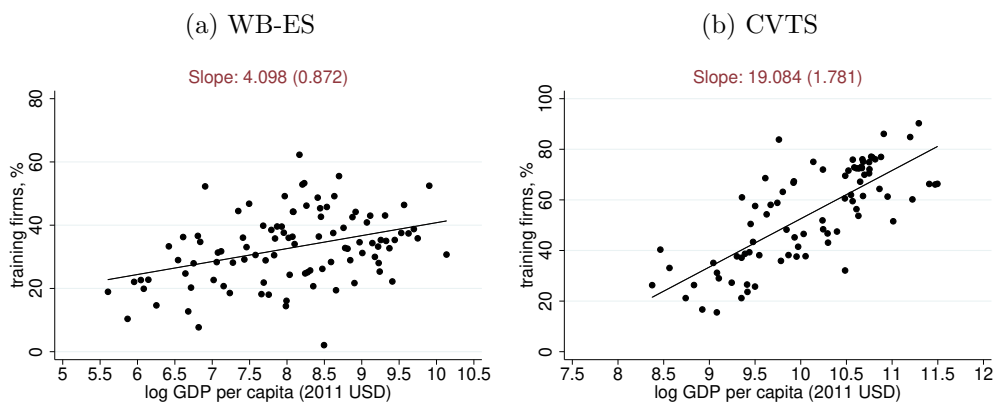
2.2 On-the-Job Training

This section complements the cross-country evidence on earnings inequality with facts on on-the-job training provision. To this end, we use data from the World Bank Enterprise Sur-

³Surveys are subject to top-coding, which is likely to be an issue when measuring the very top of the earnings distribution (Carranza et al., 2022). However, mismeasurement for high earners is primarily due to self-employment incomes (Yonzan et al., 2022), hence it is unlikely to alter our measures of wage earnings inequality.

vey (WB-ES, henceforth) and the Eurostat Continuing Vocational Training Survey (CVTS, henceforth).⁴

Figure 3: Training provision across countries



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year-fixed effects. In red we report the estimated slope in the regression. Standard errors (in parenthesis) are robust and clustered at country level. Source: World-Bank Enterprise Survey (WB-ES) and Eurostat Continuing Vocational Training Survey (CVTS).

Figure 3 shows how the percentage of firms offering job training to their employees varies with (log) GDP per capita across countries. Richer countries have a larger share of firms investing in job training. The correlation between the percentage of firms providing job training and log GDP per capita is equal to 0.52 in the WB-ES data (panel a). The coefficient from regressing the former on the latter implies that one log point higher GDP per capita is associated with a 4 percentage points more firms providing training. The correlation in the CVTS data is even higher, 0.75. The slope coefficient suggests that one log point higher GDP per capita is associated with 19 percentage points more firms offering training (panel b). In the Appendix A.3 we show that the share of workers receiving training in a given firm also increases with GDP per capita in both datasets.

Next, we show how training varies by firm size within each country. Table 1 reports the percentage of firms providing training by different firm size categories, separately for countries belonging to different regions; Latin America (LAC), Middle East and Africa (ME+AFR), and Asia in the WB-ES sample, and EU15 and non-EU15 in the CVTS sample. Training increases significantly with firm size. The share of firms investing in training more than doubles as we move from firms with less than 50 employees to more than 250 employees. This difference is robust across regions, and it is higher in countries belonging to the WB-ES

⁴Details on these two datasets are provided in Appendix A.3.

Table 1: Job training across firm size

WB-ES	LAC	ME+AFR	ASIA	Training firms, %		EU15	non-EU15
				others	CVTS		
Firm size (# employees)					Firm size (# employees)		
<20	34.84	18.42	19.32	26.35			
20-49	54.31	31.99	33.63	38.48	<49	60.27	40.71
50-249	66.94	41.31	47.02	46.47	50-249	82.37	63.94
250-449	81.13	56.86	47.32	56.65	≥ 250	93.68	84.17
≥ 500	92.12	68.45	52.28	68.88			

Notes: Each entry denotes to the percent of firms that reports to provide on-the-job training, separately for firms with different size (number of employees), and different groups of country. Firm size refers to number of employees. Source: World-Bank Enterprise Survey (WB-ES) and Eurostat Education and Training Dataset (CVTS).

sample. In Appendix A.3 we also show that conditional on investing in training, larger firms train a larger share of their workforce.

2.3 Recap

This section documented three key cross-country patterns. As countries get richer: 1. The mean-to-median ratio and Gini coefficient for earnings decline. 2. While the p50-p10 ratio increases, the p90-p50 ratio declines. 3. On-the-job training provision (measured by the share of establishments that provide training and the share of workers receiving training within a given firm) increases.

These facts complement known cross-country patterns in the firm size distribution and life-cycle earnings growth. As countries get richer: 1. Average firm size increases. 2. Dispersion and skewness of firm size distribution increase. 3. Age-earnings profiles become steeper. 4. Tenure-earnings profiles become flatter.

In the next section, we develop a model of heterogeneous firms and workers and use it to understand these patterns.

3 Model

Consider a closed economy populated by two types of agents: a unitary measure of heterogeneous workers and an endogenous measure of heterogeneous firms. Time is discrete. Workers

can live forever, but each period faces a constant probability of death (or retirement). Workers enter the economy with a given level of human capital (skill or ability). Each period they can be employed or non-employed.⁵ Labor market frictions are represented by a matching function that maps non-employed workers and open vacancies into potential matches. If a match between a worker and a firm is formed, workers' skills grow due to on-the-job learning and training. In contrast, non-employment lowers workers' skills. Firms differ along four dimensions: productivity, cost of training, the total number of employees, and skill distribution of their employees. Finally, firms face size-dependent output distortions (wedges) that are correlated with their productivity.

3.1 Workers

Workers maximize the expected present value of their lifetime utility

$$\mathcal{U} = \sum_{t=0}^{\infty} \left(\frac{1 - \delta_w}{1 + r} \right)^t c_t,$$

where c_t is consumption, $r > 0$ is a discount rate, and $\delta_w > 0$ is an exogenous probability of death (or retirement).

Workers are ex-ante heterogeneous in their initial level of human capital, denoted by $a^0 \in \mathcal{A} = \{a_0, a_1, \dots, a_A\}$, distributed according to the probability distribution, $\psi_a(a)$. Upon matching with a firm, workers improve their skills through job experience (on-the-job learning) and on-the-job training. Job experience and training cause one-step jumps in \mathcal{A} with probabilities p^e and p^t , respectively. Human capital is fully portable between jobs, so when a job is destroyed, workers retain fully their human capital. But, each period of non-employment induces one-step depreciation of skill a with probability p^d .

3.2 Firms

The industry is populated by an endogenous measure of firms, each producing a homogeneous good and characterized by a firm-specific productivity $z \in \mathcal{Z} \subset \mathcal{R}_+$. The productivity level z is drawn before entry from a probability density function, $\psi_z(z)$, and remains constant as long as the firm is active. Firms differ also by the cost they incur to train their workforce. Let $\xi \in \mathcal{E} \subset \mathcal{R}_+$ denote the per-period cost to train one worker, defined in units of final

⁵We lump self-employed, unemployed and inactive workers in a single labor market state. This approach is consistent with recent evidence on labor rationing for developing countries, where a large portion of excess labor supply is disguised as self-employment (Breza et al., 2021).

output. Like productivity, the training cost is firm-specific, it is drawn before entry from a probability density function $\psi_\xi(\xi)$, and it is time-invariant.

To produce, firms combine labor services (expressed in efficiency unit) from its employees through a linear production technology. Let $\psi(i|z, \xi, \ell)$ be the measure of worker i in a firm with productivity z , training costs ξ and ℓ workers. Then, we can write total firm output as

$$y(z, \xi, \ell, \psi) = \int_0^\ell g(z, i) \psi(i|z, \xi, \ell) di,$$

where $g(z, i) = \kappa z a(i)$ is the amount produced by a match between a firm z and a worker i with human capital $a(i)$, and $\kappa > 0$ is the aggregate productivity. Re-arranging terms, we can write the production function as

$$y(z, \xi, \ell, \psi) = \kappa z \bar{a}(z, \xi, \ell, \psi) \ell, \tag{1}$$

with

$$\bar{a}(z, \xi, \ell, \psi) = \int_0^1 a(i) \psi(i|z, \xi, \ell) di.$$

Linearity of the aggregate production function with respect to ℓ implies that each firm, independent of their productivity z , would like to hire as many workers as possible, and, as it will become clear below, are only constrained in their hiring by matching frictions and adjustment costs. This makes the problem tractable since a firm treats each of its workers as independent production units. As a result, wage bargaining and training decision take place between each worker and their employer separately. Finally, each period firms face two types of destruction shocks. They can lose a particular worker with probability δ_s , or lose all workers and exit with probability δ_f .

3.3 Distortions

Firms are subject to output distortion. Distortions are modeled as in [Bento and Restuccia \(2017\)](#) and [Guner et al. \(2018\)](#). Each firm retains a fraction $1 - \tau$ of its output, where τ depends on firm-level productivity z , given by

$$\tau(z) = 1 - z^{-\zeta}, \quad \zeta > 0. \tag{2}$$

The parameter ζ is the elasticity of a firm's distortion with respect to its productivity. This formulation implies that the net revenue function for a worker-firm pair is

$$r(z, a) = \kappa z^{1-\zeta} a. \tag{3}$$

3.4 Frictions

The labor market is subject to search and matching frictions. To hire workers, firms need to post vacancies. To find a job, workers need to search, which is random and costless. There is no on-the-job search. The number of new matches depends on the total measure of workers searching for a job, U , and the vacancies posted, v . New matches are formed according to a constant return to scale matching function, given by

$$m(U, v) = \chi \frac{Uv}{(U^\eta + v^\eta)^{\frac{1}{\eta}}}, \quad \chi > 0, \eta > 0.$$

The parameters χ and η govern the efficiency of the matching and the elasticity of new matches with respect to the pool of searchers. This matching function implies the following contact rates for workers and for firms:

$$\phi_w = \frac{\chi}{(1 + \theta^{-\eta})^{\frac{1}{\eta}}} \quad \text{and} \quad \phi_f = \frac{\chi}{(1 + \theta^\eta)^{\frac{1}{\eta}}},$$

where $\theta = v/U$ is the labor market tightness. Workers matched with a firm earn a wage equal to $w(z, \xi, a)$, which results from bargaining between the workers and the firms and depends on the productivity of the firm they work, the training costs faced, and their human capital. Workers who fail to get matched end up being non-employed, supporting themselves by means of home production, equal to b .

3.5 The Problem of the Worker

3.5.1 Value of Non-employment

The value of being not-employed at the beginning of period for a worker with ability a is equal to

$$\begin{aligned} J^u(a) = & (1 - \phi_w)[p^d J^{u,h}(a-1) + (1 - p^d) J^{u,h}(a)] \\ & + \phi_w \int_{z \in \mathcal{Z}} \int_{\xi \in \mathcal{E}} [\mathbf{1}^h(z, \xi, a) J^{e,h}(z, \xi, a) + (1 - \mathbf{1}^h(z, \xi, a)) J^{u,h}(a)] \psi_v(z, \xi) d\xi dz, \end{aligned}$$

where $\mathbf{1}^h(z, \xi, a)$ is an indicator function for match formation (*hiring*). Non-employed workers fail to match with a firm with probability $(1 - \phi_w)$ and remain without a job for the period. Non-employment can result in lower skills with probability p^d . The value of being non-employment at the end of the period, $J^{u,h}(a)$, is given by

$$J^{u,h}(a) = b + \frac{(1 - \delta_w)}{1 + r} J^u(a). \tag{4}$$

With probability ϕ_w the worker matches with a firm and takes a random draw from $\psi_v(z, \xi)$, the distribution of vacancies posted by firms with productivity z and training cost ξ , which is endogenously determined. When a worker and firm are matched and there is a positive surplus, $\mathbf{1}^h(z, \xi, a) = 1$ and employment takes places. Otherwise, a match is not formed, and the worker stays unemployed. The function $J^{e,h}(z, \xi, a)$ is the end-of-period value of employment for a worker with skill a in a firm with productivity z and training costs ξ .

3.5.2 Value of Employment

The value of being employed at the beginning of the period in a firm with productivity z and training costs ξ for a worker with skill a is equal to:

$$J^e(z, \xi, a) = \mathbf{1}^h(z, \xi, a)J^{e,h}(z, \xi, a) + (1 - \mathbf{1}^h(z, \xi, a))J^{u,h}(a). \quad (5)$$

If the surplus is positive, the value of employment is given by

$$\begin{aligned} J^{e,h}(z, \xi, a) &= w(z, \xi, a) + \frac{(1 - \delta_w)}{1 + r}(\delta_f + (1 - \delta_f)\delta_s)J^{u,h}(a) \\ &+ \frac{(1 - \delta_w)}{1 + r}(1 - (\delta_f + (1 - \delta_f)\delta_s))[p^h(z, \xi, a)J^e(z, \xi, a + 1) - (1 - p^h(z, \xi, a))J^e(z, \xi, a)], \end{aligned}$$

Note that $p^h(z, \xi, a) = p^e + \mathbf{1}^t(z, \xi, a)p^t$ sums the probability of an improvement in a due to experience and training, where $\mathbf{1}^t(z, \xi, a)$ is an indicator function for job-training provision.

3.6 The Problem of the Firm

3.6.1 Value of an Active Match

Consider a match between a type- (z, ξ) firm and a worker with ability a . The value of this match for the firm at the beginning of the period is

$$V(z, \xi, a) = \mathbf{1}^h(z, \xi, a)V^h(z, \xi, a), \quad (6)$$

with

$$\begin{aligned} V^h(z, \xi, a) &= r(z, a) - w(z, \xi, a) \\ &+ \frac{1 - \delta}{1 + r} [-\mathbf{1}^t(z, \xi, a)\xi + p^h(z, \xi, a)V(z, \xi, a + 1) + (1 - p^h(z, \xi, a))V(z, \xi, a)]. \end{aligned}$$

and $\delta = \delta_w + (1 - \delta_w)\delta_s + (1 - \delta_w)(1 - \delta_s)\delta_f$. A worker-firm match produces $r(z, a)$, as defined in equation (3), and the worker is paid $w(z, \xi, a)$, which is defined below. Next period, any active job can be destroyed due to death/retirement by the worker (δ_w), exogenous

destruction of particular job (δ_s), or exogenous destruction of the firm (δ_f). If the match is destroyed due to δ_w or δ_s , the firm keeps its remaining matches, while in case of exit all the matches are destroyed and the firm disappears. An active job can also be destroyed endogenously, if the value of match is low enough and $\mathbf{1}^h(z, \xi, a) = 0$. Finally, if training takes place, i.e. $\mathbf{1}^t(z, \xi, a) = 1$, the firm incurs into a training cost, ξ .

3.6.2 Vacancy Posting

Firms choose the amount of vacancies $v(z, \xi)$ to maximize the total value of new hires subject to convex costs, $c(v)$, given by

$$c(v) = \frac{v^{\lambda_1}}{\lambda_1}, \quad \lambda_1 > 1,$$

where λ_1 governs the degree of convexity in the cost function. Each period, the problem of a firm reads as follows:

$$\pi(z, \xi) = \max_{v(z, \xi) \geq 0} v(z, \xi) \phi_f \sum_{a \in \mathcal{A}} \mathbf{1}^h(z, \xi, a) V^h(z, \xi, a) \psi_a^u(a) - c(v(z, \xi)), \quad (7)$$

where $\psi_a^u(a)$ is the endogenous distribution of skills for unemployed workers. A firm posting $v(z, \xi)$ vacancies gets in contact with $v(z, \xi) \phi_f$ unemployed workers. Each match with a positive surplus is valued as $V^h(z, \xi, a)$. The first order condition implies the following vacancy posting rule for a type- (z, ξ) firm

$$v(z, \xi) = \left(\phi_f \sum_{a \in \mathcal{A}} \mathbf{1}^h(z, \xi, a) V^h(z, \xi, a) \psi_a^u(a) \right)^{\frac{1}{\lambda_1 - 1}}.$$

The amount of new hires for a firm- (z, ξ) are then given by $v(z, \xi) \phi_f \sum_{a \in \mathcal{A}} \mathbf{1}^h(z, \xi, a) \psi_a^u(a)$.

3.6.3 Entry

In equilibrium, the measure of firms is determined by entry decisions. Each period, a fixed measure of potential entrants, M_e , draw a productivity z and a training costs ξ from two independent distributions, ψ_z and ψ_ξ . Upon learning their type, firms decide to enter if they can cover the entry cost c^e , i.e., whenever

$$\Pi(z, \xi) \geq c^e. \quad (8)$$

The discounted sum of per-period profits is given by

$$\Pi(z, \xi) = \sum_{t=0}^{\infty} \left(\frac{1 - \delta_f}{1 + r} \right)^t \pi(z, \xi) = \frac{1 + r}{r + \delta_f} \pi(z, \xi), \quad (9)$$

where $\pi(z, \xi)$ is defined in equation (7). In an equilibrium with a positive measure of firms, there exists pairs of productivity and training costs (z^*, ξ^*) such that $\Pi(z^*, \xi^*) = c^e$. This defines a region in the space of (z, ξ) for firms that decide to enter.

3.7 The Surplus Function, Hiring, and Training

Because of search and matching frictions, each match has a potential surplus for workers and firms. The surplus, $S(z, \xi, a)$, can be written as

$$S(z, \xi, a) = \max\{0, S^h(z, \xi, a)\}, \quad (10)$$

with

$$\begin{aligned} S^h(z, \xi, a) = & r(z, a) + \frac{(1 - \delta_w)}{1 + r} (1 - (1 - \delta_f)(1 - \delta_s)) J^{u,h}(a) - J^{u,h}(a) \\ & + \frac{(1 - \delta_w)}{1 + r} (1 - \delta_f)(1 - \delta_s) [-\mathbf{1}^t(z, \xi, a)\xi + (1 - p^h(z, \xi, a))M(z, \xi, a) + p^h(z, \xi, a)M(z, \xi, a + 1)]. \end{aligned}$$

The function $M(z, \xi, a)$ denotes the joint match value at the beginning of the period, which is equal to the sum of the value of employment, $J^e(z, \xi, a)$, and the match value for the firm, $V(z, \xi, a)$, i.e.,

$$M(z, \xi, a) = J^e(z, \xi, a) + V(z, \xi, a) = S(z, \xi, a) + J^u(a). \quad (11)$$

We report the full derivation of these functions in Appendix B.1.

A match between a worker with skill $a \in \mathcal{A}$ and a firm with productivity $z \in \mathcal{Z}$ and training cost $\xi \in \mathcal{E}$ is formed upon contact (or kept alive if it already exists) as long as the match surplus is positive, i.e.

$$\mathbf{1}^h(z, \xi, a) = \begin{cases} 1 & \text{if } S^h(z, \xi, a) > 0 \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

where $S^h(z, \xi, a)$ is defined above. Hence, the amount of new hires for a firm- (z, ξ) is given by $v(z, \xi)\phi_f \sum_{a \in \mathcal{A}} \mathbf{1}^h(z, \xi, a)\psi_a^u(a)$.

Finally, each worker-firm pair decides to invest in training to maximize the joint value of the match, i.e.,

$$\mathbf{1}^t(z, \xi, a) = \arg \max_{\mathbf{1}^t \in \{0,1\}} \mathbf{1}^t p^t [M(z, \xi, a + 1) - M(z, \xi, a)] - \mathbf{1}^t \xi, \quad (13)$$

where $M(z, \xi, a)$ is defined in equation (11), which implies that

$$\mathbf{1}^t(z, \xi, a) = \begin{cases} 1 & \text{if } p^t [M(z, \xi, a + 1) - M(z, \xi, a)] > \xi \\ 0 & \text{otherwise.} \end{cases}$$

We assume firms cover the cost of training. This choice reflects the observation that firms often pay for general training, although it might not be optimal in this environment. Human capital accumulation increases the workers' future value of being employed elsewhere, and firms do not internalize this benefit. Still, in an environment with search frictions like ours, firms have incentives to provide general training - see [Acemoglu and Pischke \(1999\)](#) for a review of the literature and [Fu \(2011\)](#) for a recent model with search frictions and firm-sponsored general training.⁶

3.8 Wage Bargaining

Bargaining occurs not only at new matches, but also at continuing matches, on a period-by-period basis. Employers and employees solve the following problem,

$$\max_{w(z,\xi,a)} [J^{e,h}(z, \xi, a) - J^{u,h}(a)]^\beta V^h(z, \xi, a)^{1-\beta}, \quad (14)$$

where $\beta \in (0, 1)$ is the workers' bargaining power. This implies wages $w(z, \xi, a)$ are chosen such the worker's surplus equals a β share of the match surplus, i.e.,

$$J^{e,h}(z, \xi, a) - J^{u,h}(a) = \beta S^h(z, \xi, a).$$

A definition of a recursive competitive equilibrium for this economy and the numerical algorithm implemented to find a solution are described in Appendices B.2 and B.3, respectively.

4 Bringing the Model to the Data

We estimate the model parameters by matching a set of facts on firms and workers from the UK for the 2011-2018 period. The choice of the UK reflects two considerations: First, it is a high-income economy that we contrast with poorer economies in the counterfactuals. Second, the availability of data on firm- and worker-level job training allow us to identify parameters governing human capital accumulation due to experience and training.

We take the UK as a distortion-free economy and set ζ to zero. Moreover we normalize the aggregate productivity shifter κ and the efficiency of the matching function χ to one.

⁶Incentives to invest in general human capital could increase if the matched pair could credibly commit to long term contracts, preventing workers to leave the job at any moment and search for better employers. However, this would reduce the efficiency gains obtained from workers moving to firms with higher productivity. See [Flinn et al. \(2017\)](#) for a discussion.

These three values should be interpreted as normalization against which the counterfactual economies will be compared.

A few parameters are set directly to their data counterparts. The interest rate, r , is 0.0033 to match an annual return of 4%. On average, workers stay in the labor force for forty years, corresponding to ages 22 to 62, so δ_w is 0.0099. The firm destruction rate δ_f is chosen to match an annual firm exit rate of 10.5%.⁷ Finally, the parameter governing the elasticity of the matching function, η , is estimated with the generalized method of moments (GMM), by minimizing the distance between new matches formed according to the model’s matching function (given data on vacancies and unemployed workers) and the number of new hires in the data. The estimated value for η is 0.5416, with a standard error 0.0134. Details of data, estimation, and sensitivity analysis following [Andrews et al. \(2017\)](#) and [Jørgensen \(2020\)](#) are reported in Appendix C.

4.1 Estimated Parameters

The initial human capital of workers and productivity of firms are draws from log-normal distributions, $a \sim \log\mathcal{N}(0, \sigma_a)$, with $\sigma_a > 0$, and $z \sim \log\mathcal{N}(0, \sigma_z)$, with $\sigma_z > 0$. The training costs come from a uniform distribution, given by $\xi \sim \mathcal{U}(\underline{\xi}, \bar{\xi})$, with $0 < \underline{\xi} < \bar{\xi}$. Given these parametric assumption, there are 13 parameters to be estimated, collected in the following vector

$$\vartheta = \{c_e, \sigma_z, \underline{\xi}, \bar{\xi}, \lambda_1, M_e, \beta, \sigma_a, p^d, p^e, p^t, b, \delta_s\}.$$

These parameters are estimated using the method of simulated moments, with standard errors calculated following [Chernozhukov and Hong \(2003\)](#). Table 2 reports estimates and their standard errors.

The estimated values imply significant heterogeneity in training costs across firms; the maximum ($\bar{\xi}$) is about 15 times the minimum ($\underline{\xi}$). The average cost of training one worker equals 21% of the output produced by a worker-firm pair and 8% of the output produced by a pair undertaking training. [Flinn et al. \(2017\)](#) find workers spent 11% of their working time undertaking on-the-job training. Within their model, this corresponds to a comparable value for training cost, 11% of forgone worker-firm output. Large dispersion in training costs across firms is consistent with recent evidence provided by [Arellano-Bover and Saltiel \(2021\)](#).

Consistent with available estimates (see, among others, [Merz and Yashiv \(2007\)](#)), the hiring costs are highly convex, with $\lambda_1 = 2.525$. We estimate firm-productivity dispersion,

⁷Data on firm exit rate come from Business Demographic Statistics of the Office for National Statistics (ONS) for 2011-2018.

Table 2: Estimated parameters

Parameters	Description	Estimate	St.Error
c_e	Entry cost	39.26	3.665
σ_z	Firm-productivity dispersion	1.204	0.106
$\underline{\xi}$	Training cost (lower bound)	1.735	0.157
$\bar{\xi}$	Training cost (upper bound)	26.69	2.304
λ_1	Hiring costs, convexity	2.525	0.166
M_e	Measure of potential entrants	0.013	0.044
β	Workers' bargaining power	0.457	0.042
σ_a	Initial human capital dispersion	1.195	0.111
p^e	Experience jump	0.223	0.019
p^t	Training jump	0.028	0.003
p^d	Depreciation jump	0.432	0.040
b	Home production	20.94	1.824
δ_s	Match separation, %	1.235	0.120

Notes: The entries show the parameters estimated by the method of simulated moments. The standard errors are computed following [Chernozhukov and Hong \(2003\)](#).

σ_z , to 1.20, which implies a coefficient of variation for firm-level labor productivity of 1.38, consistent with the estimates reported by the ONS ([Black et al., 2019](#)).

We estimate workers' bargaining power to be 0.46, a value consistent with the estimates in [Flinn \(2006\)](#), which has a similar wage-setting mechanism for non-employed workers. Estimated values of p^e and p^d imply that for each period of employment, there is about 22% chance that workers' skills can jump by one level. In contrast, for each period of non-employment, they decline by one level with 43% probability. Using a similar process for human capital, [Jarosch \(2021\)](#) estimates a monthly probability of skill depreciation equal to 0.24 for the U.S., corresponding to a quarterly value of 55%. The training jumps skills by one level with a small probability, about 3%. Finally, the value of non-employment, b is about 22.5% of average earnings in the economy, while the entry costs, c_e is 19.9% of the per-capita income in the economy.

4.2 Model Fit

The estimation uses 40 moments, reported in Table 3 and Figures 4 and 5. Overall, the model does remarkably well in fitting the data with an average log-deviation of 0.086. The first column of Table 3 pertains to firm-level targets: i) average firm size, ii) mean and

standard deviation of log employment, iii) fraction of firms that offer training by firm size, and iv) fraction of employees receiving training. The average firm in the data has about 16.4 employees, and the standard deviation of log employment is about 1.2. About 65% of firms offer training. The share increases sharply by firms size, 85% of firms with more than 250 employees offer training to their workers. Across all establishments, about 46% of employees receive training. The model matches all these targets and also generates a firm size distribution that is in line with the data (Figure 4).

Table 3: Selected Targeted Moments

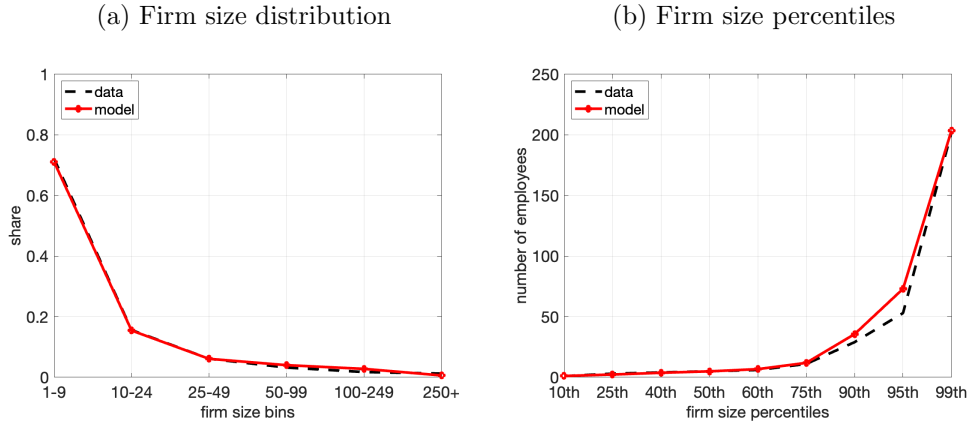
	Data	Model		Data	Model
<i>Firm-level employment</i>			<i>Worker earnings distribution</i>		
Average firm size, $E(\ell_t)$	16.42	16.19	Average earnings at entry, $E[\log(w_1/\bar{w})]$	-0.518	-0.505
Average log-firm size, $E(\log \ell_t)$	1.739	1.700	Average earnings after 20 y.o., $E[\log(w_{20}/\bar{w})]$	0.107	0.109
Dispersion log-firm size, $\text{std}(\log \ell_t)$	1.220	1.392	Average earnings at re-emp, $E[\log(w_R/\bar{w})]$	-0.301	-0.170
			Earnings dispersion at entry, $\text{sd}[\log w_1]$	0.582	0.675
<i>Firm training provision</i>			Earnings dispersion after 20 y.o., $\text{sd}[\log w_{20}]$	0.796	0.795
$E\left(\frac{\#\text{training firms}}{\#\text{firms}}\right)$			Earnings dispersion at re-emp, $\text{sd}[\log w_R]$	0.834	0.833
All firms	0.646	0.650			
Firms with 1-49 employees	0.611	0.644	<i>Worker-level training return</i>		
Firms with 20-249 employees	0.776	0.714	$\log w_{it} = \beta_1 \mathbf{1}_{it-1}^t + \epsilon_{it}$	0.199	0.208
Firms with 250+ employees	0.855	0.888			
			<i>Aggregate moments</i>		
$E\left(\frac{\#\text{trained employees}}{\#\text{employees}}\right)$			Job duration (years)	6.700	6.185
All firms	0.436	0.482	Employment rate	0.776	0.788

Notes: This table reports a set of firm-level and worker-level empirical moments used in the estimation, together with their model counterparts.

While the estimation does not provide with a one-to-one map between parameters and targets, specific targets have more impact on specific parameter in ϑ . In particular, the entry cost c_e determines the average firm size, while firm-productivity dispersion, σ_z , maps into dispersion in firm size. The convexity of the hiring costs, λ_1 , is identified by the different percentiles of firm size distribution. Finally, the boundaries in the support of training costs, ξ and $\bar{\xi}$, are identified by the share of firms providing training and number of workers trained within the firm for different firm size.

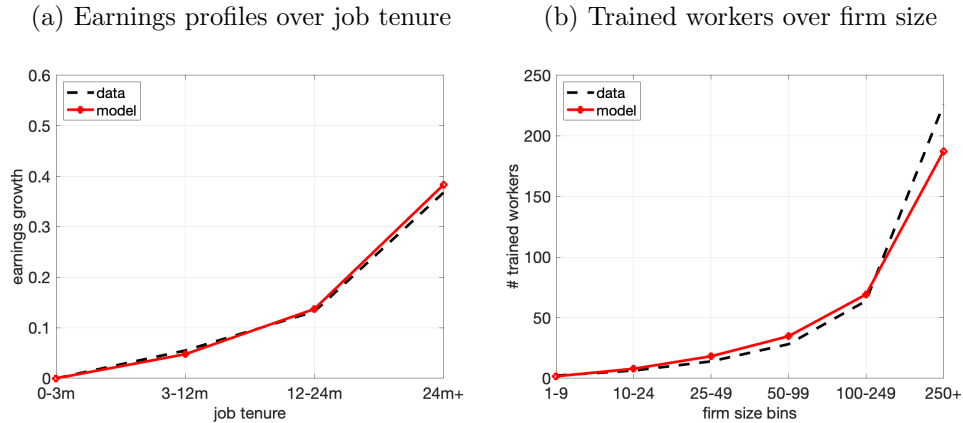
The second column of Table 3 presents a set of worker-level moments: i) earnings level and dispersion, conditional on labor market experience, and at re-employment after an unemployment spell, ii) returns to training, iii) returns to tenure, and iv) average job duration and the employment rate (the share of wage and salary earners in the population). The model does a great job matching the worker-level moments. Workers enter the labor market at an average earnings that is about 50% below the mean, and after 20 years in the labor

Figure 4: Firm size distribution



Notes: Left and right panels show selected empirical moments used in the estimation together with their model counterparts.

Figure 5: Earnings profile and training provision



Notes: Left and right panels show selected empirical moments used in the estimation together with their model counterparts.

market, their earnings grow by about 10%. After a non-employment spell, re-employed workers' earnings are lower than the mean, both in the data and in the model (although the model underestimates the decline). The dispersion of earnings is relatively small when workers enter the labor market, but as their labor market histories diverge, it increases by 20 log points higher after 20 years in the labor market. The returns to past training, calculated with a simple Mincerian regression in the data and the model, are large, about 20%. So are the returns to job tenure; workers with more than two years of job tenure earn almost 40% more than the entrants (panel (a) in Figure 5).

Table 4: Non-targeted Moments

	Data	Model		Data	Model
<i>Earnings-size regression</i>			<i>Earnings inequality</i>		
<10 employees	0	0	Log-earnings dispersion, $\text{sd}[\log w_{it}]$	0.779	0.852
$\in [10, 25)$ employees	0.151	0.183	Mean-median earnings ratio, $E[w_{it}]/p^{50}[w_{it}]$	1.276	1.207
$\in [25, 50)$ employees	0.244	0.342	90-50 pct. earnings ratio, $p^{90}[w_{it}]/p^{50}[w_{it}]$	2.410	2.551
$\in [50, 250)$ employees	0.407	0.680	50-10 pct. earnings ratio, $p^{50}[w_{it}]/p^{10}[w_{it}]$	2.938	5.262
≥ 250 employees	0.586	1.039			

Notes: The entries show a set of empirical moments not included in the estimation, together with their model counterparts.

As far as the identification of different parameters is concerned, the exogenous separation rate, δ_s , determines the average job duration, about 6.7 years. Moreover, the measure of potential entrants M_e maps, given all other parameters, into a value wage and salary employment of about 78% of the population, through its effect on the aggregate vacancies posted and job finding probability, ϕ_w . The parameters governing how skills change during employment and non-employment, p^e and p^d , are disciplined by the earnings profile of workers, while the probability of skill accumulation due to training, p^t is identified by the earnings premium of training workers.⁸ Finally, the distribution of the initial human capital, σ_a , and bargaining power for workers, β are identified, given all other parameters, by the dispersion of earnings at entry and along workers' life cycle.

4.2.1 Non-targeted Moments

Table 4 reports two sets of non-targeted moments: the relation between firm size and earnings and different moments of earnings distribution. Standard search and matching models with large firms and concave production function fail to generate a positive and large earnings-size premium - see [Elsby and Michaels \(2013\)](#) for a discussion. In contrast, the linearity in production function allows the estimated model to deliver a earnings-size premium close to the one observed in the data.

The model also replicates well the observed earnings inequality in the UK. Two-sided heterogeneity, workers and firms, and human capital accumulation allows the model to match

⁸The estimates for the parameters governing human capital accumulation, i.e. p^e, p^t and p^d , together with match separation rate δ_s , are particularly sensitive to calibrated value of η . This is the case because η determines the elasticity of matches to vacancies, hence the rate at which jobs are formed. Changes in match formation are key determinant of job duration and wage dynamics over the life cycle. See Table 13 in Appendix C.2.5.

dispersion in log earnings and the mean to median earnings, even though we abstract from the on-the-job search - see [Hornstein et al. \(2011\)](#) for a discussion.⁹ On the other hand, while the model captures the magnitude of the dispersion in the upper tail of the earnings distribution very well, it generates a more left-skewed earnings distribution.

5 Cross-Country Facts, Revisited

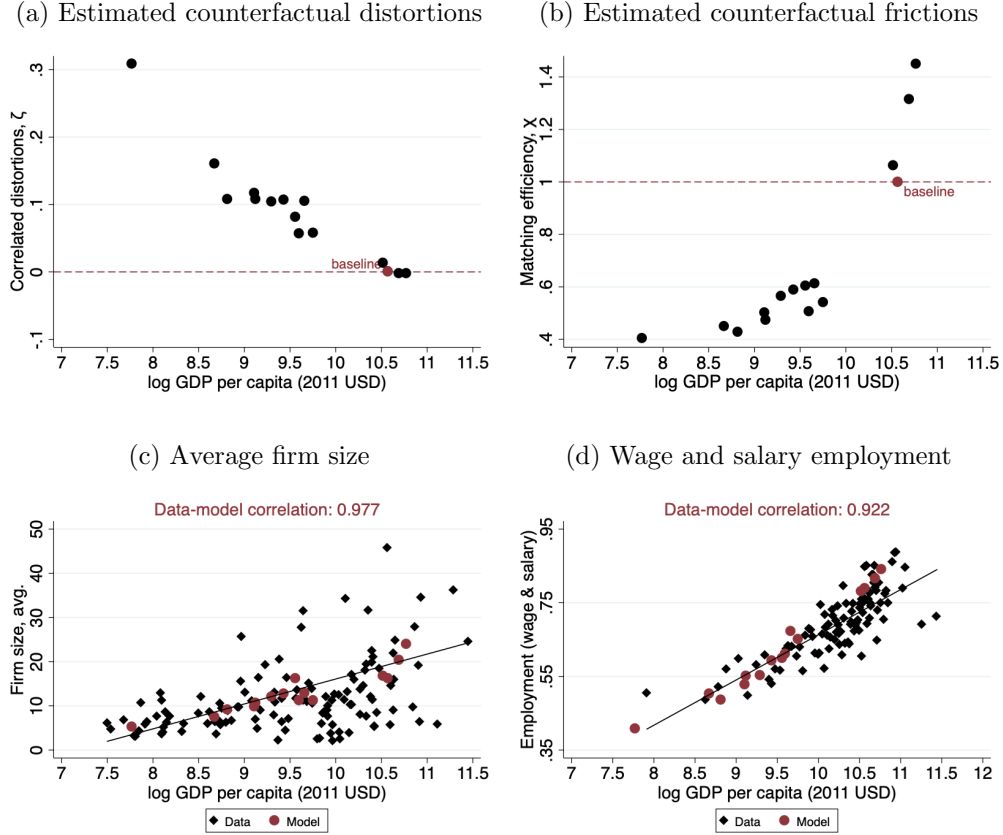
We are now ready to interpret the cross-country facts documented in Section 2 through the lens of the model. To this end, we construct counterfactual economies that differ from the benchmark along two key features: size-dependent distortions, captured by the parameter ζ , and matching frictions captured by the parameter, χ . In Appendix D, we discuss how the data allows us to identify ζ and χ separately.

In the benchmark economy, ζ was set to zero, i.e., there were no size-dependent distortions, while the efficiency of the matching function χ was normalized to one. Keeping all other parameters fixed at their benchmark values, we search for values of ζ and χ that generate the average firm size and wage and salary employment observed in other countries. As a result, the counterfactual economies are replicas of the UK, except for differences in two key parameters that we focus on. We consider eight countries: Brazil, Georgia, Indonesia, Mexico, Peru, Poland, Serbia, and South Africa. We complement these eight countries with six representative economies to span the range of GDP per capita levels observed in the data. The representative economies have the average firm size and wage employment rate of countries with log GDP per capita in following brackets: [8,9), [9,9.5), [9.5,10), [10,10.5), [10.5,11), [11,12). For each counterfactual economy, we adjust the value of home production b , so that it is about 22.5% of the average earnings, the value estimated in the baseline economy.

Figure 6 shows ζ and χ values for each counterfactual economy (panels a and b), together with the targeted moments, average firm size and employment (panels c and d). The range of calibrated values is quite wide, which is necessary to match the data. While ζ is zero for the UK, it is as high as 0.3 for the poorest counterfactual economy (Indonesia). As panel (a) in Figure 6 shows, ζ increases quickly for poorer countries. In particular, they increase substantially for countries with a log GDP per capita lower than 8.5 (about 3000 USD). Similarly, χ is as low as 0.4 in poorest countries and increases sharply for countries that have larger wage and salary employment than the UK.

⁹In absence of on-the-job search, on-the-job human capital accumulation operates similarly, as it provides a way for workers' earnings to grow on the job as in [Burdett and Mortensen \(1998\)](#) and [Postel-Vinay and Robin \(2002\)](#).

Figure 6: Distortions and matching frictions across countries



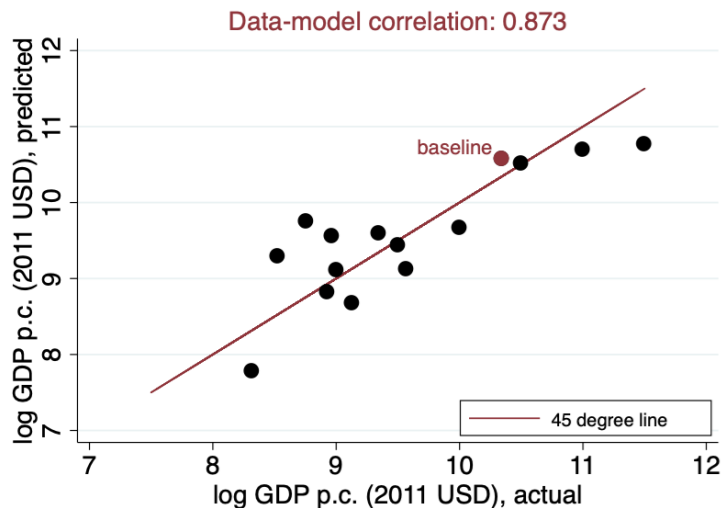
Notes: Panels (a) and (b) report the estimated counterfactual parameters, correlated distortions ζ , and matching efficiency, χ , for every targeted country. Panels (c) and (d) report cross-country moments targeted in the estimation of the counterfactual parameters, average firm size and wage and salary employment, respectively.

We do not directly target the GDP per capita in counterfactual economies. Yet, as illustrated in Figure 7, the model endogenously generates the levels of GDP per capita that are almost perfectly aligned with the data. This is achieved without exogenous productivity differences in the aggregate productivity, which is normalized to one for all countries.

5.1 Workers and Firms around the World

We now have several economies that differ in average firm size, employment rate, and GDP per capita. Each of these economies also provides us with measures of earnings inequality that we can compare with the data. We first show, however, that the counterfactual economies, which only differ in ζ and χ , display several cross-country patterns that are remarkably consistent with available cross-country evidence, which is a strong validation check. In

Figure 7: GDP p.c. across countries: Model vs. Data



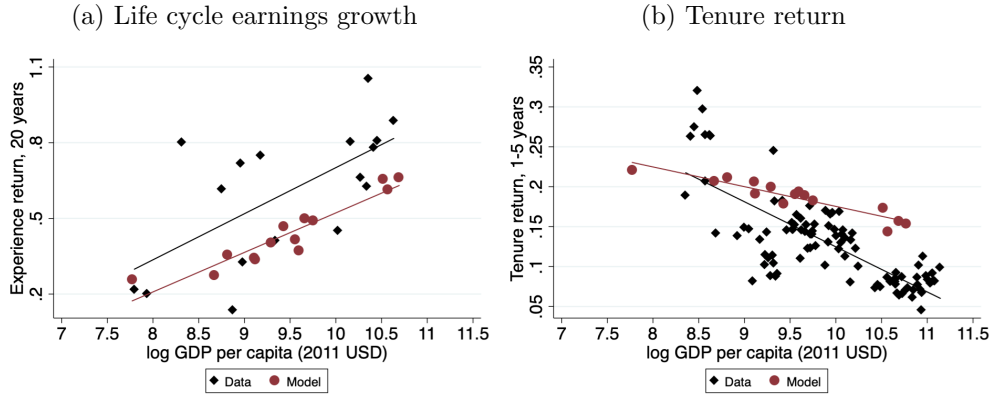
Notes: Each dot compares the observed GDP per capita for a targeted country against the value predicted by the model.

particular, we focus on life-cycle earnings profiles, moments of the firm size distribution, and training provisions.

We start by looking at how earnings change along the life cycle. Panel (a) in Figure 8 shows earnings growth. The data, from [Lagakos et al. \(2018\)](#), is the earnings growth between ages 22 (labor market entry) and 42. The model counterpart is the average earnings growth during the first 20 years of working life. In the data age-earnings trajectories are much steeper in richer countries. In the model, frictions and distortions depress wage and salary employment, preventing workers from accumulating human capital and depressing wage growth as much as in the data.

Earnings grow much faster, however, during initial years of employment in poorer countries. This is documented in panel (b) of Figure 8. The data, from [Donovan et al. \(2020\)](#), shows the average earnings for those between 1 and 5 years of tenure relative to those with less than 6 months of tenure. We calculate the same statistics in the model economy and again the model predictions are in line with the data. Earnings increase sharply with tenure, and do so more in poorer countries, where the share of wage and salary earners is significantly lower, and only the high-ability workers are able to find wage and salary jobs. On the other hand, workers in poorer countries spend longer time non-employed, and as a result these gains do not translate into lifetime earnings growth. In contrast, workers are less likely to stay non-employed in richer countries, and despite their returns to tenure are lower, they

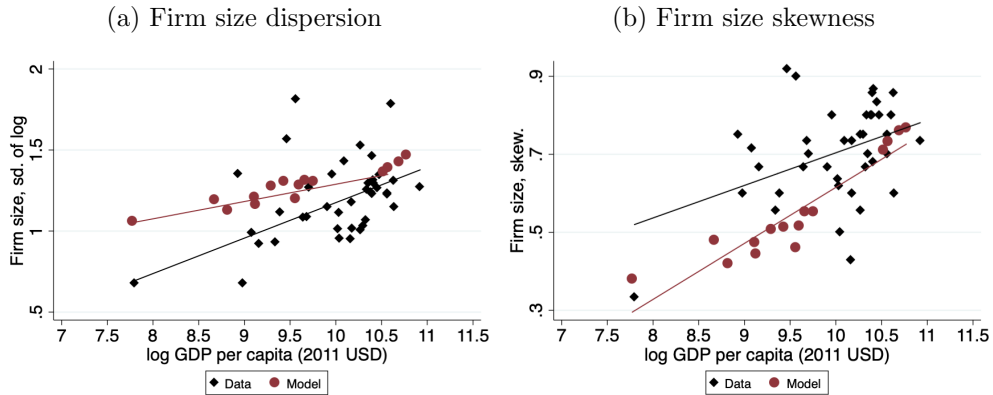
Figure 8: Earnings profile across countries



Notes: Panel (a) shows the earnings growth during the first 20 years of a worker’s life cycle for countries with different log GDP per capita. Panel (b) shows the earnings growth during the first 5 years of employment. The black diamonds represent the data and red dots the model.

experience higher earnings growth along the life cycle.

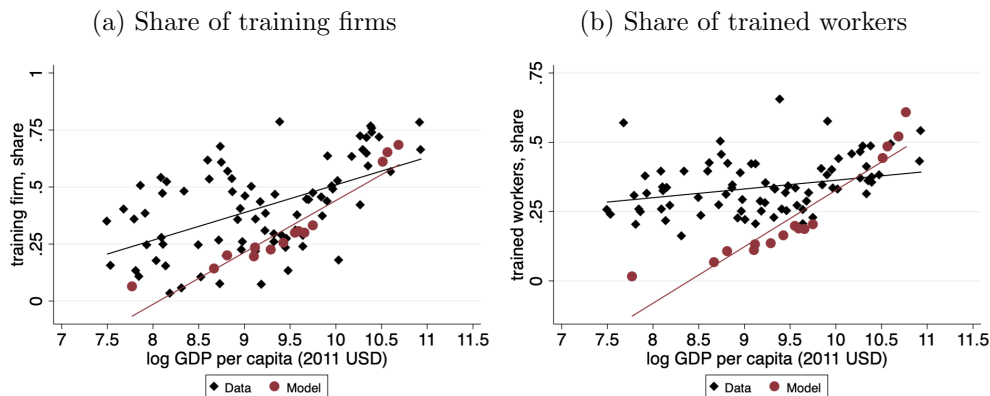
Figure 9: Firm distribution across countries



Notes: Panels (a) and (b) show how the dispersion (measured as the standard deviation of log size) and the skewness of firm-size distribution change with log GDP per capita. The black diamonds represent the data and red dots the model.

Figure 9 shows that dispersion and skewness of the firm size distribution are significantly higher in richer countries. The data, from [Poschke \(2018\)](#), shows the standard deviation of log firm size (panel a) and the 90/10 percentile skewness (panel b). The model is again in line with the data. In the model, distortions and frictions prevent firms from growing, squeezing the size distribution towards the left of the support, cutting the distribution’s right tail, and reducing size dispersion. Moreover, distortions and frictions reduce the gains from training,

Figure 10: Training provision across countries



Notes: Panels (a) and (b) show how the share of firms that provide on-the-job training and the share of workers who received on-the-job training change with log GDP per capita. The black diamonds represent the data and red dots the model.

which leads to a lower number of firms that offer training and to a lower share of trained workers within each firm, as documented in Section 2. Figure 10 show that the model can also account for these cross-country patterns.

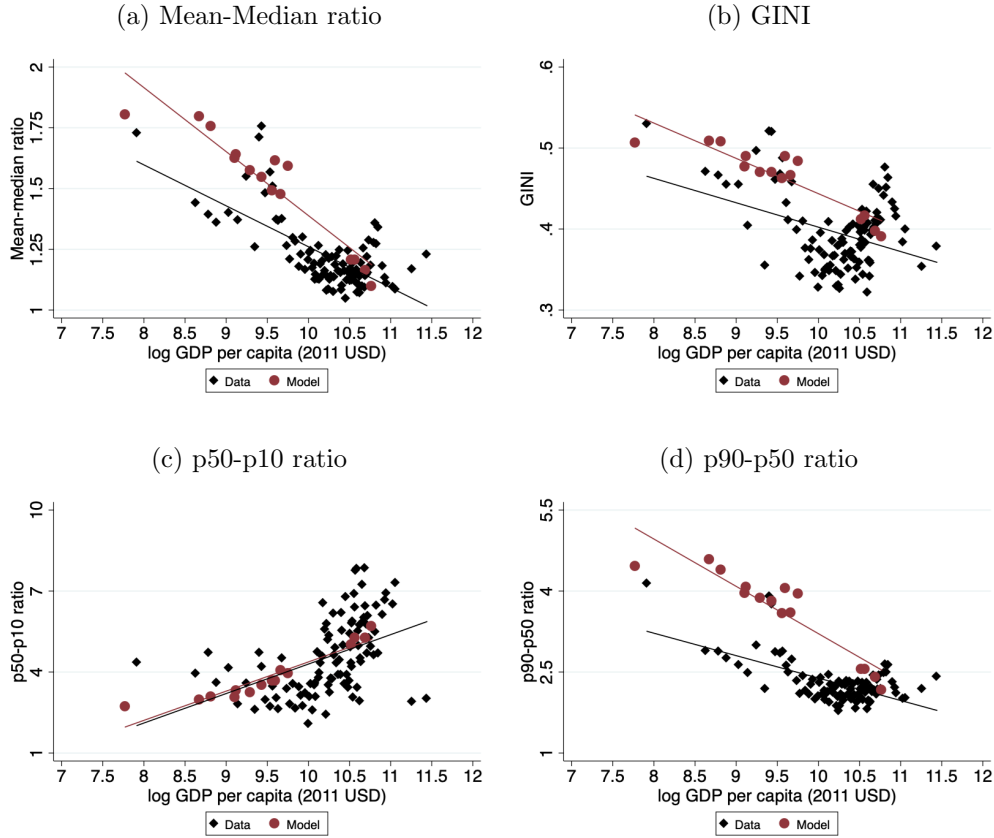
5.2 Earnings Inequality

We are now ready to look at how earnings inequality changes with economic development. As Figure 11 shows, the model does a great job replicating how earnings inequality changes with higher GDP per capita. As in the data, mean-to-median ratio and the Gini index declines as countries get richer. But p50-p10 ratio increases sharply and p90-p50 ratio declines. The workers at the bottom of the earnings distribution are not able to catch up with the median workers, but the median workers are getting closer to those at the top. Hence, just by re-calibrating two parameters (χ and ζ), model is able to generate changes in earnings inequality that we observe in the data.

5.3 Discussion

Our quantitative strategy is to understand how far we can go by focusing on cross-country differences in distortions and matching efficiency. The results in the previous section show that this strategy is very successful. As discussed in the next section, it is also transparent and allows us to highlight how we generate cross-country patterns in earnings inequality. An alternative strategy would be to recalibrate several parameters for different countries, not

Figure 11: Earnings inequality across countries: Model vs. Data



Notes: Each panel shows how a particular measure of earnings inequality changes with log GDP per capita across countries. The black diamonds represent the data and red dots the model.

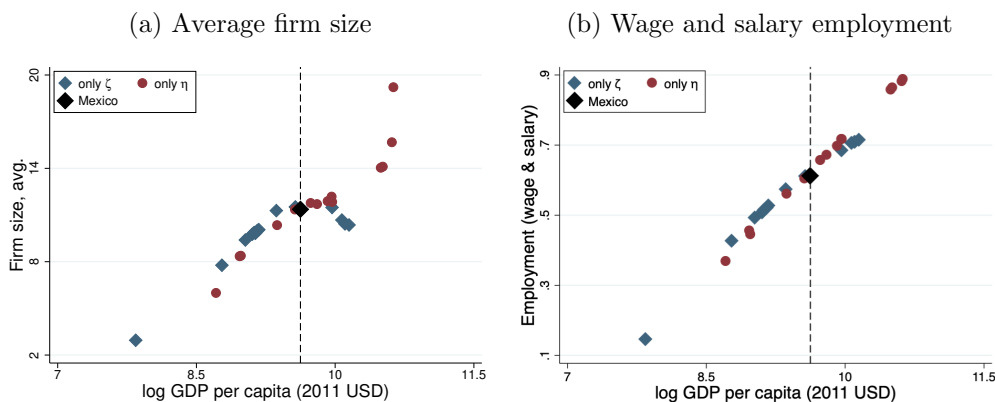
just χ and ζ . However, we do not get much additional mileage from such a strategy and lose transparency. We illustrate this in Appendix E comparing the benchmark economy (UK) with Indonesia, a country with one-tenth of UK's GDP per capita (4,095 USD vs. 39,000 USD). Table 14 shows the outcomes of our baseline counterfactual against the alternative counterfactual where multiple parameters are recalibrated. In particular, together with matching efficiency, χ and correlated distortion, ζ , we re-estimate the aggregate productivity shifter, κ to match the GDP per capita, the experience probability jump, p^e to match the life-cycle wage growth after 20 years, and the training probability jump, p^t to match the share of firms providing training. Moreover, we adjust the values of home production, b , the boundaries for training costs, $\underline{\xi}$ and $\bar{\xi}$, and the entry cost c_e such that they are equal to the value estimated in the baseline economy, as % of average earnings. The changes in earnings inequality in the benchmark counterfactual, with only differences in χ and ζ , are very close to what we obtain from this alternative counterfactual.

Also, the cross-country differences in distortions and frictions can take different forms. In Appendix F, Table 15, we consider differences in worker separation rates, δ_s , and firm exit rates, δ_f . While their effects are qualitatively similar, differences in ζ and χ are quantitatively more effective in generating the observed cross-country patterns in the data.

6 Frictions vs Distortions

How does the model generate these cross-country patterns? We answer this question by first discussing the relative importance of correlated distortions and matching frictions. To this end, we consider a country in the middle of the GDP per capita distribution, Mexico (the GDP per capita of Mexico is 11,400 USD vs. 39,000 USD of the UK). Then, for each counterfactual economy in Section 5, we only impose the calibrated values of distortions, ζ , or frictions, χ , from Figure 6. If we impose the calibrated value of ζ for a country, we keep the value of non-employment, b , and χ at their calibrated values for Mexico. If, instead, the calibrated value of χ is imposed, b and ζ are kept at their values for Mexico. As a result, starting from Mexico, we increase or decrease the GDP per capita due to changes in ζ or χ alone.

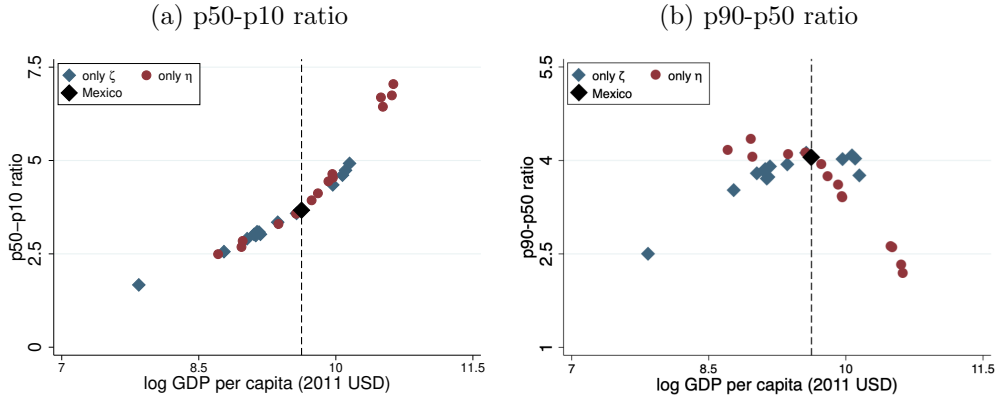
Figure 12: Firm size and employment across countries: distortions vs. frictions



Notes: Panel (a) and (b) shows the average firm size and the share of wage and salary workers across countries when only χ (red dots) or ζ (blue diamonds) are allowed to change, while the other parameter are fixed at Mexico's value.

Figure 12 shows the average firm size (panel a) and wage and salary employment (panel b) when either correlated distortions (blue diamonds) or search frictions (red dots) are allowed to change across countries. Either correlated distortions, ζ , or matching frictions, χ , alone can generate a positive correlation between GDP per capita and both average firm size and

Figure 13: Earnings inequality across countries: distortions vs. frictions



Notes: Panel (a) and (b) shows the p50-10 and p90-p50 earnings ratios when only χ (red dots) or ζ (blue diamonds) are allowed to change, while the other parameter is fixed at Mexico's value.

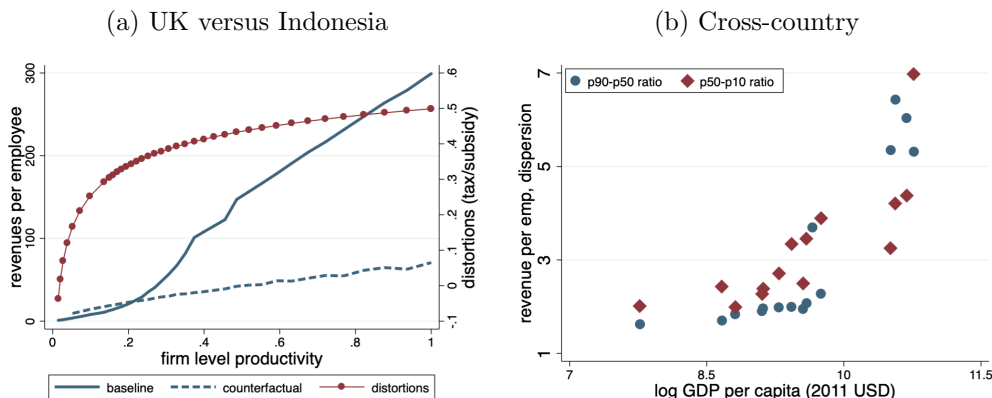
wage and salary employment. Yet, their impact differs markedly depending on whether we look at countries poorer or richer than Mexico.

While search frictions are quantitatively more important in richer countries, correlated distortions are necessary to generate observed patterns in poorer ones. Consider, for example, panel (a) of Figure 12. Reductions in correlated distortion alone can only account for changes in firm size for countries poorer than Mexico (blue diamonds). Indeed, there is very little change in the average firm size as we move to countries with a lower ζ . In contrast, lower search frictions generate larger average firm size in richer countries, but they cannot quantitatively account for changes in firm size observed in very low-income countries. Both search frictions and correlated distortions are necessary to generate variation in firm size and employment across the entire spectrum of GDP per capita. The same argument applies for changes in employment rate, reported in panel (b) of Figure 12.¹⁰

What about earnings inequality? Start with the role of correlated distortions. Changes in ζ alone generate a positive relationship between GDP per capita and inequality at both ends of the earnings distribution. Both the p50-p10 and p90-p50 ratios increase with GDP per capita (blue diamonds in Figure 13). Search frictions instead affect the distribution of earnings asymmetrically. As matching in the labor market becomes more efficient, GDP per capita increases. But while earnings become more dispersed at the bottom and the p50-p10 ratio rises, they become less dispersed at the top, and the p90-p50 ratio drops (red dots in

¹⁰Note that for high levels GDP per capita, the relation between GDP per capita and firm size becomes flat or can even be slightly negative. This is due to general equilibrium effects. A low ζ encourages entry and, because of matching frictions, firm growth can slow down.

Figure 14: Firm-level revenues per employee



Notes: The red dots in Panel (a) shows the correlated distortions in Indonesia (left axis). The solid and dashed-blue lines show revenue per worker in the benchmark (UK) and counterfactual (Indonesia), respectively. Panel (b) shows p90-p50 (blue dots) and p50-p10 (red diamonds) ratios for the revenue-per-worker distribution.

Figure 13).

6.1 Mechanisms

Why is this happening? We shed light on this by focusing on three key determinants of earnings inequality in the model: i) how revenues are distributed across firms with different productivity levels, ii) how long it takes for workers to match with a firm (non-employment duration), and iii) sorting between workers and firms. We do that by comparing the UK with Indonesia. The calibrated values of ζ and χ for Indonesia are 0.312 (versus 0 in the UK) and 0.403 (versus 1 in the UK).

Distribution of firm-level revenues. The first channel operates through changes in the distribution of firm-level revenues. Panel (a) in Figure 14 shows the average distortions, τ , faced by firms of different productivity levels z in Indonesia (red dots). Recall that the implicit tax rate, τ , is zero in the baseline economy. In Indonesia, with ζ equal to 0.312, implicit taxes on firms' output increase sharply with firm productivity. The employment-weighted average tax rate is around 51% in Indonesia and the tax rate increases from 40% for firms with less than 10 workers to 70% for those with more than 25 employees.

Panel (a) of Figure 14 also reports the average revenues per employee in the UK and Indonesia (blue lines). A higher value of ζ implies more progressive output taxes, which reduce the difference between productive and unproductive firms, making them more similar

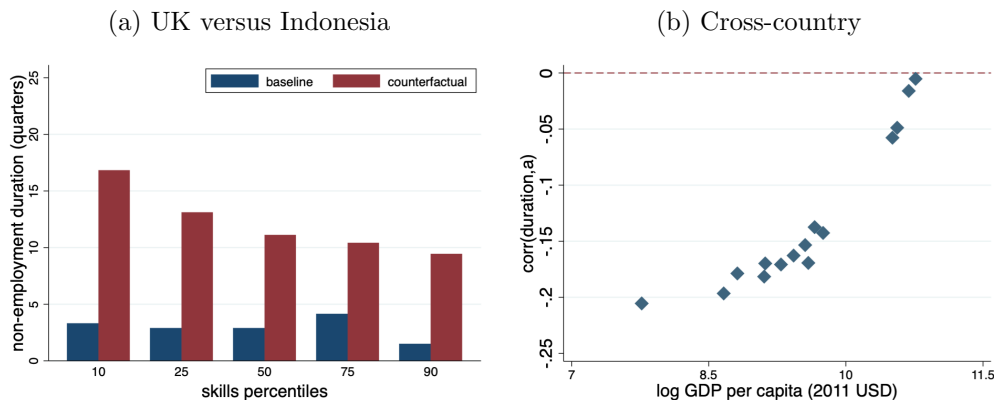
as potential employees for workers in Indonesia. Notice that in the UK, a firm at the top 10% of the productivity distribution has about 7 times higher revenues per employee than the median firm. A firm with median productivity, in turn, has about 5 times higher revenue per worker than a firm at the bottom 10%. As we move to poorer countries, these differences shrink significantly. Panel (b) in Figure 14 compares revenue per worker for firms at different points of the revenue-per-worker distribution. Both p50-p10 (red diamonds) and p90-p10 (blue circles) increase with the GDP per capita. The earnings distribution mirrors the distribution of revenue per employee. As a result, both p50-p10 and p90-p50 earnings ratios also increase as we move from countries with higher to lower distortions (blue diamonds in Figure 13).

Non-employment duration. Let's now turn to the second channel, which operates through changes in non-employment duration across workers. Panel (a) in Figure 15 shows the average non-employment duration for workers of different skills at the time of re-employment in the benchmark (UK) and the counterfactual (Indonesia). In the UK, re-employed workers have spent on average 3 quarters before transiting from non-employment to employment. Moreover, duration is relatively similar across workers with different skill levels. Non-employment spells are much longer in Indonesia, about 11 quarters for the average re-employed worker. Furthermore, the average spells decline significantly by worker's skills. Hence, as a country gets richer, non-employment duration shrinks and becomes more uniform across re-employed workers with different skills. The decline is most significant for workers with low skills: higher labor market efficiency prevents skill depreciation during non-employment and this effect is stronger for low-ability workers because it allows them to stay attached to the labor market.

Panel (b) in Figure 15 shows the correlation between the skills of the re-employed workers, a , and their non-employment duration across countries. The correlation is always negative, i.e., more skilled workers get out of non-employment faster. But the correlation is stronger in the poorest countries, it increases monotonically with GDP per capita, and it is almost zero in countries with the highest GDP per capita. In the model, non-employment duration decreases as frictions in the labor market are removed, and the decline is stronger for low-skilled workers. This effect implies more employment opportunities and relatively higher human capital accumulation for the low-skilled, and, as a result, a compression in the distribution of skills and lower earnings inequality.

Worker-firm sorting. Finally, Figure 16 describes the sorting between firms and workers. In panel (a), the horizontal axis ranks workers again by their skills while the vertical axis

Figure 15: Non-employment duration



Notes: Panel (a) shows non-employment duration in the benchmark (UK) and counterfactual (Indonesia) for workers in different percentiles of the skill distribution. Panel (b) shows the correlation between GDP per capita and non-employment duration across countries in the model.

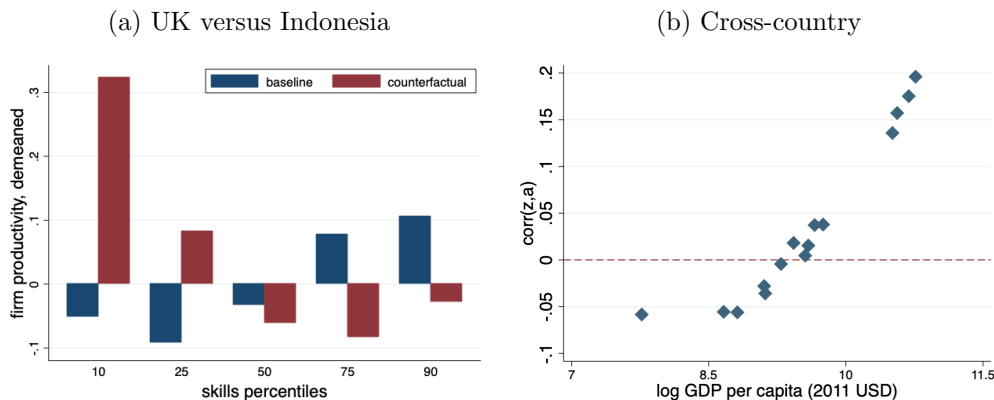
shows the average productivity of firms that employ these workers (as log deviations from the mean firm productivity). Workers at the bottom of the skill distribution in the benchmark economy are employed by firms with about 10% lower productivity than the average firm in the economy (blue bars in panel a). As workers become more productive, so do their employers, generating positive assortative matching. This is not the case in the counterfactual (red bars in panel a). Low-skill workers are matched to high productivity firms, and a high-productivity firm does not necessarily have a more skilled workforce, resulting in negative assortative matching.

Stronger labor market frictions make it costly for high-ability workers to wait for better employers, and they end up accepting any job they can find.¹¹ It takes about 3.5 quarters for a vacancy to encounter a potential hire in the UK. In Indonesia, a firm has to wait on average 5.6 quarters. Panel (b) in Figure 16 shows the correlation between workers' skills, a , and the productivity of their employers, z , for each counterfactual economy. The correlation is negative for the poorest countries, increases monotonically with GDP per capita, and eventually becomes positive as frictions and distortions vanish. This effect on sorting increases earning inequality.

A more fluid labor market increases firm-worker sorting and reduces workers' non-employment spells. The former favors workers who are more skilled, who reap the benefits from matching with high productivity firms. The latter is relatively stronger for less skilled workers, who

¹¹Higher distortions also lower sorting by making firms more similar from workers' point of view. This effect is, however, quantitatively small.

Figure 16: Sorting



Notes: Panel (a) shows the level of firm productivity, measured as deviations from the overall mean, for workers in different percentiles of the skill distribution for the benchmark (UK) and counterfactual (Indonesia). Panel (b) shows the correlation between firm productivity(z) and their workers' skills (a) across countries in the model.

progressively populate employment and accumulate human capital as GDP per capita increases. In the simulations, the net effect is a distribution of skills with a much larger mass in the middle, generating an increase in the p50-p10 earnings ratio and a reduction in the p90-p50 earnings ratio over development (red dots in Figure 13).

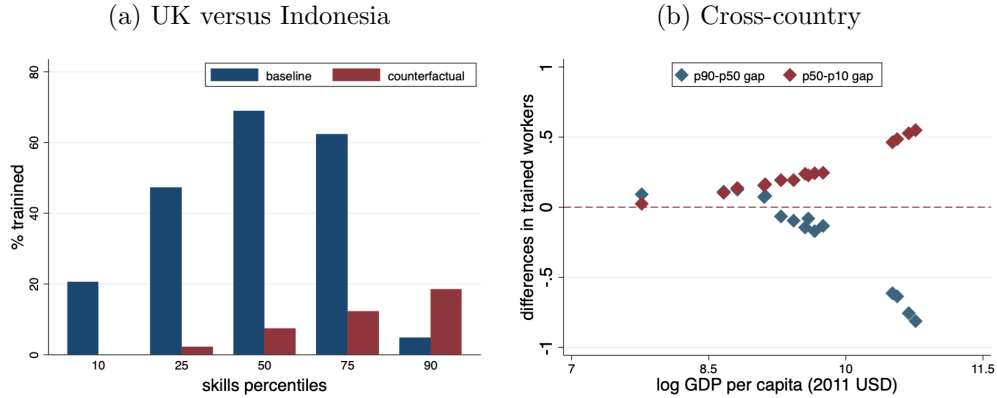
6.2 Role of Training

What is the role of OTJ training in the model? Panel (a) in Figure 17 shows the fraction of workers who receive training in the model economy, conditional on their skills.

In the benchmark (UK), there is an inverted-U relation between worker skills and intensity of training, i.e., workers in the middle of skill distribution are more likely to receive training than those at the bottom or the top. Firms do not have strong incentives to train low-skilled workers since productivity gains can't always cover the cost of training. On the other hand, training improves the matching opportunities of all workers, in particular, the high-skilled ones. As a result, their outside options significantly improve, reducing gains from training these workers.

In a poor country, training provision shrinks dramatically. As a percentage point decline, the training reduction mainly affects workers in the middle of the skill distribution. Furthermore, the intensity of training is now increasing in workers' skills. The training of low-skilled workers vanishes since it is even harder for the firm to cover the cost of training. But with more significant labor market frictions, workers have no incentives to leave the firm. As a

Figure 17: Training



Notes: Panel (a) shows the fraction of workers receiving on-the-job training in the benchmark (UK) and counterfactual (Indonesia) for workers in different percentiles of the skill distribution. Panel (b) shows the differences between the shares receiving training for workers who are at the 90th versus the 50th percentile (blue diamonds) and the 50th versus the 10th percentile (red diamonds) of the skill distribution.

result, firms now have strong incentives to invest heavily in their skilled workers.

Panel (b) in Figure 17 illustrates how the share of workers with different skills who receive training vary with GDP per capita. We compare the share of workers with median skills with those at the bottom (p10) and the top (p90). As countries get richer, the share of trained workers with median skills relative to the lower-skilled ones increases, i.e., median-skilled workers are more likely to get training than low-skilled workers. In contrast, the high-skilled workers who receive training relative to the median-skilled ones decline. These cross-country differences in training are reflected in how the p50-p10 and p90-p50 earnings ratios change with development.

How much does OTJ training account for the cross-country patterns of inequality? To answer this question, we re-calibrate the baseline economy (the UK) without training, i.e., we set $p^t = 0$ so that no one receives training. The re-calibration uses the same targets in Tables 3, except the ones on training. The moments and parameter estimates are reported in Appendix G. Columns 1 and 3 in Table 5 show the baseline economy with and without training. We then move to Indonesia, i.e., impose $\chi = 0.403$ and $\zeta = 0.308$. Columns 2 and 4 in Table 5 reports the counterfactual results with and without training.

A comparison between Columns 1 and 2 versus 3 and 4 shows that changes in inequality are more muted. When we allow for the OTJ training, the mean-median earnings ratio declines from 1.805 in Indonesia to 1.207 in the UK. Without training, the decline is from 1.67 to 1.28. Hence, training accounts for about 35% of the total change. Similarly, the

Table 5: A World without OJT Training

	Baseline with OTJ training (1)	Counterfactual (2)	Baseline w/o OTJ training (3)	Counterfactual (4)	Explained
Matching friction: χ	1	0.403	1	0.403	-
Distortion correlation: ζ	0	0.308	0	0.308	-
Home production: b	20.94	3.505	20.94	3.505	-
<i>Aggregates</i>					
Employment rate	0.788	0.407	0.797	0.561	38.05%
Average earnings	1	0.124	1	0.140	1.84%
Income per capita	1	0.061	1	0.086	2.69%
<i>Earnings profile over experience/tenure</i>					
Earnings growth, $E[\log(w_{25}/\bar{w}_1)]$	0.801	0.280	0.731	0.363	29.45%
<i>Earnings inequality</i>					
Mean-median ratio, $E[w_{it}]/p^{50}[w_{it}]$	1.207	1.805	1.280	1.667	35.13%
GINI	0.416	0.506	0.416	0.487	20.99%
90-50 pct. ratio, $p^{90}[w_{it}]/p^{50}[w_{it}]$	2.551	4.462	2.815	3.990	33.87%
50-10 pct. ratio, $p^{50}[w_{it}]/p^{10}[w_{it}]$	5.262	2.729	4.118	2.854	58.51%

Notes: The entries in columns (1) and (2) show the benchmark (UK) and the counterfactual (Indonesia). The entries in columns (3) and (4) show the benchmark (UK) and the counterfactual (Indonesia) when there is no on-the-job training. The last column shows the ratio of differences between (3) and (4) compared with (1) and (2).

difference in the Gini coefficient is also smaller. The Gini declined by 9 percentage points for the benchmark calibration, while without on-the-job training, the decline is about 7 percentage points. Training also magnifies changes in p90-p50 and, in particular, p50-p10 ratios.

In Appendix H, we also report outcomes from an alternative experiment where we impose training decisions from the counterfactual economy (Indonesia) on the UK firms. Hence, if a match between a type- a worker and type- (z, ξ) implies training (or no training) in the counterfactual economy, the pair behaves the same way in the baseline economy, even if such behavior is not optimal for the match. With this experiment, we find that endogenous training decisions account for about 11% of changes in the mean-to-median ratio and the Gini coefficient.

7 A Re-training Program

We next assess the value of a re-training program for unemployed workers. Until recently, the accepted view on the active labor programs, e.g., vocational training, wage subsidies, or job search assistance, was that they had little impact on employment or earnings (see reviews by McKenzie (2017) and Card et al. (2018)). Alfonsi et al. (2020) show, however, that an intensive intervention in Uganda aimed at providing unemployed young workers with vocational or firm-sponsored training increased both employment and wages significantly.¹² The vocational training arm of their randomized control trial (RCT) provides unemployed young workers, between ages 18 and 25, with a six-months, fully-subsidized, sector-specific training. Motivated by their design, we introduce a fully subsidized re-training program available for all non-employed workers in the model, scaling up their program to the entire economy.

We assume that non-employed workers have the option of either searching for a job or participating in a re-training program and postpone their search. Compared to those who choose to search for jobs, re-trained workers do not face any skill depreciation. Instead, because of re-training, their skills can increase with probability p^t . The other features of the model are kept the same. We report details in Appendix I. Table 6 compares benchmark (the UK, Column 1) and counterfactual outcomes. Column 2 is Indonesia without a re-training program, i.e., the counterfactual discussed in Sections 5 and 6, while column 3 shows the outcomes for Indonesia with the re-training program. We calibrate the cost of re-training following Alfonsi et al. (2020). They report a training cost of 470 USD per participant in a six-month-long training session in Uganda. Using their costs as a fraction of GDP per capita, training a worker for a model period of a quarter should cost 1024 USD in Indonesia. The program is fully subsidized and financed by a lump sum tax on everyone (employed and non-employed).

Not every non-employed workers choose to receive training, even if it is fully subsidized. But the re-training is popular; more than 40% of non-employed workers participate in the program.¹³ Figure 18 reports the probability of choosing to re-train by non-employment duration (panel a) and pre non-employment earnings (panel b). Workers who have just lost their jobs prefer to search since their human capital remains relatively intact. As the unemployment duration increases and workers' human capital depreciates, they are more likely to re-train instead of looking for a job. On the other hand, workers with low pre

¹²Attanasio et al. (2011) evaluate a program that combines vocational and firm-provided training in Colombia and find significant effects on employment and earnings for women, but not for men.

¹³Alfonsi et al. (2020) report that 68% of workers assigned to vocational training starts the program. It is reassuring that our take-up rate is lower since their program targets disadvantaged youth.

Table 6: Re-training program

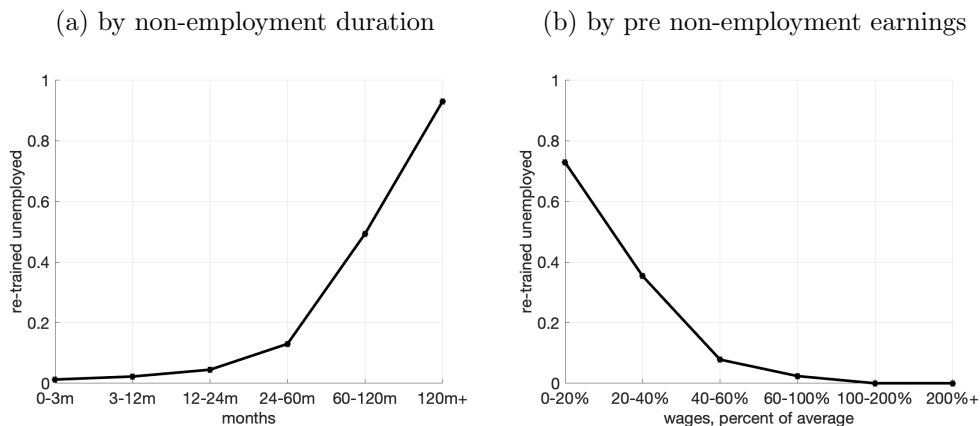
	UK	Indonesia	
	Baseline	Counterfactual	
	(1)	(2)	(3)
Matching friction: χ	1	0.403	0.403
Distortion correlation: ζ	0	0.308	0.308
Home production: b	20.94	3.505	3.505
Re-training under non-employment	no	no	yes
Cost per re-trained individual:	-	-	1024 USD
<i>Re-trained workers</i>			
$E\left(\frac{\# \text{re-trained workers}}{\# \text{non-employed workers}}\right)$, %	0	0	43.07
<i>Aggregates</i>			
Employment rate	0.788	0.407	0.529
Average earnings	1	0.124	0.140
Income per capita	1	0.061	0.095
Income per capita (net of re-training costs)	1	0.061	0.070
<i>Earnings profile over experience</i>			
Earnings growth, $E[\log(w_{25}/\bar{w}_1)]$	0.801	0.280	0.329
<i>Earnings inequality</i>			
Mean-median ratio, $E[w_{it}]/p^{50}[w_{it}]$	1.207	1.805	1.787
GINI	0.416	0.506	0.500

Notes: The entries in columns (1) and (2) show the benchmark (UK) and the counterfactual (Indonesia). The entries in column (3) show the counterfactual economy with a re-training program.

non-employment earnings, hence low human capital, are more likely to opt for a re-training program, while those with high pre non-employment earnings are more likely to look for jobs. These patterns endogenously replicate the sample selection implemented in the RCT of [Alfonsi et al. \(2020\)](#). Compared to labor-market active workers, the targeted sample in their RCT is worse off in terms of labor market outcomes at baseline. The selected workers were less likely to have any wage employment in the week prior and had on average lower total earnings from wage employment in the previous month (see Table A.II in their Appendix). This suggests that the incentives provided by this program within our model are able to scale up the sample restriction in the RCT to the entire economy.

How valuable is the re-training program? It increases employment opportunities significantly since there are substantial gains in human capital accumulation for the participants;

Figure 18: Selection into re-training



Notes: Panel (a) and (b) show the fraction of non-employed workers who choose to participate in the re-training program by non-employment duration and pre-non-employment earnings.

recall that p^t , the probability of a one-step jump in a with training is about 2.8% per month while p^d , the likelihood of a one-step decline in a is more than 40%. As a result, re-training opportunities increase employment from 41% to 53% (excluding re-trained workers). With better employment opportunities, average earnings and income per capita increase significantly. After considering the program's cost, income per capita is around 15% higher (0.061 vs. 0.070). Not surprisingly, the re-training program also implies a 5 percentage point steeper age-earnings profile for workers.

8 Conclusion

A growing literature in macroeconomics has been emphasizing how the misallocation of resources at the micro-level can generate aggregate income and productivity differences. This literature has been built around the idea of distortions, either modeled as explicit policies or implicit taxes on firms' production decisions. If correlated with firms' productivity, distortions result in smaller firms on average and lower incomes. However, this literature has been silent on how misallocation might affect earnings distribution since distortions often are embedded within competitive labor markets. Yet, there is growing evidence that firm-level drivers are fundamental to understanding earnings inequality.

Search and matching models provide a natural framework to study firm-level drivers of earnings inequality. In these models, labor market frictions determine how workers are matched with firms and affect firms' and workers' incentives to invest in their skills. Yet, search and matching models often focus on one-worker with one-firm abstraction and do not

necessarily speak to cross-country differences in firm dynamics.

We combine these two approaches to study how misallocation affects earnings inequality. The benchmark economy speaks to a large set of facts on firms (size distribution, size-earnings, and size-training decisions) and workers (age-earning, tenure-earnings profiles, and training provision). The model also delivers a natural framework to study how the distribution of earnings changes with economic development. In the data, the dispersion of earnings at the bottom increases with development, it declines at the top. We show that the model replicates this pattern when poorer countries are characterized by higher correlated distortions and labor market frictions.

References

- Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.
- Acemoglu, D. and Pischke, J.-S. (1999). Beyond becker: Training in imperfect labor markets. *Economic Journal*, 109(453):F112–F142.
- Akcigit, U., Alp, H., and Peters, M. (2021). Lack of selection and limits to delegation: Firm dynamics in developing countries. *American Economic Review*, 111(1):231–275.
- Alfonsi, L., Bandiera, O., Bassi, V., Burgess, R., Rasul, I., Sulaiman, M., and Vitali, A. (2020). Tackling youth unemployment: Evidence from a labor market experiment in uganda. *Econometrica*, 88(6):2369–2414.
- Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.
- Arellano-Bover, J. and Saltiel, F. (2021). Differences in on-the-job learning across firms. *working paper*.
- Attanasio, O., Kugler, A., and Meghir, C. (2011). Subsidizing vocational training for disadvantaged youth in colombia: Evidence from a randomized trial. *American Economic Journal: Applied Economics*, 3(3):188–220.
- Bento, P. and Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.

- Black, R., Evans, C., Hill, L., Masson, J. L., Radcliffe-Brown, B., and Shafat, M. (2019). Firm-level labour productivity measures from the annual business survey, great britain: 2017. Technical report, Office for National Statistics.
- Breza, E., Kaur, S., and Shamdasani, Y. (2021). Labor rationing. *American Economic Review*, 111(10):3184–3224.
- Buera, F. J., Kaboski, J. P., and Shin, Y. (2011). Finance and development: A tale of two sectors. *American economic review*, 101(5):1964–2002.
- Burdett, K. and Mortensen, D. T. (1998). Wage differentials, employer size, and unemployment. *International Economic Review*, pages 257–273.
- 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.
- Card, D., Kluve, J., and Weber, A. (2018). What works? a meta analysis of recent active labor market program evaluations. *Journal of the European Economic Association*, 16(3):894–931.
- Carranza, R., Morgan, M., and Nolan, B. (2022). Top income adjustments and inequality: an investigation of the eu-silc. *Review of Income and Wealth*, 0(0).
- Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.
- David, J. M., Hopenhayn, H. A., and Venkateswaran, V. (2016). Information, misallocation, and aggregate productivity. *Quarterly Journal of Economics*, 131(2):943–1005.
- David, J. M. and Venkateswaran, V. (2019). The sources of capital misallocation. *American Economic Review*, 109(7):2531–67.
- Donovan, K., Lu, W. J., and Schoellman, T. (2020). Labor market dynamics and development. *Yale University Economic Growth Center Discussion Paper*, (1071).
- Elsby, M. W. L. and Michaels, R. (2013). Marginal jobs, heterogeneous firms, and unemployment flows. *American Economic Journal: Macroeconomics*, 5(1):1–48.
- Engbom, N. (2020). Labor market fluidity and human capital accumulation. *working paper*.
- Flinn, C., Gemici, A., and Laufer, S. (2017). Search, matching and training. *Review of Economic Dynamics*, 25:260–297.

- Flinn, C. J. (2006). Minimum wage effects on labor market outcomes under search, matching, and endogenous contact rates. *Econometrica*, 74(4):1013–1062.
- Fu, C. (2011). Training, search and wage dispersion. *Review of Economic Dynamics*, 14(4):650–666.
- Gopinath, G., Kalemli-Özcan, Ş., Karabarbounis, L., and Villegas-Sanchez, C. (2017). Capital allocation and productivity in south europe. *The Quarterly Journal of Economics*, 132(4):1915–1967.
- Gregory, V. (2021). Firms as learning environments: Implications for earnings dynamics and job search. *Federal Reserve Bank of St.Louis*.
- Gulyas, A. et al. (2018). Firm dynamics and labor market sorting.
- Guner, N., Parkhomenko, A., and Ventura, G. (2018). Managers and productivity differences. *Review of Economic Dynamics*, 29:256–282.
- Guner, N., Ventura, G., and Xu, Y. (2008). Macroeconomic implications of size-dependent policies. *Review of Economic Dynamics*, 11(4):721–744.
- Heise, S. and Porzio, T. (2021). The aggregate and distributional effects of spatial frictions. Technical report, National Bureau of Economic Research.
- Hopenhayn, H. A. (2014). Firms, misallocation, and aggregate productivity: A review. *Annual Review of Economics*, 6(1):735–770.
- Hopenhayn, H. A. (2016). Firm size and development. *Economía*, 17(1):27–49.
- Hornstein, A., Krusell, P., and Violante, G. L. (2011). Frictional wage dispersion in search models: A quantitative assessment. *American Economic Review*, 101(7):2873–98.
- Hsieh, C.-T., Hurst, E., Jones, C. I., and Klenow, P. J. (2019). The allocation of talent and us economic growth. *Econometrica*, 87(5):1439–1474.
- Hsieh, C.-T. and Klenow, P. J. (2009). Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448.
- Jarosch, G. (2021). Searching for job security and the consequences of job loss. Technical report, National Bureau of Economic Research.
- Jørgensen, T. H. (2020). Sensitivity to calibrated parameters. *The Review of Economics and Statistics*, pages 1–30.

- Jovanovic, B. (2014). Misallocation and growth. *American Economic Review*, 104(4):1149–71.
- Karahan, F., Ozkan, S., and Song, J. (2022). Anatomy of lifetime earnings inequality: Heterogeneity in job ladder risk vs. human capital. *FRB St. Louis Working Paper*, (2022-2).
- Lagakos, D., Moll, B., Porzio, T., Qian, N., and Schoellman, T. (2018). Life cycle wage growth across countries. *Journal of Political Economy*, 126(2):797–849.
- Lentz, R. and Roys, N. (2015). Training and search on the job. Technical report, National Bureau of Economic Research.
- Lise, J., Meghir, C., and Robin, J.-M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, 19:63–87.
- Ma, X., Nakab, A., Vidart, D., et al. (2021). Human capital investment and development: The role of on-the-job training. *University of Connecticut, Department of Economics*.
- Martellini, P. and Menzio, G. (2021). Jacks of all trades and masters of one: Declining search frictions and unequal growth. *American Economic Review: Insights*, 3(3):339–352.
- McKenzie, D. (2017). How effective are active labor market policies in developing countries? a critical review of recent evidence. *The World Bank Research Observer*, 3(2):127–261.
- Merz, M. and Yashiv, E. (2007). Labor and the market value of the firm. *American Economic Review*, 97(4):1419–1431.
- Midrigan, V. and Xu, D. Y. (2014). Finance and misallocation: Evidence from plant-level data. *American economic review*, 104(2):422–58.
- Moll, B. (2014). Productivity losses from financial frictions: Can self-financing undo capital misallocation? *American Economic Review*, 104(10):3186–3221.
- Poschke, M. (2018). The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics*, 10(3):1–41.
- Poschke, M. (2019). Wage employment, unemployment and self-employment across countries. *working paper*.
- Postel-Vinay, F. and Robin, J.-M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. *Econometrica*, 70(6):2295–2350.

- Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720.
- Restuccia, D. and Rogerson, R. (2017). The causes and costs of misallocation. *Journal of Economic Perspectives*, 31(3):151–174.
- Song, J., Price, D. J., Guvenen, F., Bloom, N., and Von Wachter, T. (2019). Firming up inequality. *The Quarterly journal of economics*, 134(1):1–50.
- Yonzan, N., Milanovic, B., Morelli, S., and Gornick, J. (2022). Drawing a line: comparing the estimation of top incomes between tax data and household survey data. *The Journal of Economic Inequality*, 20(1):67–95.

Online Appendix

A Data

A.1 Earnings inequality

We use data from three different sources: the IPUMS-International, the EU Statistics on Income and Living Conditions (EU-SILC) and the The Luxembourg Income Study Database (LIS) datasets. In all three sources, the sample is restricted to individuals between 18 and 64 who are not students. Table 7 provides a list of countries, years and sources.

IPUMS-International provides harmonized census microdata for a large set of countries. Details on sampling and stratification design and harmonization of variables across countries can be in <https://international.ipums.org/international/>. The surveys identify whether or not a respondent was working over a specified period of time (variable "EMP-STAT"). When this is missing, we use information on the average number of hours worked per week, in total (variable "HRSWORK1") or in the main job (variable "HRSMAIN"), and define a person as employed if she reports positive number of hours worked in at least one of these variables. To distinguish between employees and self-employed workers, we use the variable "INCWAGE", which records the respondent's weekly, monthly or annual wage and salary income. We annualize weekly or monthly wage and salary income estimates by multiplying them by 52 or 12 respectively. The final sample includes all individuals with non missing information on employment and wage and salary income. We consider anyone with strictly positive wage and salary income as an employee.

The EU Statistics on Income and Living Conditions (EU-SILC) database collects comparable cross-sectional microdata on income and other living conditions of households in European Union countries. Details are provided in <https://ec.europa.eu/eurostat/web/income-and-living-conditions/data>.

For each household member, the survey collects information on several demographic characteristics - age, gender, marital status, citizenship and head of households - education attainment, and labor market outcomes. The survey uses self-declared current labor market status to distinguish working and non-working individuals (variable PL040). Employee income is defined as the total cash remuneration payable by an employer to an employee in return for work done by the latter during the income reference period, typically calendar year prior to the survey date (variable PY010G). The sample includes all individuals with non-missing information on employment status and employee income, and define as earners anyone with a strictly positive employee income.

The Luxembourg Income Study Database (LIS) collects and harmonizes household and person-level micro-data for about 50 countries in Europe, North America, Latin America, Africa, Asia, and Australia, spanning five decades. Besides basic demographics, this dataset provides information on employment, labor and capital income, social benefits, private transfers, taxes and social security contributions. More information about sample and stratification is reported at www.lisdatacenter.org/our-data/lis-database. The dataset identifies the labor market status (employed, unemployed or inactive) of the surveyed individual (variable "lfs": labour force status). If this information is missing, we use information on the average number of hours worked per week (variable "hourstot": total weekly hours worked), or the number of weeks worked per year (variable "weeks": annual weeks worked), and define a person as employed if she reports positive number of hours or weeks worked. Conditional on being employed, we distinguish whether the surveyed individual is employed as a salaried worker or not in his main job using the variable "status1" (status in employment, main job). For these individuals, we compute a measure of annual wage and salary earnings using the variable "pi11" (Wage income, annual), which include any monetary payments received from regular and irregular dependent employment, i.e., cash wage and salary income (gross of social security contributions and income taxes) and monetary supplements to the basic wage, such as overtime pay, employer bonuses, 13th-month bonus, profit-share, tips.

Table 7: Data source

Country	Year	Source
Australia	1981, 1985, 1989, 1995, 2001, 2003, 2004, 2008, 2010, 2014	LIS
Austria	1994, 1997, 2000, 2004, 2005, 2007, 2010, 2013, 2016	EU-SILC, LIS
Belgium	1995, 1997, 2000, 2003-216	EU-SILC, LIS
Bulgaria	2007, 2009	EU-SILC
Brazil	2006, 2009, 2011, 2013, 2016	LIS
Canada	1981, 1987, 1991, 1994, 1997, 1998, 2004, 2007, 2010, 2012-2016	LIS
Chile	1990, 1992, 1994, 1996, 1998, 2000, 2003, 2006, 2009, 2011, 2013, 2015	LIS
China	2002, 2013	LIS
Colombia	2004, 2007, 2010, 2013, 2016	LIS
Croatia	2010	EU-SILC
Cyprus	2005, 2010	EU-SILC
Czech republic	1992, 1996, 2002, 2004, 2006, 2007, 2009, 2010, 2013, 2016	EU-SILC, LIS
Denmark	1987, 1992, 1995, 2000, 2004, 2005, 2007, 2009, 2010, 2013, 2016	EU-SILC, LIS
Dominican Republic	1981, 2007	IPUMS, LIS
Egypt	2012	LIS
Estonia	2004, 2005, 2007, 2010, 2013, 2016	EU-SILC, LIS
Finland	1987, 1991, 1995, 2000, 2004, 2005, 2007, 2009, 2010, 2013, 2016	EU-SILC, LIS
France	1978, 1984, 1989, 1994, 2000, 2005, 2010	EU-SILC, LIS
Germany	1973, 1978, 1981, 1983, 1984, 1987, 1989, 1991, 1994, 1995, 1998, 2000-2016	EU-SILC, LIS
Georgia	2010, 2013, 2016	LIS
Greece	1995, 2000, 2004, 2005, 2007, 2009, 2010, 2013, 2016	EU-SILC, LIS
Guatemala	2006, 2011, 2014	LIS
Hungary	1991, 1994, 1999, 2005-2007, 2009, 2010, 2012, 2015	EU-SILC, LIS
Iceland	2004, 2005, 2007, 2010	EU-SILC , LIS
Israel	1979, 1986, 1992, 1995, 1997, 2001-2016	IPUMS, LIS
Italy	1986, 1987, 1989, 1991, 1993, 1995, 1998, 2000, 2004, 2005, 2008-2010, 2014	EU-SILC, LIS
India	1993, 1999, 2004, 2011	IPUMS, LIS
Indonesia	1976, 1995	IPUMS
Ireland	1994-1996, 2000, 2002-2013	EU-SILC, LIS
Jamaica	1981, 1991, 2001	IPUMS
Japan	2008, 2010, 2013	LIS
Latvia	2006, 2010	EU-SILC
Lithuania	2006, 2009-2016	EU-SILC, LIS
Luxembourg	1985, 1991, 1994, 1997, 2000, 2004, 2005, 2007, 2010, 2013	EU-SILC, LIS
Malta	2007, 2010	EU-SILC
Mexico	1984, 1989, 1992, 1994, 1996, 1998, 2000, 2002, 2004, 2008, 2010, 2012, 2014, 2016	LIS
Netherlands	1983, 1987, 1990, 1993, 1999, 2004, 2006, 2007, 2010, 2013	EU-SILC, LIS
Norway	1979, 1986, 1991, 1995, 2000, 2004, 2005, 2007, 2010, 2013, 2016	EU-SILC, LIS
Panama	2007, 2010, 2013, 2016	IPUMS, LIS
Peru	2004, 2007, 2010, 2013, 2016	LIS
Poland	1986, 1992, 1995, 2004, 2005, 2007, 2009, 2010, 2013, 2016	EU-SILC, LIS
Portugal	2005, 2010	EU-SILC
Puerto Rico	1990, 2000, 2005	IPUMS
Paraguay	2000, 2004, 2007, 2010, 2013, 2016	LIS
Romania	2007, 2009	EU-SILC
Russia	2000, 2004, 2007, 2010, 2011, 2013-2016	LIS
Serbia	2006, 2010, 2013, 2016	LIS
Slovakia	1992, 2004, 2006, 2007, 2009, 2010, 2013, 2014-2016	EU-SILC, LIS
Slovenia	1997, 1999, 2004, 2006, 2007, 2009, 2010, 2012, 2015	EU-SILC, LIS
Spain	1980, 1990, 1995, 2000, 2004, 2005, 2007, 2009, 2010, 2013, 2016	EU-SILC, LIS
Sweden	1975, 1981, 1987, 1992, 1995, 2000, 2005, 2009	EU-SILC, LIS
Switzerland	1982, 1992, 2006-2016	EU-SILC, LIS
Trinidad and Tobago	2000	IPUMS
USA	1974, 1979, 1980, 1986, 1990-2016	IPUMS, LIS
Uruguay	2004, 2006, 2007, 2010, 2013, 2016	IPUMS, LIS
United Kingdom	1974, 1979, 1986, 1991, 1994-2016	EU-SILC, LIS
South Africa	2008, 2010, 2012, 2015	LIS

A.2 Robustness

In this section of the Online Appendix, we show that the cross-country inequality patterns are robust. In Table 8 we report the outcomes of the following regression:

$$y_{it} = \alpha + \mu_t + \beta \log \text{GDP p.c.}_{it} + \gamma X_{it} + \epsilon_{it}$$

where y_{it} are various measures of earnings inequality for country i at time t , μ_t are time fixed effects and X_{it} are various controls, including average years of schooling, women labor force participation rate, share of self-employment, shares of agricultural and industrial employment, average number of hours worked per year, aggregate capital stock and the value of trade as a share of GDP.

Table 8: Earnings inequality across countries

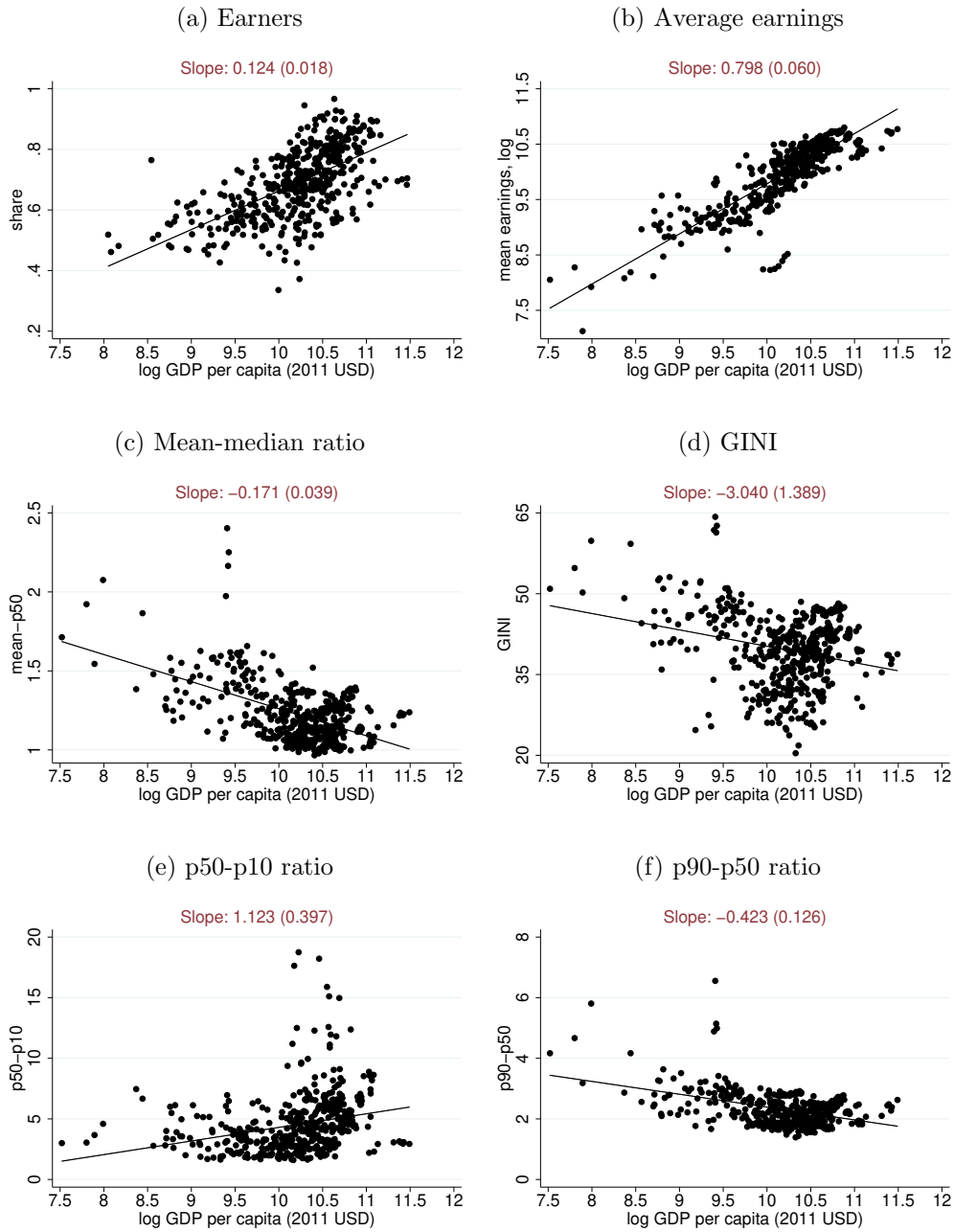
	Mean-median ratio			GINI		
	(1)	(2)	(3)	(1)	(2)	(3)
log GDP p.c.	-0.171*** (0.0386)	-0.189*** (0.0429)	-0.229*** (0.0549)	-3.040** (1.389)	-3.346** (1.493)	-4.551* (2.603)
Observations	497	497	420	430	497	420
R-squared	0.286	0.420	0.690	0.067	0.194	0.499
Time FE		✓	✓		✓	✓
Controls			✓			✓

	p50-p10 ratio			p90-p50 ratio		
	(1)	(2)	(3)	(1)	(2)	(3)
log GDP p.c.	1.123*** (0.397)	1.248*** (0.440)	1.797*** (0.437)	-0.423*** (0.126)	-0.469*** (0.136)	-0.570*** (0.203)
Observations	497	497	420	497	497	420
R-squared	0.069	0.136	0.308	0.201	0.323	0.557
Time FE		✓	✓		✓	✓
Controls			✓			✓

Notes: GDP p.c. is expressed in 2011 USD. Controls include average years of schooling (Penn-Word), women labor force participation rate (World Bank), share of self-employment (World-Bank), shares of agricultural and industrial employment (World Bank), average number of hours worked per year (Penn-Word), aggregate capital stock (Penn-Word), and the value of trade as a share of GDP (World Bank). Standard errors are robust and clustered as country level. *** p<0.01, ** p<0.05, * p<0.1

In Figure 19 we report the raw shares of earners, average earnings and inequality.

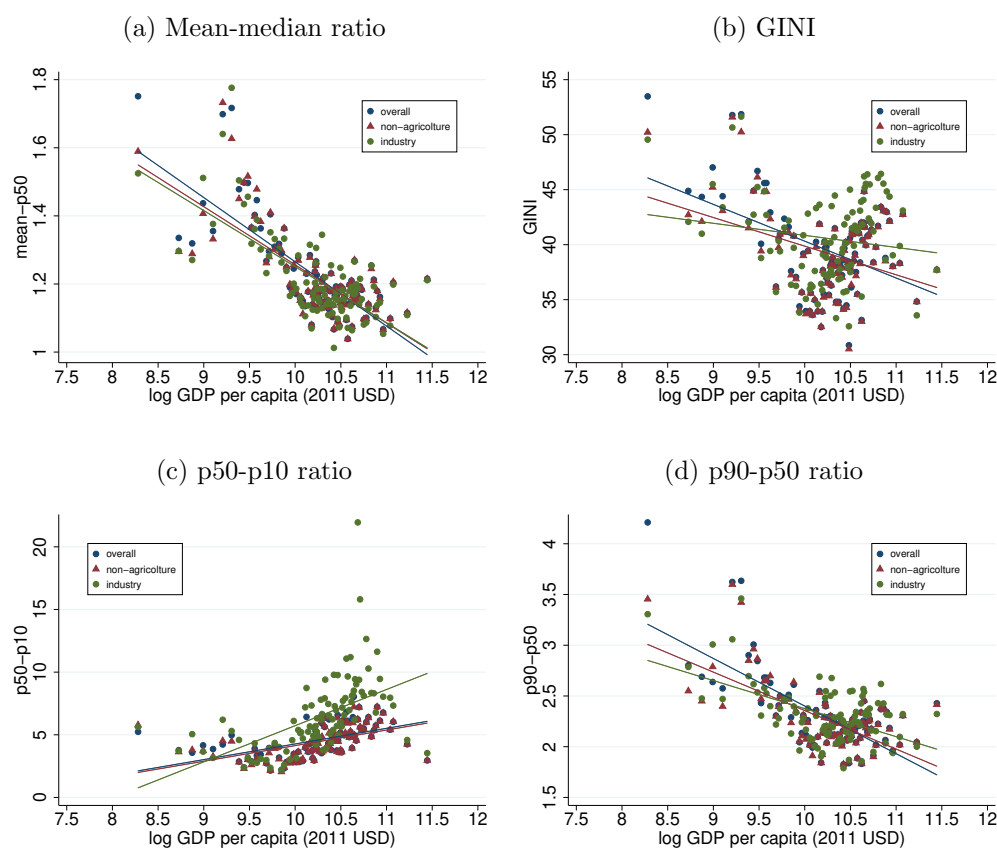
Figure 19: Earners and earnings across countries, raw data



Notes: Each dot corresponds to country-year data points. In red we report the estimated slope in the regression. Standard errors (in parenthesis) are robust and clustered at country level. Source: IPUMS, EU-SILC, LIS and author's calculations.

In Figure 20 we report the mean-median earnings ratio, the GINI coefficient, and the p50-p10 and p90-p50 earnings ratios across countries separately for individuals 1) working in any sector (blue dots), ii) working in any non-agriculture sectors (red dots), and iii) working only in the industrial sector (green dots).

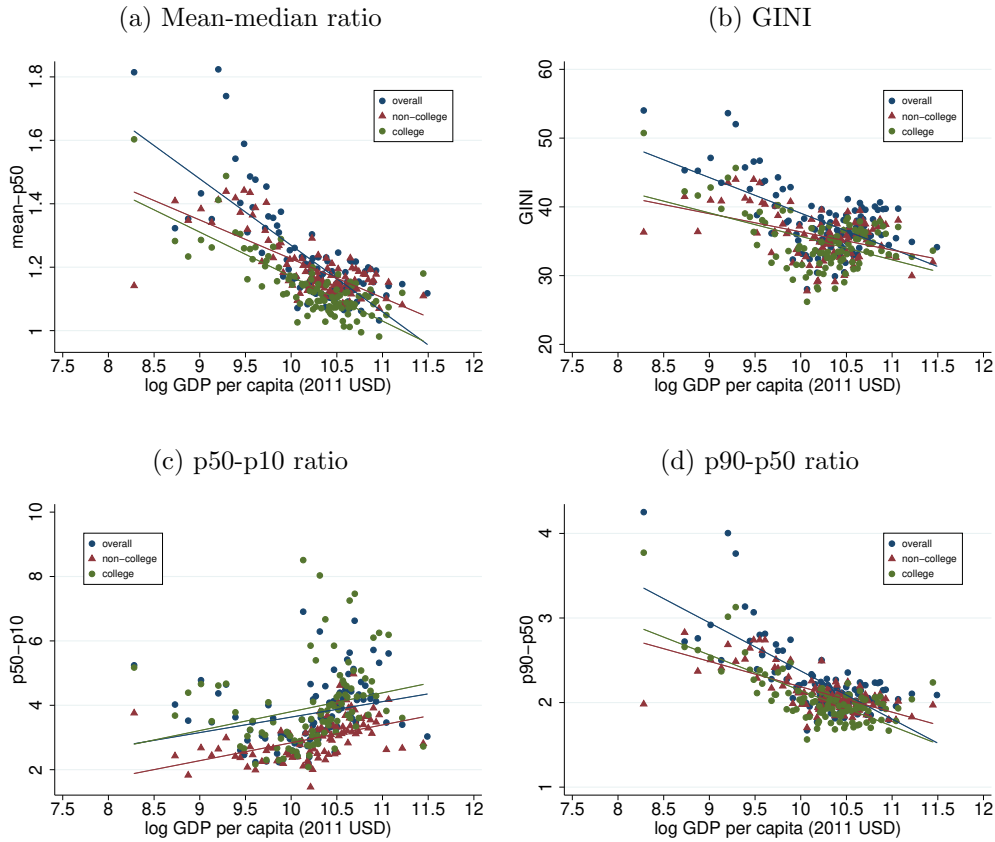
Figure 20: Earnings inequality across countries, by sectors



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year- fixed effects. Source: IPUMS, EU-SILC, LIS and author's calculations.

In Figure 21 we report different measures of earnings inequality across countries separately for individuals i) with (red dots) and ii) without a college degree (green dots).

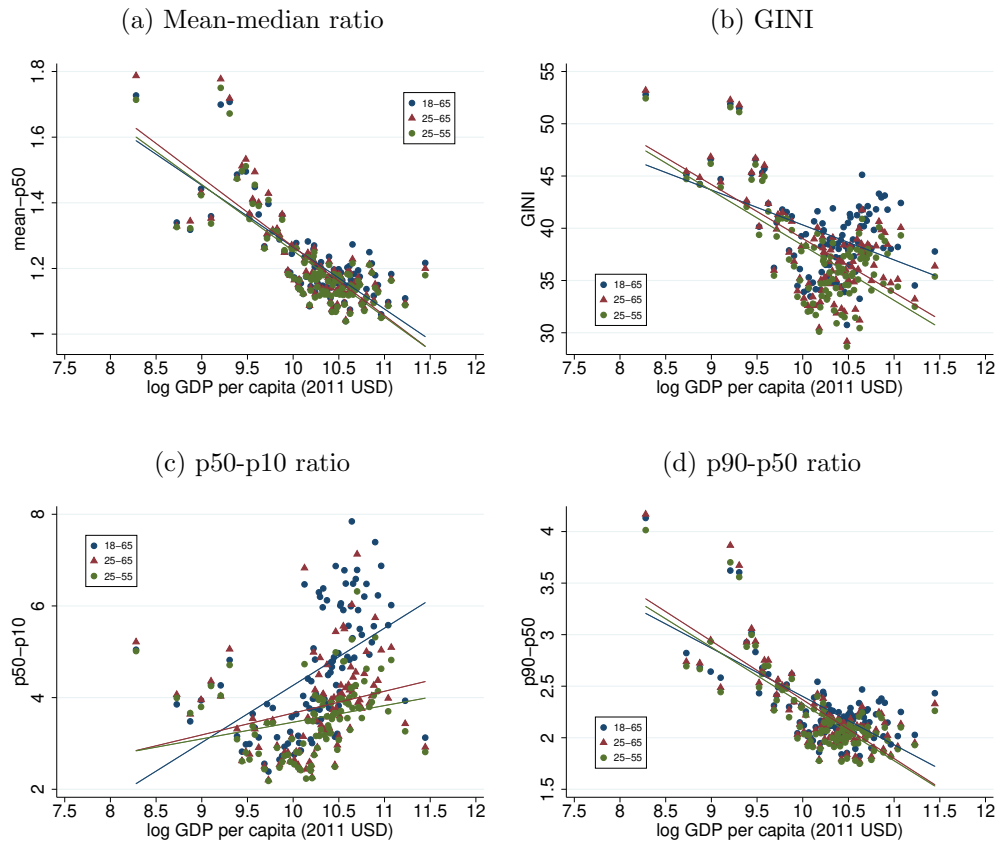
Figure 21: Earnings inequality across countries, by education



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year- fixed effects. Source: IPUMS, EU-SILC, LIS and author's calculations.

In Figure 22 we consider workers within different age categories: i) the 18-65 years old (blue dots), ii) 25-65 years old (red dots), and iii) the group of prime age individuals, 25-55 years old (green dots)

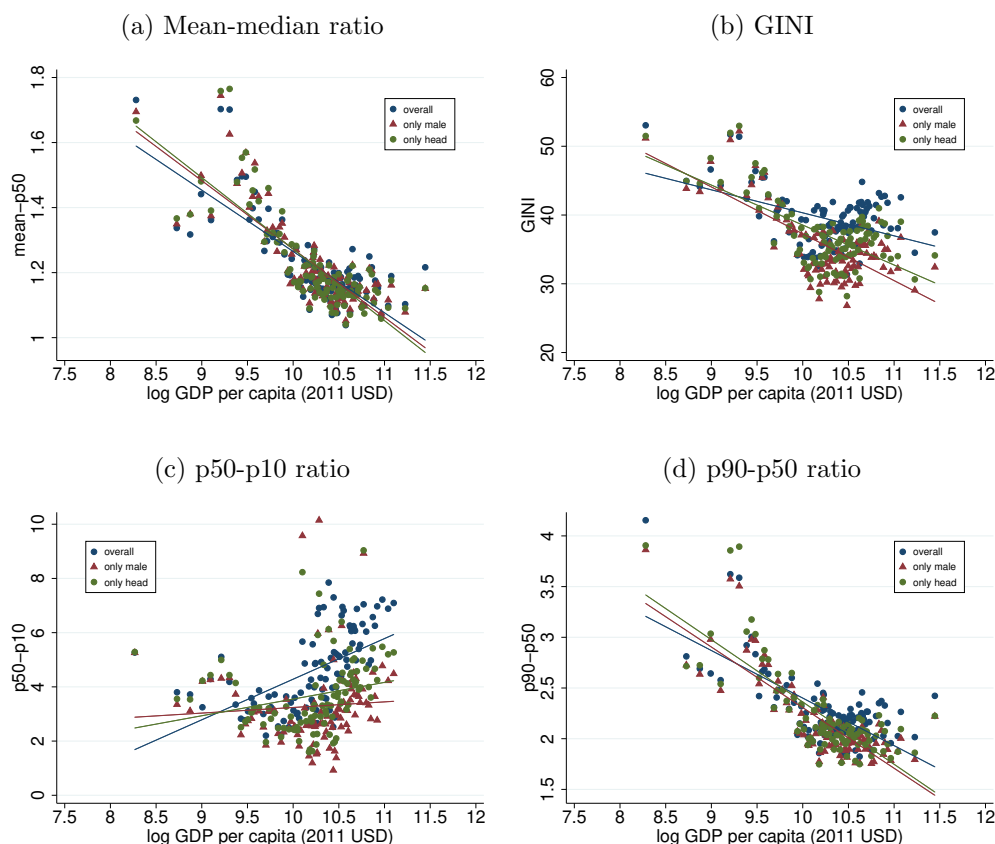
Figure 22: Earnings inequality across countries, by age



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year- fixed effects. Source: IPUMS, EU-SILC, LIS and author's calculations.

In Figure 23 we restrict the sample to workers in different demographic groups: i) all (blue dots), ii) males (red dots), and iii) household heads, regardless their gender (green dots)

Figure 23: Earnings inequality across countries, by demographic groups



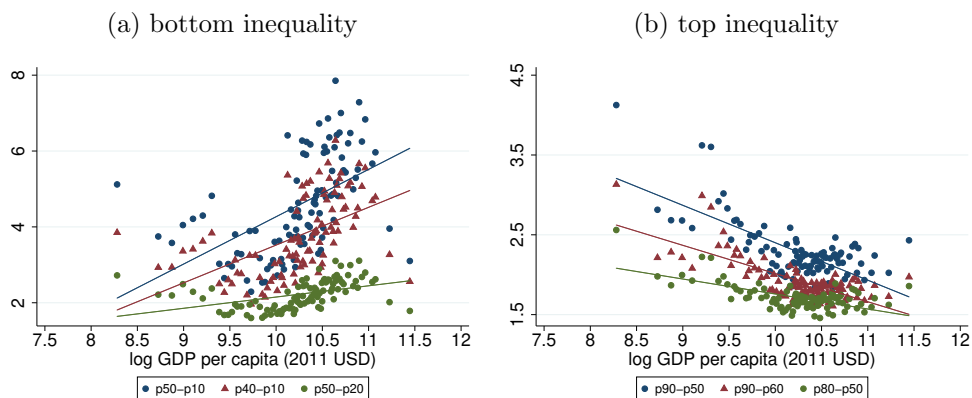
Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year- fixed effects. Source: IPUMS, EU-SILC, LIS and author's calculations.

Finally, in Figure 24 we report alternative measures of bottom (the p40-p10 and p50-p20 earnings ratios) and top (the p90-p60 and p80-p50) earnings ratios.

A.3 On-the-job Training

The World-Bank Enterprise Survey (WB-ES) is a firm-level survey of a representative sample of an economy's private sector. It is a cross-sectional survey and targets formal (registered) companies with 5 or more employees, operating in the manufacturing and services sectors. For more details about the sampling methodology, see

Figure 24: Earnings inequality across countries



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year- fixed effects. Source: IPUMS, EU-SILC, LIS and author’s calculations.

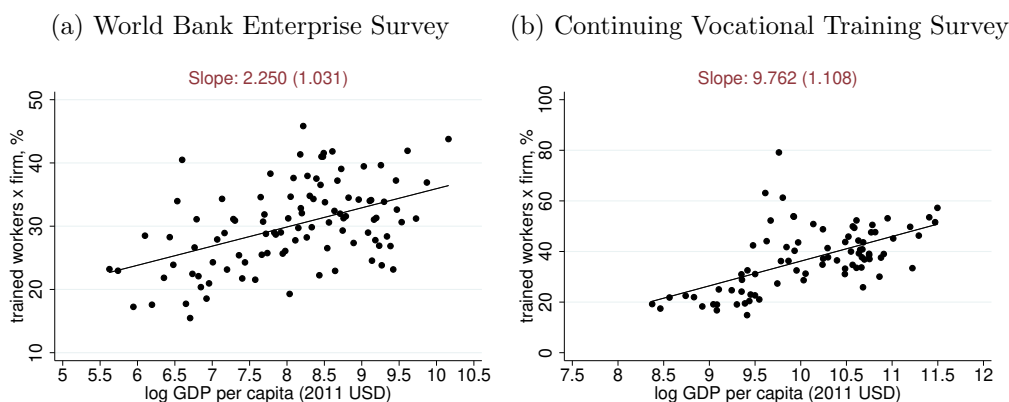
<https://www.enterprisesurveys.org/en/methodology>. For each firm, the dataset records demographic information (age, region of operation, ownership status), number of employees, annual sales, annual wage bills (firm-level average wage is constructed using wage bill divided by the number of employees). The survey also provides different measures of training provision: 1) whether a firm has provides training to all or some of its workforce, and 2) the share of workforce who received training in a given year. To construct our main empirical evidence, we use the March-04-2019 survey release. This version of the survey covers firms in 139 countries surveyed during the period 2006-2018. We remove countries lacking information on firm-level training, or countries where firm-level number of employees or wage bills are either missing, or inconsistent with the aggregate indicators reported by the World Bank. We remove also Sweden (which is instead included in the Eurostat CVTS dataset). This leaves us with 122 countries: Afghanistan, Albania, Angola, Antigua and Barbuda, Argentina, Armenia, Azerbaijan, Bahamas, Bangladesh, Barbados, Belarus, Belize, Bhutan, Bolivia, Bosnia-Herzegovina, Botswana, Brazil, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Cape-Verde, Central African Republic, Chad, Chile, China, Colombia, Congo, Costa Rica, Cote d’Ivoire, Croatia, Czech Republic, D.R.C., Djibouti, Dominica, Dominican Republic, Ecuador, Egypt, El Salvador, Eritrea, Estonia, Eswatini, Ethiopia, Fiji, Gabon, Gambia, Georgia, Ghana ,Grenada, Guatemala, Guinea, Guinea Bis-sau, Guyana, Honduras, Hungary, India, Indonesia, Iraq, Israel, Jamaica, Jordan, Kaza-khstan, Kenya, Kosovo, Kyrgyzistan, Lao P.D.R., Latvia, Lebanon, Liberia, Lithuania, Macedonia, Madagascar, Malaysia, Mali, Mauritius, Mexico, Micronesia, Moldova, Mon-golia, Montenegro, Morocco, Myanmar, Namibia, Nicaragua, Nigeria, Pakistan, Panama,

Papua New Guinea, Paraguay, Peru, Philippines, Poland, Romania, Russia, Samoa, Serbia, Sierra Leone, Slovakia, Slovenia, Solomon Islands, South Africa, South Sudan, Sri Lanka, St.Kitts and Nevis, St.Lucia, St.Vincent and Grenadines,, Tanzania, Tonga, Trinidad and Tobago, Tunisia, Turkey, Uganda, Ukraine, Uruguay, Uzbekistan, Vanuatu, Vietnam, West-Bank, Yemen, Zambia and Zimbabwe.

The Continuing Vocational Training Survey (CVTS) is a firm-level survey that covers a representative sample of formal enterprises with 10 or more employees in 27 EU countries plus Norway, North Macedonia and United Kingdom, for the years 2005, 2010 and 2015. Besides demographic information, the survey includes information about firm-level provision of on-the-job vocational training, and share of employees participating in vocational training for each firms. To construct our main empirical evidence, we use the aggregate statistics reported by the Eurostat, available here. Statistics are constructed for each country and year overall, and broken by firm size categories.

We merge the WB-ES and the CVTS with information on GDP per capita and population from the World Bank Indicators. GDP per capita is expressed in constant 2011 USD. Finally, we use the World Bank PPP deflator to convert firm-level average wages from local currency units to current international dollars.

Figure 25: Training provision across countries



Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year-fixed effects. In red we report the estimated slope in the regression. Robust standard errors are in parenthesis.

Source: World-Bank Enterprise Survey and Eurostat Education and Training Dataset.

A.3.1 Further empirical evidence on job training

Using the WB-ES, we can measure the share of trained workers in firm i as

$$\text{trained-workers}_{it} = \mathbf{1}_{it}^{\text{training}} \frac{\% \text{permanent full-time workers trained}_{it}}{100}$$

The Eurostat reports this variable constructed using data from the CVTS. Figures 25 (a) and (b) show the average share of workers within each firm receiving formal job training. In both figures, the measure of training provision is scattered over the country-average real GDP per capita. The correlation between the share of workers trained and the country log GDP per capita for is between 0.49 for more developed countries and 0.57 for developing countries. The slope coefficient from a regression of the average share of workers trained within each country and log GDP per capita is around 0.11 for developing countries, and 0.08 for more developed countries, and is statistically significant at the five percent level in both cases. These coefficients implies that one log point higher GDP per capita is associated with about 10% percent more workers receiving training. Table 9 reports the share of trained worker by firm size. In both data sets, larger firms provide training to a larger set of their workforce.

Table 9: Job training across firm size

		Trained workers within firms, %						
		WB-ES				CVTS		
		LAC	ME+AFR	ASIA	others		EU15	non-EU15
Firm size						Firm size		
(# employees)						(# employees)		
<20		34.36	21.01	27.95	29.63			
20-49		40.06	25.56	29.72	30.18	<50	29.31	21.96
50-249		44.35	26.68	35.51	30.36	50-249	37.92	30.13
250-449		52.51	30.30	32.22	28.86	≥250	49.71	46.25
≥500		50.73	32.37	34.34	28.98			

Notes: Each entry denotes to the average share of workers (in percent) receiving training within firms reporting to provide training, separately for firms with different size (number of employees), and different groups of country. Source: World-Bank Enterprise Survey (WB-ES) and Eurostat Education and Training Dataset (CVTS).

B Model

B.1 The Surplus function, Hiring and Training decisions

Each match has a potential surplus, $S(z, \xi, a)$, given by

$$S(z, \xi, a) = M(z, \xi, a) - J^u(a), \quad (15)$$

where $M(z, \xi, a)$ denotes the joint match value at the beginning of the period, equal to the sum of the value of employment $J^e(z, \xi, a)$ and the match value for the firm $V(z, \xi, a)$,

$$\begin{aligned} M(z, \xi, a) &= J^e(z, \xi, a) + V(z, \xi, a) \\ &= \mathbf{1}^h(z, \xi, a)[J^{e,h}(z, \xi, a) + V^h(z, \xi, a)] + (1 - \mathbf{1}^h(z, \xi, a))J^{u,h}(a). \end{aligned}$$

Using equations (5) and (6), $M(z, \xi, a)$ can be express using the following recursive formulation

$$M(z, \xi, a) = \mathbf{1}^h(z, \xi, a)M^h(z, \xi, a) + (1 - \mathbf{1}^h(z, \xi, a))J^{u,h}(a) \quad (16)$$

where $M^h(z, \xi, a)$ is the match value at the end of the period, defined as

$$\begin{aligned} M^h(z, \xi, a) &= r(z, a) + \frac{(1 - \delta_w)}{1 + r}(1 - (1 - \delta_f)(1 - \delta_s))J^{u,h}(a) + \\ &\frac{(1 - \delta_w)}{1 + r}(1 - \delta_f)(1 - \delta_s) [-\mathbf{1}^t(z, \xi, a)\xi + (1 - p^h(z, \xi, a))M(z, \xi, a) + p^h(z, \xi, a)M(z, \xi, a + 1)] \end{aligned}$$

Combining equations (17) and (4), we can write the surplus function as

$$S(z, \xi, a) = \max\{0, S^h(z, \xi, a)\} \quad (17)$$

where $S^h(z, \xi, a)$ is the difference between the value of an active match and the value of being non-employed, i.e.

$$S^h(z, \xi, a) = M^h(z, \xi, a) - J^{u,h}(a). \quad (18)$$

or equivalently

$$\begin{aligned} S^h(z, \xi, a) &= r(z, a) + \frac{(1 - \delta_w)}{1 + r}(1 - (1 - \delta_f)(1 - \delta_s))J^{u,h}(a) - J^{u,h}(a) + \\ &\frac{(1 - \delta_w)}{1 + r}(1 - \delta_f)(1 - \delta_s) [-\mathbf{1}^t(z, \xi, a)\xi + (1 - p^h(z, \xi, a))M(z, \xi, a) + p^h(z, \xi, a)M(z, \xi, a + 1)]. \end{aligned}$$

A match between a worker with skill $a \in \mathcal{A}$ and a firm with productivity $z \in \mathcal{Z}$ and training cost $\xi \in \mathcal{E}$ is formed upon contact (or kept alive if it already exists) as long as the match surplus is positive, i.e.

$$\mathbf{1}^t(z, \xi, a) = \arg \max_{\mathbf{1}^t \in \{0,1\}} \mathbf{1}^t p^t [M(z, \xi, a + 1) - M(z, \xi, a)] - \mathbf{1}^t \xi,$$

where $M(z, \xi, a)$ is defined in equation (16), which implies the following indicator function for training provision:

$$\mathbf{1}^t(z, \xi, a) = \begin{cases} 1 & \text{if } p^t[M(z, \xi, a + 1) - M(z, \xi, a)] > \xi \\ 0 & \text{otherwise.} \end{cases}$$

B.2 Equilibrium

A stationary recursive competitive equilibrium for this economy consists of workers' value functions for employment and unemployment, firms' value functions for active jobs, policy functions for job creation, training, entry and vacancies posted, wage schedule, job contact probabilities for workers and firms, unemployment rate, distribution of employed and unemployed workers across states, distribution of open vacancies and firms across states, such that:

1. *optimality*: the value functions attain their maximum;
2. *bargaining*: the wage schedule is the solution of the problem (14);
3. *training*: training decision is the solution of the problem (13);
4. *market clearing*: goods and labor market are cleared;
5. *measure of entrants*: for all Borel sets $\mathcal{Z} \times \mathcal{E} \subset \mathcal{R}^+ \times \mathcal{R}^+$ it must be that

$$E(\mathcal{Z} \times \mathcal{E}) = M \int_{z \in \mathcal{Z}} \int_{\xi \in \mathcal{E}} \mathbf{1}^e(z, \xi) \psi_z(z) \psi_\xi(\xi) dz d\xi$$

where $\mathbf{1}^e(z, \xi)$ is the solution to the problem of potential entrant (8).

6. *measure of incumbent*: for all Borel sets $\mathcal{Z} \times \mathcal{E} \subset \mathcal{R}^+ \times \mathcal{R}^+$ it must be that

$$\Gamma(\mathcal{Z} \times \mathcal{E}) = \frac{1}{\delta_f} E(\mathcal{Z} \times \mathcal{E})$$

7. *aggregate consistency*: workers' and vacancies' distributions replicate themselves through workers' and firms' policy functions.

B.3 Solution algorithm

To compute the value functions, we discretize the state space using using 50 equally-spaced grid points for firm productivity, 20 equally-spaced grid points for firm-specific training costs,

and 60 equally-spaced grid points for workers human capital. While we directly calibrate the boundaries for training costs, we fix minimum and maximum (log) productivity and (log) human capital to -4 and 4 respectively, covering 99.9% of both calibrated distributions.¹⁴

To find an equilibrium, we employ the following algorithm:

1. Formulate a guess for the labor market tightness, θ^0 . Use the definition of matching function to compute the workers' job contact rate, $\phi_w^0 = \chi(1 + (\theta^0)^{-\eta})^{-\frac{1}{\eta}}$, and job contact rate for firms, $\phi_f^0 = \chi(1 + (\theta^0)^\eta)^{-\frac{1}{\eta}}$.
2. Formulate a guess for the distribution of vacancies, $\psi_v^0(z, \xi)$
 - 2.1. Given ϕ_w^0 and $\psi_v^0(z, \xi)$, solve for the surplus function, $S^h(z, \xi, a)$.
 - 2.2. Obtain the policy functions for job creation, $\mathbf{1}^h(z, \xi, a)$ and on-the-job training $\mathbf{1}^t(z, \xi, a)$
 - 2.3. Use ϕ_w^0 , $\psi_v^0(z, \xi)$, $\mathbf{1}^h(z, \xi, a)$ and $\mathbf{1}^t(z, \xi, a)$ to simulate a large panel of workers and construct a distribution of non-employed workers over human capital, $\psi_a^u(a)$, and the aggregate measure of workers who are non-employed, U .
 - 2.4. Given ϕ_f^0 , $\mathbf{1}^h(z, \xi, a)$, $\psi_a^u(a)$, and the bargaining splitting rule, solve the vacancy posting problem of the firm, $v(z, \xi)$.
 - 2.5. Compute the value at entry, $\Pi(z, \xi)$, and obtain a solution to the entry decision, $\mathbf{1}^e(z, \xi)$
 - 2.6. Given $v(z, \xi)$ and $\mathbf{1}^e(z, \xi)$, construct a new guess for the distribution of vacancy, $\psi_v^1(z, \xi)$
 - 2.7. Check for convergence:
 - if $\psi_v^1(z, \xi)$ and $\psi_v^0(z, \xi)$ are close enough, store $\psi_v^*(z, \xi) = \psi_v^1(z, \xi)$ and go ahead.
 - if not, set if $\psi_v^0(z, \xi) = \psi_v^1(z, \xi)$ and go back to step 2
 - 2.8. Iterate till convergence
3. Compute the measure of entrant firms

$$M = M_e \int_{z \in \mathcal{Z}} \int_{\xi \in \mathcal{E}} \mathbf{1}^e(z, \xi) \psi_z(z) \psi_\xi(\xi) dz d\xi \quad (19)$$

and use stationarity condition to compute total number of firms

$$N = \frac{M}{\delta_f} \quad (20)$$

¹⁴This makes the parameters p^e , p^t and p^d interpretable as the probability of having a 8.5% change in human capital.

4. Construct the aggregate measure of vacancy posted as $v = N\bar{v}$, with the average number of vacancy posted given by

$$\bar{v} = \int_{z \in \mathcal{Z}} \int_{\xi \in \mathcal{E}} \mathbf{1}^e(z, \xi) v(z, \xi) \psi_z(z) \psi_\xi(\xi) dz d\xi \quad (21)$$

5. Use U , v and the definition of matching function to obtain a new guess for the job contact rate of workers, θ^1

6. Check for convergence:

- if θ^1 and θ^0 are close enough, store $\theta^* = \theta^1$ and go ahead.
- if not, set if $\theta^0 = \theta^1$ and go back to step 1

7. Iterate till convergence

Use ϕ_w^* , ϕ_f^* , $\psi_v^*(z, \xi)$, and relevant policy functions to simulate a large panel of firms and workers and construct firm-level and worker-level statistics

C Estimation

C.1 Estimation of matching elasticity

We estimate the matching elasticity outside of the main estimation algorithm. To compute quarterly new hires, we use employment gross inflows from the ONS Labor Force Survey Flows Estimates (dataset X02, available here). From the same source, we also obtain data on aggregate open vacancies (dataset AP2Y, available here) and stock of non-employed workers (dataset ANZ6, available here). The GMM minimizes the following function:

$$\hat{x} = \arg \max_{\{\eta, x_1, x_2, x_3\}} \left[\left(\frac{1}{T} \sum_{t=1}^T Z_t' \epsilon_t(x) \right)' W_T \left(\frac{1}{T} \sum_{t=1}^T Z_t' \epsilon_t(x) \right) \right],$$

where $\epsilon_t(x)$ denotes the moment conditions, given by

$$\epsilon_t(x) = \left[h_t - \frac{u_t v_t}{(u_t^\eta + v_t^\eta)^{\frac{1}{\eta}}} - \sum_{i=1}^3 x_i \mathbf{1}_t^{q=i} \right],$$

with h_t is the number of new hires at time t , v_t the number of open vacancies, and u_t the number of non-employed workers.

Table 10: Matching function estimation

Parameters	Description	Estimates	St.Error
η	Elasticity	0.542	0.013
$\mathbf{1}_t^{q=1}$	Dummy first quarter	64189.29	36374.74
$\mathbf{1}_t^{q=2}$	Dummy second quarter	44722.20	41908.83
$\mathbf{1}_t^{q=3}$	Dummy third quarter	59070.01	40683.91

We also remove seasonal effects by including quarter dummies. The vector of instruments, $Z'_t = [u_{t-4}, v_{t-4}, \mathbf{1}_t^{q=1}, \mathbf{1}_t^{q=2}, \mathbf{1}_t^{q=3}, \mathbf{1}_t^{q=4}]$ includes four lags for non-employment and active vacancies, while W_T is a weighting matrix. Hence, an estimate for η is obtained by simply minimising the distance between new hires implied by the matching function and the data. For the estimation, we use data from the first quarter of 2002 till the fourth quarter of 2019. This makes the total number of observations used equal to 68. Table 10 reports estimates and standard errors obtained using the robust GMM weighting matrix in the second step.

C.2 Model estimation

C.2.1 Data

Table 11 reports descriptive statistics for the sample of households in the Five-Quarter Longitudinal LFS. The sample is restricted to individuals between 22 and 62 who report to be currently employed at the time of interview.

The statistics used in the calibration are computed using the sample of employed workers with non-missing information on hourly pay, on-the-job training and tenure on the job. The ultimate sample is made of 85,524 observations. About 78% of the individuals reports to be full-time employed, and work on average 37 hours in a week. Around 25% of the respondents who are employed reports to have received on-the-job training in the current quarter. The LFS reports information for tenure on the job using indicators for whether an individual has been employed in the same firm for < 3 months, for a period $\in [3, 12)$ months, $\in [12, 24)$ months, and for ≥ 24 months.

The LFS also records average hourly pay in the current quarter for individuals who are employed. We remove all the observations reporting negative hourly pay, or hourly pay lower than 40% the statutory minimum wage in that year. Therefore we deflated it using a first stage regression where we control for year and quarters fixed effects, i.e.

$$w_{it}^h = \delta_{y(t)} + \delta_{q(t)} + \epsilon_{it}$$

Table 11: Descriptive Statistics

	Mean	SD	Min	Max	N
<i>Employed workers</i>					
Age	41.63	11.64	22	62	85,524
Female	0.506	0.500	0	1	85,524
Full-time	0.756	0.423	0	1	85,524
Hours worked	37.04	12.10	1	97	85,524
Log Hourly pay	2.385	0.599	0.025	7.248	85,524
Log Quarterly Earnings	8.457	0.824	3.956	13.39	85,524
Training	0.244	0.430	0	1	85,524
Tenure<3 months	0.038	0.191	0	1	85,524
Tenure∈[3,12) months	0.039	0.192	0	1	85,524
Tenure∈[12,24) months	0.109	0.311	0	1	85,524
Tenure≥24 months	0.815	0.388	0	1	85,524

where w_{it}^h denotes the hourly pay of individual i at time t while $\delta_{y(t)}$ and $\delta_{q(t)}$ are respectively year and quarter dummies for each time t . Hourly pay are then expressed in 2010-q1 LCU. This variable - together with weekly hours - allows us to construct average weekly earnings in the current quarter. Finally, we construct average quarterly earnings by multiplying average weekly earnings by 12.6, which accounts for the average number of weeks in a quarter.

C.2.2 Parameter Estimates

The model is solved at a quarterly frequency, and the population is normalized to one. A few parameters are determined based on available evidence or set to their data counterparts a priori, without solving the model. These parameters are reported in Table 12.

The remaining parameters are estimated by method of simulated moments to minimize the sum of square residuals between the model-implied values and the data for a set of worker- and firm-level targets. Let \bar{m} be a vector of data targets and $m(\vartheta)$ their model counterparts. Let $\bar{d}(\vartheta)$ be a vector of $g \geq \dim[\vartheta]$ moment conditions (deviations between model and the data), $\bar{d}(\vartheta) = m(\vartheta) - \bar{m}$. The vector of parameter values, $\hat{\vartheta}$, is given by

$$\hat{\vartheta} = \arg \min_{\vartheta \in \Theta} \mathcal{L}(\vartheta). \quad (22)$$

where $\mathcal{L}(\vartheta) = \bar{d}(\vartheta)'W\bar{d}(\vartheta)$, and W is a positive semi-definite matrix. To improve the stability of our estimator while maintaining consistency, we opted for the identity matrix instead of the optimal weighting matrix. See also Lise et al. (2016) for a similar approach.

Table 12: Parameters directly calibrated

Parameters	Description	Value	Sources/Targets
ζ	Correlated distortion	0	Assumption
κ	Productivity shifter	1	Normalization
χ	Matching efficiency	1	Normalization
r	Interest rate	0.0033	Annual return of 4%
δ_w	Workers retirement	0.0099	Life-span of 40 years, ages 22-62
δ_f	Firm exit	0.0253	Annual exit rate of 10.50% (ONS)
η	Matching function	0.5416	Estimated using GMM (Appendix C.1)

Notes: The entries show the parameters set a priori without simulating the model, and their sources and/or targets.

C.2.3 Algorithm

In the estimation algorithm we exploit the definition of matching function, i.e.

$$m(U, v) = \chi \frac{Uv}{(U^\eta + v^\eta)^{\frac{1}{\eta}}}$$

to treat the equilibrium market tightness, θ as a parameter to be estimated, and let the measure of potential entrants, M_e , be an equilibrium object, equal to the solution of the following equilibrium equation:

$$\theta = \frac{U}{\bar{v} \frac{M}{\delta_f}} \quad (23)$$

where M is defined in equation (19). To estimate the parameters, we follow this algorithm:

1. Guess the following set of parameters:

$$\vartheta^0 = \{c_e^0, \sigma_z^0, \underline{\xi}^0, \bar{\xi}^0, \lambda_1^0, \theta^0, \beta^0, \sigma_a^0, p^{d0}, p^{e0}, p^{t0}, b^0, \delta_s^0\}$$

Let $J = \dim[\vartheta^0]$.

2. Given θ^0 , compute job contact rate for workers, ϕ_w^0 , and firms, ϕ_f^0 as follows

$$\phi_w^0 = (1 + (\theta^0)^{-\eta})^{-\frac{1}{\eta}} \quad \text{and} \quad \phi_f^0 = (1 + (\theta^0)^\eta)^{-\frac{1}{\eta}}.$$

3. Proceed as in the solution algorithm, step 2 and obtain the equilibrium measure of potential entrants M_e using equations (19) and (23).
4. Use the parameter guesses together with M_e , $\psi_v^*(z, \xi)$, and the relevant policy functions to simulate a large panel of firms and workers.

5. Compute relevant moment condition using simulated data, i.e.

$$\bar{d}(\vartheta^0) = m(\vartheta^0) - \bar{m}$$

Let $g = \dim [\bar{d}(\vartheta^0)] \geq \dim [\vartheta]$.

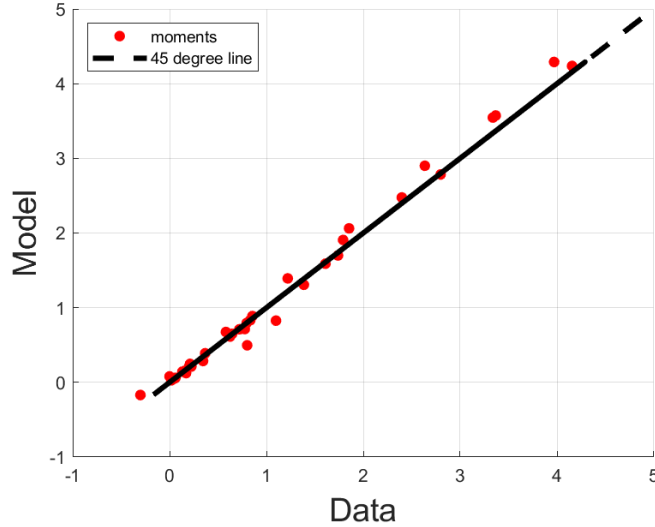
6. Evaluate the distance function:

$$\mathcal{L}(\vartheta^0) = \bar{d}(\vartheta^0)'W\bar{d}(\vartheta^0)$$

7. Update guesses and iterate to minimize the distance function.

We follow a genetic algorithm to update the vector of guesses. Figure 26 displays the estimation fit by scattering model-generated statistics against their data counterpart. At the obtained minimum, the log deviation between empirical and simulated moments is 0.086.

Figure 26: Estimation fit



C.2.4 Standard errors

To obtain estimates standard errors, we follows Chernozhukov and Hong (2003) methodology. This procedure consists of simulating a chain of parameters that has a quasi-posterior density equal to

$$f(\vartheta) = \frac{e^{\mathcal{L}(\vartheta)}p(\vartheta)}{\int e^{\mathcal{L}(\vartheta)}p(\vartheta)d\vartheta}$$

where $\mathcal{L}(\vartheta)$ is the objective function while $p(\vartheta)$ denotes a prior distribution. Standard errors are computed as the standard deviation of the sequence of elements in the converged MCMC chain. To simulate a chain that converges to the quasi posterior, we use the Metropolis–Hastings algorithm. This algorithm generates a chain of parameters $\vartheta^0, \vartheta^1, \vartheta^2, \dots$ as follows. First, we choose a starting value ϑ^j . Next, we impose the proposal density to be a uniform and we extract a new guess ϑ^p from it. Finally, we update from ϑ^{j+1} from ϑ^j for $j = 1, 2, \dots$, using the following rule:

$$\vartheta^{j+1} = \begin{cases} \vartheta^p & \text{with probability } \min\left\{1, \frac{e^{\mathcal{L}(\vartheta^p)}}{e^{\mathcal{L}(\vartheta^j)}}\right\} \\ \vartheta^j & \text{with probability } 1 - \min\left\{1, \frac{e^{\mathcal{L}(\vartheta^p)}}{e^{\mathcal{L}(\vartheta^j)}}\right\} \end{cases}$$

where we use a uniform also as our prior distribution. The quasi-posterior density is obtained using a chain of 3000 model evaluations after discarding the first 10000. As in Lise et al. (2016), standard errors are computed as standard deviation of the MC chains.

C.2.5 Identification sensitivity

In this section we explore how sensitive are parameters estimates to changes in the value of η , which is calibrated outside the model. To do this, we use we use the sensitivity measure proposed by Jørgensen (2020), defined as follows:

$$\mathbf{S} = \Lambda \mathbf{D}$$

where \mathbf{D} is the Jacobian of the moment condition w.r.t. to η , whose (i)th entry is equal to

$$\mathbf{D}_{(i)} = \left. \frac{\partial \bar{d}_{(i)}(\vartheta)}{\partial \eta} \right|_{\eta = \hat{\eta}}$$

while Λ is the measure of parameters' sensitivity to moment conditions constructed by Andrews et al. (2017):

$$\Lambda = -(\mathbf{J}'\Sigma\mathbf{J})^{-1}\mathbf{J}'$$

where Σ is a $g \times g$ weighting matrix used to construct the distance function (in our case the identity matrix), while the matrix \mathbf{J} is a $g \times J$ Jacobian of the moment conditions w.r.t. the estimated parameters, whose (i,j)th entry is equal to:

$$\mathbf{J}_{(i,j)} = \left. \frac{\partial \bar{d}_{(i)}(\vartheta)}{\partial \vartheta_{(j)}} \right|_{\vartheta_{(j)} = \hat{\vartheta}_{(j)}}$$

Table 13 reports the elasticity of each estimated parameter j to the value of η , which is equal to

$$\mathbf{S}_{(j)} \frac{\hat{\eta}}{\hat{\vartheta}_{(j)}}$$

Table 13: Estimate sensitivity to matching elasticity

Parameters	Description	Elasticity to changes in η
c_e	Entry cost	0.016
σ_z	Firm-productivity dispersion	0.060
$\underline{\xi}$	Training cost (lower bound)	0.061
$\bar{\xi}$	Training cost (upper bound)	-0.010
λ_1	Hiring costs, convexity	0.195
M	Measure of potential entrants	0.186
β	Bargaining power	-0.225
σ_a	Initial human capital dispersion	0.004
p^e	Experience jump	0.778
p^t	Training jump	-4.916
p^d	Depreciation jump	-1.391
b	Home production	0.004
δ_s	Match separation	-40.45

This final measure can be interpreted as the percent bias in parameter j estimate for a one percent perturbation of η .

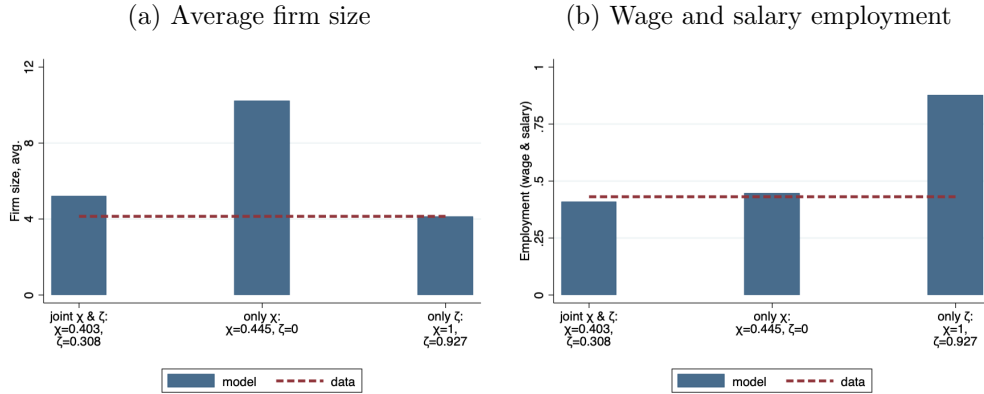
D Identification of ζ and χ

To illustrate how we identify ζ and χ , consider Indonesia, a country with one-tenth of UK's GDP per capita (4,095 USD vs. 39,000 USD). The average firm size in Indonesia is just 4.14 employees, roughly 12 employees less than the average firm in the UK.¹⁵ The share of wage and salary earners, based on our calculation from Section 2, is 43.11%, around 35 percentage points lower than the UK. To match the Indonesian firm size and the wage employment, the model requires a value of ζ around 0.308 (in contrast to 0 for the UK), while χ is about 0.403 (in contrast to 1 for the UK).

The identification of ζ and χ relies on the differential effect these two parameters have on average firm size and employment. Compare our approach with two alternative calibration strategies where these parameters are pinned down one at a time to match only one target, the average firm size for ζ , and the employment rate for χ , while still re-calibrating b . The calibrated values of ζ and χ would be 0.927 and 0.445, respectively. Targeted and non-targeted moments for these three approaches are shown in Figure 27. Panel (a) reports the

¹⁵Data for average firm size across countries is from Bento and Restuccia (2017).

Figure 27: Identification of counterfactual parameters



Notes: Blue bars in left and right panels refer to the simulated average firm size and wage and salary employment for different calibration strategies of the counterfactual parameters ζ and χ . The red lines refer to the empirical targets.

average firm size, panel (b) reports wage and salary employment. The dashed lines refer to data. The bars refer to the moments generated by the model in each calibration strategies.

Consider only finding a value for χ . Relative to joint calibration, for the same drop in employment rate, from 77.58 to 44.44% (second bar in panel b), higher labor market frictions alone generate a much smaller drop in average firm size, from 16.4 to 10.2 employees (second bar in panel a). Suppose, instead, we only look for ζ . Now, for the same drop in average firm size from 16.1 to 4.1 employees (third bar in panel a), an increase in correlated distortions alone would increase wage and salary employment to 87.67%, instead of reducing it from 77.58 to 40.75% (third bar in panel b). Correlated distortions generate many small firms that would hire workers, if the matching process in Indonesia was as efficient as it is in the UK.

E Alternative counterfactual

In this section we report the outcomes of an alternative counterfactual exercise where we re-estimate a larger set of parameters to match empirical targets for Indonesia.

In particular, together with matching efficiency, χ and correlated distortion, ζ , we re-estimate: i) the aggregate productivity shifter, κ to match the GDP per capita; ii) the experience probability jump, p^e to match the life-cycle wage growth after 20 years; iii) the training probability jump, p^t to match the share of firms providing training. Moreover, we adjust the values of home production, b , boundaries for training costs, $\underline{\xi}$ and $\bar{\xi}$, and entry

Table 14: Alternative counterfactual

	UK Baseline	Indonesia Counterfactual Joint (χ, ζ)	Indonesia Full	Indonesia Data
	(1)	(2)	(3)	(4)
Matching frictions: χ	1	0.403	0.382	-
Distortion correlation: ζ	0	0.308	0.252	-
Aggregate Productivity shifter: κ	1	1	0.938	-
Experience jump: p^e	0.223	0.223	0.205	-
Training jump: p^e	0.028	0.028	0.003	-
Home production: b	20.94	3.505	4.020	-
Training costs (lower bound): $\underline{\xi}$	1.735	1.735	0.232	-
Training costs (upper bound): $\bar{\xi}$	26.69	26.69	2.212	-
Entry cost: c_e	39.26	39.26	3.161	-
Average firm size, $E[\ell_t]$	16.19	5.177	3.681	4.141
Employment rate	0.788	0.408	0.461	0.431
Income per capita	1	0.061	0.087	0.100
Training provision, overall %	65.02	6.210	7.006	6.291
Earnings growth, $E[\log(w_{25}/w_1)]$	0.801	0.280	0.222	0.216
Mean-median ratio, $E[w_{it}]/p^{50}[w_{it}]$	1.207	1.805	1.772	1.687
GINI	0.416	0.506	0.503	0.502

Notes: The entries in columns (1) show the outcomes for the benchmark (UK). The entries in columns (2) and (3) show the outcome for the counterfactual (Indonesia) after changing in matching efficiency and correlated distortions together (column 2). and after changing in a larger set of parameters, including matching efficiency, correlated distortions, aggregate productivity shifter, learning probability, training probability, training costs and entry cost (column 3).

cost c_e such that they are equal to the value estimated in the baseline economy, as % of average earnings.

F Alternative mechanisms

In this section, we explore the effects of two alternative factors that affect the decisions of the firms and the workers in the model. Again, we take the UK as a benchmark economy and compare it against Indonesia.

Worker separation. Donovan et al. (2020) show that worker separation rates decline with GDP per capita. To study this channel, we conduct an alternative experiment where we calibrate δ_s to match an average job tenure of 3.3 year in Indonesia (Marinescu and

Table 15: Alternative mechanisms

	UK		Indonesia		Indonesia
	Baseline	Counterfactual			Data
		Joint (χ, ζ)	Joint (δ_s, ζ)	Joint (χ, δ_f)	
	(1)	(2)	(3)	(4)	(5)
Matching frictions: χ	1	0.403	1	0.501	-
Distortion correlation: ζ	0	0.308	0.659	0	-
Separation rate: δ_s , %	1.235	1.235	5.179	1.235	-
Firm exit rate: δ_f , %	2.526	2.526	2.526	3.253	-
Home production: b	20.94	3.505	1.400	11.84	-
Average firm size, $E[\ell_t]$	16.19	5.177	4.421	10.11	4.141
Employment rate	0.788	0.408	0.666	0.452	0.431
Income per capita	1	0.061	0.051	0.232	0.100
Training provision, overall %	65.02	6.210	0	27.59	6.291
Earnings growth, $E[\log(w_{25}/w_1)]$	0.801	0.280	0.614	0.327	0.216
Mean-median ratio, $E[w_{it}]/p^{50}[w_{it}]$	1.207	1.805	1.327	1.835	1.687
GINI	0.416	0.506	0.427	0.513	0.502

Notes: The entries in columns (1) show the outcomes for the benchmark (UK). The entries in columns (2) to (4) show the outcome for the counterfactual (Indonesia) across different experiments: changes in matching efficiency and correlated distortions together (column 2), changes in worker separation and correlated distortions (column 3), changes in firm exit rate and matching efficiency (column 4). The entries in column (5) refer to the empirical outcomes for Indonesia.

Triyana 2016, Table 1), and re-calibrate ζ to match the average firm size. Recall that in the benchmark (the UK), δ_s was estimated to be 1.235% per month, and average job tenure was 6.2 years. All the other parameters are kept at their UK level, except for home production b which is adjusted to be the same share of the average wage. The outcomes are in column 3 of Table 15. With a higher separation rate, employment rate and GDP per capita decline significantly. However, while an increase in separation rate can qualitatively account for inequality patterns observed over development, quantitatively, changes in the mean-to-median ratio and the GINI are much smaller compared to the counterfactual with changes in χ and ζ (column 2).

Firm turnover. Next, we focus on differences in firm exit rate, δ_f . Bartelsman et al. (2009) document that firm turnover declines with GDP per capita. We study this channel by calibrating δ_f so that the yearly firm exit rate in Indonesia equal to 13.66% (Hallward-Driemeier and Rijkers 2013, Table 1, for 2001) and re-calibrating χ to match the wage and salary employment rate in Indonesia (and b is again adjusted). Recall that the firm exit

rate in the UK was 10.5% per year. This experiment does generate the observed patterns in inequality. It is not able to generate, however, the very small firm size observed in the data (column 4).

G Baseline estimation without OTJ training

To study the role of OTJ training along development, we re-estimate a version of the model without OTJ training.

Table 16: Parameters estimates (without OTJ training)

Parameters	Description	Value
c_e	Entry cost	44.75
σ_z	Firm-productivity dispersion	1.221
λ_1	Hiring costs, convexity	2.532
M_e	Measure of potential entrants	0.031
β	Bargaining power	0.427
σ_a	Initial human capital dispersion	1.035
p^e	Experience jump	0.209
p^d	Depreciation jump	0.430
b	Home production	22.26
δ_s	Match separation, %	1.226

In this version of the model, human capital accumulation when employed only happens through on-the-job learning. The outcomes of this experiment are discussed in Section 6.2. Without the parameters and targets pertaining to OJT, 10 parameters are estimated to match 30 moments. Estimates are reported in Table 16. Targeted moments are reported in Tables 17.

H The role of OTJ training

In this section, we present an alternative experiment that illustrates the role of training. We impose training decisions from the counterfactual economy (Indonesia) on UK firms. If a match between a type- a worker and type- (z, ξ) implies training (or no training) in the counterfactual economy, the pair behaves the same way in the baseline economy, even if such behavior is not optimal for the match.

Given these imposed decision rules, firms still make hiring decisions to maximize their profits facing the benchmark values of ζ and χ . Hence, this experiment isolates the impact

Table 17: Targeted Moments (without OTJ training)

	Data	Model		Data	Model
<i>Firm-level moments</i>			<i>Earnings distribution</i>		
Average firm size, $E(\ell_t)$	16.42	16.18	Average earnings at entry, $E[\log(w_1/\bar{w})]$	-0.518	-0.479
Average log-firm size, $E(\log \ell_t)$	1.739	1.789	Average earnings after 20 y.o., $E[\log(w_{20}/\bar{w})]$	0.107	0.108
Dispersion log-firm size, $\text{std}(\log \ell_t)$	1.220	1.371	Average earnings at re-emp, $E[\log(w_R/\bar{w})]$	-0.301	-0.163
			Earnings dispersion at entry, $\text{sd}[\log w_1]$	0.582	0.571
<i>Firm size distribution</i>			Earnings dispersion after 20 y.o., $\text{sd}[\log w_{20}]$	0.796	0.738
1-9 employees	72.12	69.11	Earnings dispersion at re-emp, $\text{sd}[\log w_R]$	0.834	0.735
10-24 employees	15.95	15.68			
25-49 employees	6.12	7.310	<i>Job tenure return</i>		
50-99 employees	3.21	4.621	tenure<3 months	1	1
100-249 employees	1.73	3.080	tenure \in [3,12) months	1.055	1.053
250+ employees	0.88	0.210	tenure \in [12,24) months	1.132	1.136
			tenure \geq 24 months	1.368	1.369
<i>Firm size percentiles</i>			<i>Aggregate moments</i>		
10th percentile	1	1.181			
25th percentile	3	2.689	Job duration	6.700	6.217
40th percentile	4	3.984	Employment rate	0.776	0.764
50th percentile	5	5.098			
60th percentile	6	7.111			
75th percentile	11	13.60			
90th percentile	29	39.89			
95th percentile	53	72.54			
99th percentile	202	175.4			

Notes: The entries show the full set of firm-level and worker-level empirical moments used in the estimation of the model without OTJ training, together with their simulated counterparts.

of correlated distortions and higher labor market frictions on training decisions. Column 1 in Table 18 shows the results for the benchmark economy, and column 3 shows the outcomes for Indonesia. In column 2, the results for the UK under Indonesia's training decisions are reported.

The income per capita and average earnings are lower by about 10% with the training decision rules fixed at the counterfactual economy. Workers now receive much less training, which lowers the human capital accumulation. With fixed training policies, changes in earnings inequality induced by correlated distortions and search frictions are muted. Although firms in the UK are forced to take constrained training decisions, they do not face size-dependent distortions or higher search frictions as the firms do in Indonesia. A comparison between columns 1 (UK) and 3 (Indonesia) versus column 1 (UK) and 2 (UK with Indonesia's training decision) suggests that the endogenous training decisions account for about 11% of changes in the mean-to-median ratio and the Gini coefficient, 17% of the change in p90-p50 earnings ratio and around 25% of the change in p50-p10 earnings ratio.

Table 18: The Benchmark with Counterfactual Training Policies

	Baseline		Counterfactual	Explained
	with	with		
	baseline	counterfactual		
	training	training		
	(1)	(2)	(3)	(4)
Matching frictions: χ	1	0.403	0.403	-
Distortion correlation: ζ	0	0	0.308	-
Home production: b	20.94	20.94	3.505	-
<i>Aggregates</i>				
Employment rate	0.788	0.764	0.408	6.312%
Average earnings	1	0.932	0.124	7.729%
Income per capita	1	0.903	0.061	10.33%
<i>Earnings profile over experience/tenure</i>				
Earnings growth, $E[\log(w_{25}/\bar{w}_1)]$	0.801	0.760	0.280	7.994%
<i>Earnings inequality</i>				
Mean-median ratio, $E[w_{it}]/p^{50}[w_{it}]$	1.207	1.269	1.805	10.37%
GINI	0.416	0.426	0.506	11.11%
90-50 pct. ratio, $p^{90}[w_{it}]/p^{50}[w_{it}]$	2.551	2.876	4.462	17.06%
50-10 pct. ratio, $p^{50}[w_{it}]/p^{10}[w_{it}]$	5.262	4.610	2.729	25.74%

Notes: The entries in columns (1) and (3) show the outcomes for the benchmark (UK) and the counterfactual (Indonesia). The entries in column (2) show outcomes when on-the-job training decisions from the counterfactual economy are imposed on the benchmark. The last column shows the ratio of differences between (1) and (2) compared with (1) and (3).

I A Re-training program

We model re-training by assuming non-employed workers have the option of either searching for jobs or to participate to a re-training a program which increases human capital with probability p^t . The value of being non-employed at the beginning of period for a worker with ability a is equal to

$$J^u(a) = \max\{J^r(a), J^s(a)\},$$

where $J^r(a)$ is the value of re-training during non-employment, given by

$$J^r(a) = p^t J^{u,h}(a+1) + (1-p^t) J^{u,h}(a),$$

while $J^s(a)$ is the value of searching for a job, which is unchanged and given by

$$J^s(a) = (1 - \phi_w)[p^d J^{u,h}(a - 1) + (1 - p^d)J^{u,h}(a)] \\ + \phi_w \int_{z \in \mathcal{Z}} \int_{\xi \in \mathcal{E}} [\mathbf{1}^h(z, \xi, a)J^{e,h}(z, \xi, a) + (1 - \mathbf{1}^h(z, \xi, a))J^{u,h}(a)]\psi_v(z, \xi)d\xi dz.$$

A solution to this problem is an indicator function for re-training, $\mathbf{1}^r(a)$ defined as:

$$\mathbf{1}^r(a) = \begin{cases} 1 & \text{if } J^r(a) \geq J^s(a) \\ 0 & \text{otherwise} \end{cases}$$

All the other features of the model are kept the same as in the benchmark.

References

- Altonji, J. G. and Segal, L. M. (1996). Small-sample bias in gmm estimation of covariance structures. *Journal of Business & Economic Statistics*, 14(3):353–366.
- Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2009). Measuring and analyzing cross-country differences in firm dynamics. *Producer dynamics: New evidence from micro data*, 15-76.
- Bento, P. and Restuccia, D. (2017). Misallocation, establishment size, and productivity. *American Economic Journal: Macroeconomics*, 9(3):267–303.
- Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.
- Cosar, A. K., Guner, N., and Tybout, J. (2016). Firm dynamics, job turnover, and wage distributions in an open economy. *American Economic Review*, 106(3):625–63.
- Donovan, K., Lu, W. J., and Schoellman, T. (2020). Labor market dynamics and development. *Yale University Economic Growth Center Discussion Paper*, 1071.
- Guner, N. and Ruggieri, A. (2022). Misallocation and Inequality. *CEPR Discussion Paper*, 17113.
- Hallward-Driemeier, M., and Rijkers, B. (2013). Do crises catalyze creative destruction? Firm-level evidence from Indonesia. *Review of Economics and Statistics*, 95(5), 1788-1810.
- Lise, J., Meghir, C., and Robin, J.M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, 19:63–87.
- Marinescu, I., and Triyana, M. (2016). The sources of wage growth in a developing country. *IZA Journal of Labor & Development*, 5(1), 1-43.