I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15174

# Misallocation and Inequality

Nezih Guner
Alessandro Ruggieri

MARCH 2022

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 15174

# Misallocation and Inequality

**Nezih Guner**
*UAB, ICREA, BSE and IZA*

**Alessandro Ruggieri**
*University of Nottingham*

MARCH 2022



**IZA – Institute of Labor Economics**

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

# ABSTRACT

# Misallocation and Inequality*

For a large set of countries, we document how the 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 and matching frictions, and skill accumulation of workers through learning-by-doing and on-the-job 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.

**JEL Classification:** E23, E24, J24, O11

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

**Corresponding author:**
Nezih Guner
Universitat Autonoma de Barcelona
Bellaterra Campus
Cerdanyola del Valles 08193
Spain
E-mail: gunermail@gmail.com

* We would like to thank Jaime Arellano-Bover, Niklas Engbom, David Lagakos, Todd Schoellman, and seminar and workshop participants at 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. The usual disclaimers apply.

# 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 also 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. Other workers will be less fortunate and work for firms with lower productivity. The higher a worker's human capital is, the higher her chances of employment in a high productivity firm. 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 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-dependent or correlated distortions make high and low-productivity firms more similar, compressing the overall earnings distribution. While the link between distortions and inequality is intuitive, it has not been studied so far.

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 firms less willing to wait for the right workers, 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. Finally, 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.

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). Flinn et al. (2017) study investment in general and firm-specific human capital within an equilibrium search and matching model.

In the model, countries differ with respect to correlated distortions and labor market frictions. We calibrate the correlated distortions to match the average firm size and the

[1] Beyond 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.

matching function elasticity to match the share of wage and salary workers for different countries. Lower distortions or frictions alone would increase the firm size, employment, and GDP per capita in the model economy. But to match the observed cross-country differences in firm size and employment, we need both. Distortions help us generate observations for poor countries, i.e., lower frictions alone would not be able to match the small firm size and low levels of wage and salary employment in the data. Similarly, frictions help us generate observations for rich countries, i.e., lower distortions alone would not match the large firm size and high levels of wage and salary earners in the data.

Then we zoom into these economies and calculate inequality measures, firm size distribution, training outcomes, and life-cycle earnings profiles. 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, both distortions and friction can generate this pattern. But again, both features are needed to match the observed range, with higher distortions pushing p50-p10 down in poor countries and lower frictions pushing it up in rich ones. In the data, the p90-p50 ratio falls by GDP per capita. In the model, this is generated by lower matching frictions in richer countries.

Furthermore, the model also fits a large set of cross-country facts on firm-size distributions, life-cycle earning profiles, and training. 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 significantly 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 turns out to be a higher p50-p10 and lower p90-p50 in richer countries since workers in the middle of the skill distribution benefit the most.

The 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 a high-skilled worker 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 between 10% and 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 significantly 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, the mean earnings of managers tend to 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 for 57 countries from 1974 to 2016, for a total of 502 country-year pairs. 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 (which we refer to as earnings below), which accounts for any payment received as an employee, including extra pay, tips, commissions, bonuses, piece-rate payments, and occasional earnings. For each country-year, wage and salary earners (which we refer to 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.

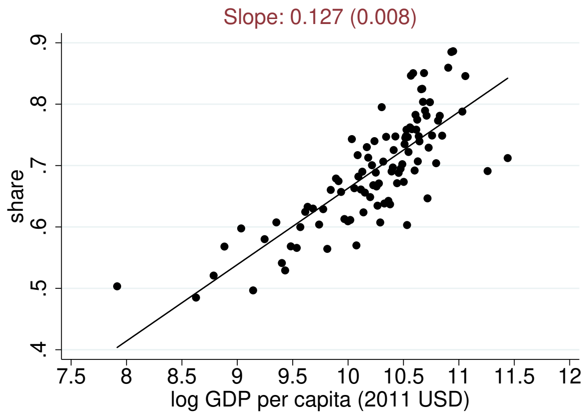
Figure 1 shows how the share of earners (panel a) and the average earnings (panel b) change with (log) GDP per capita across countries. Each dot corresponds to the average share for countries in a specific bin of GDP per capita. We construct 100 bins corresponding to the percentiles of the GDP-per-capita distribution. Outcomes for the y-axis, the share of earners, are reported as residuals from a regression with year-fixed effects. As we move from poorer to richer countries, workers become significantly more likely to work as an employee and report positive earnings. The share ranges from around 50 for the poorest countries in the sample to almost 90 percent for the richer ones. Moreover, average log earnings exhibit a steep and significant gradient as they increase almost one-to-one with log GDP per capita.

Figure 2 documents earnings inequality. Each dot again corresponds to the average values of the dependent variable (different inequality measures) for countries in a 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

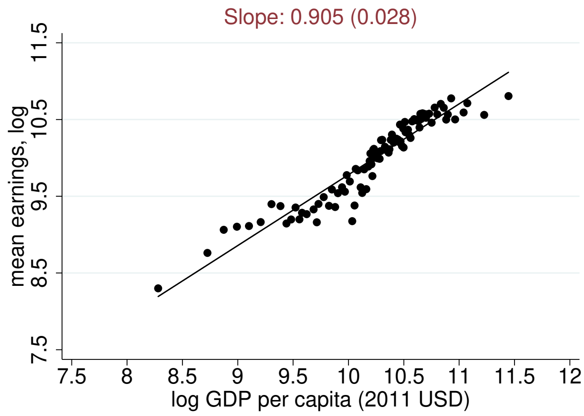
[2]Details on these three datasets, sample restriction, and variable definitions are provided in Appendix A.1.

Figure 1: Wage and salary earners and earnings across countries

(a) Earners




(b) Average earnings



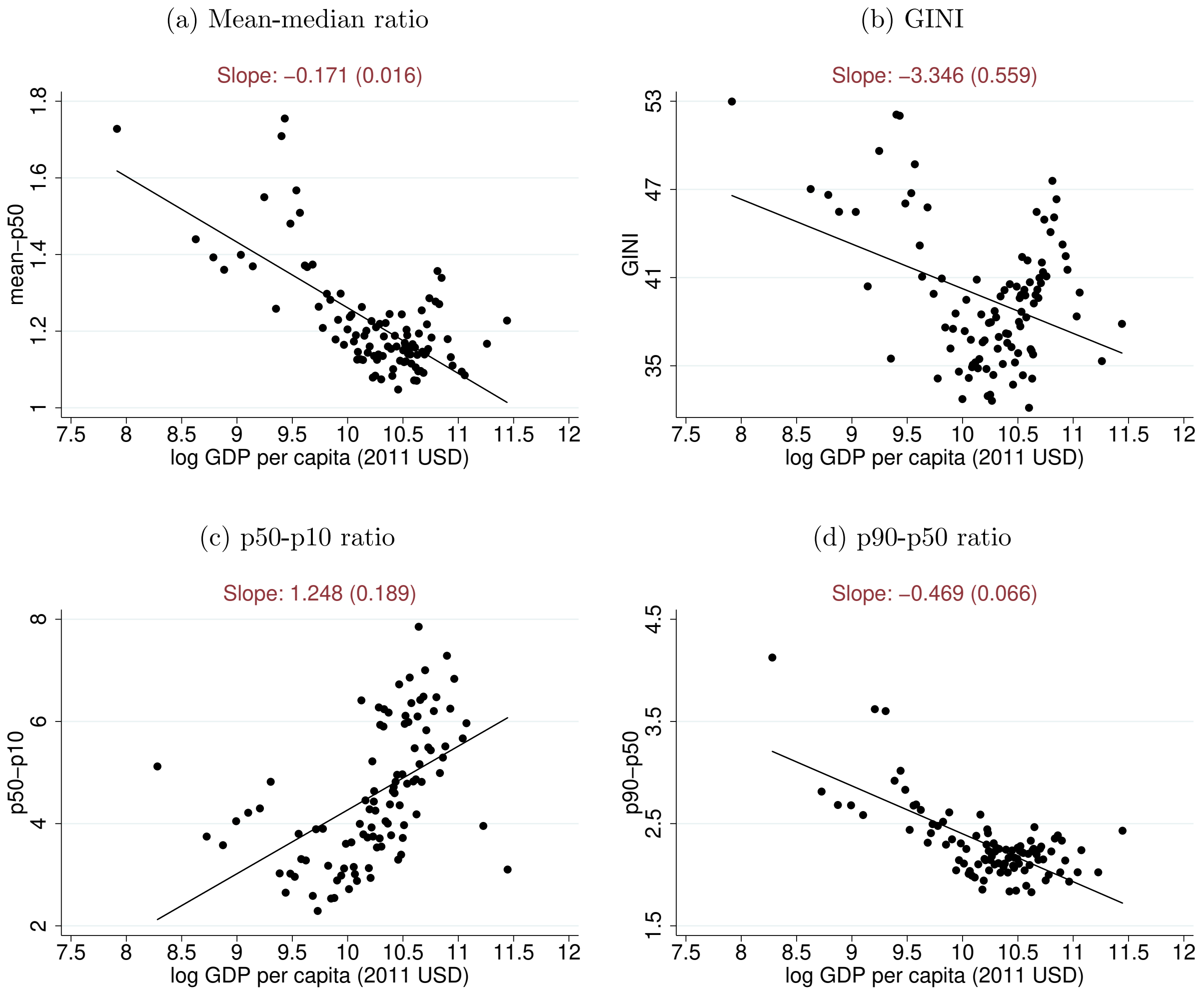

Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year-fixed effects. In red we report the estimated slope in the regression. Robust standard errors are in parenthesis. Source: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

the poorest countries in the sample to about 1.1 for the richest ones. Hence, as countries get richer and the mean of the earnings distribution increases, 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 around 50% to 35%.

The decline in these two measures of inequality, however, 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 holds in the upper tail, as shown in panel (d). Labor income for workers in the 90th percentiles does not grow as fast as the median, and the 90-to-50 ratio declines as countries get richer.

In Appendix A.1, we show that these findings are robust. They hold when we restrict the sample to workers employed in specific sectors (non-agriculture or industry) or workers holding college or non-college degrees. We also find the same patterns when the sample is restricted to include only males, only household heads, or only workers in prime working ages (25 to 55 y.o.). Finally, we zoom into different percentiles of the earnings distribution and 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.

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. Robust standard errors are in parenthesis. Source: IPUMS, EU-SILC, LIS and author's calculations.

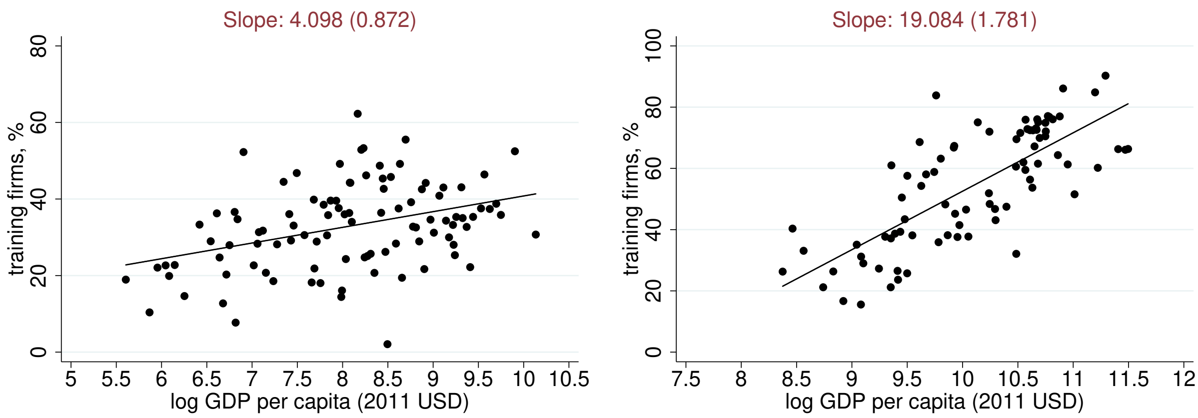
## 2.2 On-the-Job Training

This section complements our cross-country evidence on earnings inequality with facts on on-the-job training provision. To this purpose, we use the World Bank Enterprise Survey (WB-ES, henceforth) dataset. WB-ES, which is representative of the population of establishments with at least five employees, collects data from a wide range of developing countries through face-to-face surveys. It contains several establishment-level outcomes for over 100 countries for at least one year since 2002, and, in particular, information on firms' demographics (industry, age, and the number of employees) and firms' training provision.

We complement WB-ES with information from the Eurostat Database on Education and Training. This dataset provides information on the participation of individuals in education and training activities for a sample of 30 middle- and high-income countries. Within this database, the Continuing Vocational Training Survey (CVTS, henceforth) provides information on on-the-job training provision by firms.[3]

Figure 3: Training provision across countries

(a) World Bank Enterprise Survey (b) Continuing Vocational Training Survey




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.

Figure 3 shows how the percentage of firms offering job training to their employees vary 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 CV-TS data is even higher, 0.745, and 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.2 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 job 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 job

[3]Further details on these two datasets are provided in Appendix A.2.

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 sample. In Appendix A.2 we also show that conditional on investing in job training, larger firms train a larger share of their workforce.

Table 1: Job training across firm size

| | Training firms, % | | | | | | |
|---|---|---|---|---|---|---|---|
| | WB-ES | | | | | CVTS | |
| | LAC | ME+AFR | ASIA | others | | EU15 | non-EU15 |
| 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. Source: World-Bank Enterprise Survey (WB-ES) and Eurostat Education and Training Dataset (CVTS).

### 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 by the share of workers receiving training within a given firm, increases.

These facts complement known cross-country patterns on 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). Each worker enters the economy with a given level of human capital (skill or ability). Each period workers can be employed or non-employed.[4] Labor market frictions are represented by a matching function that maps non-employed workers and opens vacancies into potential matches. If a match between a worker and a firm is formed, workers' skills can grow due to learning-by-doing and on-the-job 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 $a_0 \in \mathcal{A} = \{a_0, a_1, ..., a_A\}$. Initial skills are distributed according to an exogenous probability distribution function, $\psi_a(a)$. Upon matching with a firm, workers improve their skills through job experience (learning by doing) 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 skill depreciation of $a$ 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 output. Like

[4] Within those without a wage and salary job, we do not distinguish the unemployed from the inactive or from those who are self-employed. We will therefore refer to them as non-employed or unemployed, interchangeably.

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 $\xi$ and $\ell$ workers. Then, we can write total firm output as

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

where $g(z,i)$ is the amount produced by a match between a firm $z$ and a worker $i$ with human capital $a(i)$, given by

$$g(z,i) = za(i),$$

Re-arranging terms, we can write the production function as

$$y = g(z,\bar{a})\ell, \tag{1}$$

where $\ell$ is the number of employees, and $g(z,\bar{a})$ is the average amount of production, defined as follows:

$$g(z,\bar{a}) = z\bar{a},$$

with

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

Linearity of the aggregate production function with respect to $\ell$ 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 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 $\delta_s$, or lose all workers and exit with probability $\delta_f$.

## 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 $\tau$ depends on firm-level productivity $z$, given by

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

The parameter $\zeta$ is the elasticity of a firm′s distortion with respect to its productivity. This formulation implies that the net revenue function for a firm producing $y$ units of goods is

$$r(z,\ell,\bar{a}) = (1-\tau)y(z,\ell,\bar{a}) = (1-\tau)g(z,\bar{a})\ell = z^{1-\zeta}\bar{a}\ell$$

or, equivalently

$$r = \int_0^{\ell} r(z,i)\psi(i|z,\xi,\ell)di,$$

where $r(z,i) = z^{1-\zeta}a(i)$ denotes the net revenue generated by a firm-worker pair.

## 3.4 Frictions

The labor market is subject to search and matching frictions as in Mortensen and Pissarides (1999). To hire workers, firms need to post vacancies. To find a job, workers need to search. Search is random and costless. There is no on-the-job search. Each period, 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, following Den Haan et al. (2000), given by

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

This matching function implies that contact rates for workers and for firms are equal to

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

respectively. Workers matched with a firm earn a wage equal to $w(z,\xi,a)$, which depends on the productivity of the firm they work, the training costs faced, and the level of 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 in the industry 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*), defined below. 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

$$J^{u,h}(a) = b + \frac{(1-\delta_w)}{1+r} J^u(a). \tag{3}$$

With probability $\phi_w$ 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 $\xi$, which is endogenously determined. When a worker and firm are matched, if there is a positive surplus, employment takes places, i.e, $\mathbf{1}^h(z,\xi,a) = 1$. Otherwise, a match is not formed, and the worker stays unemployed. The function $J^{e,h}(z,\xi,a)$, which is defined below, is the value for a worker with skill $a$ of being employed at the end of the period in a firm with productivity $z$ and training costs $\xi$.

### 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 $\xi$ for a worker with skill $a$ is equal to:

$$J^e(z,\xi,a) = \mathbf{1}^h(z,\xi,a) J^{e,h}(z,\xi,a) + (1 - \mathbf{1}^h(z,\xi,a)) J^{u,h}(a), \tag{4}$$

where again $\mathbf{1}^h(z,\xi,a)$ is an indicator function for a positive surplus. If the surplus is positive, the value of employment is given by

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

where $w(z,\xi,a)$ is the wage rate. 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, defined below.

## 3.6 The Problem of the Firm

### 3.6.1 Value of an Active Match

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

$$V(z,\xi,a) = \mathbf{1}^h(z,\xi,a) V^h(z,\xi,a), \tag{5}$$

with

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

and

$$\delta = \delta_w + (1-\delta_w)\delta_s + (1-\delta_w)(1-\delta_s)\delta_f$$

A worker-firm match produces $r(z,a)$ and the worker is paid $w(z,\xi,a)$. Next period, any active job can be destroyed due to death/retirement by the worker ($\delta_w$), exogenous destruction of particular job ($\delta_s$), or exogenous destruction of the firm ($\delta_f$). If the match is destroyed due to $\delta_w$ or $\delta_s$, the firms 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 $\mathbf{1}^h(z,\xi,a)=0$.

### 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, $c(v)$, equal to

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

where $\lambda_1$ governs the degree of convexity in the cost function. Each period, the problem of a firm reads as follows:

$$\pi(z,\xi) = \max_{v(z,\xi)\geq 0} \quad v(z,\xi)\phi_f \sum_{a\in\mathcal{A}} \mathbf{1}^h(z,\xi,a)V^h(z,\xi,a)\psi^u_a(a) - c(v(z,\xi)), \tag{6}$$

where $\psi^u_a(a)$ 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 conditions for vacancies $v$ is then given by

$$c'(v(z,\xi)) = \phi_f \sum_{a\in\mathcal{A}} \mathbf{1}^h(z,\xi,a)V^h(z,\xi,a)\psi^u_a(a),$$

which implies the following vacancy posting rule for a type-$(z,\xi)$ firm

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

### 3.6.3 Entry

In general equilibrium, the measure of firms is determined by the entry decision. Before entry, a fixed measure of potential employers, $M_e$ draw a productivity $z$ and a training costs $\xi$ from two independent pdfs, $\psi_z$ and $\psi_\xi$. Upon learning their type, firms decide to enter if they can cover the entry cost $c^e$, i.e., they enter whenever

$$\Pi(z,\xi) \geq c^e. \tag{7}$$

The discounted sum of per-period aggregate profits of a firm with productivity $z$ and training costs $\xi$, $\Pi(z,\xi)$, is given by

$$\Pi(z,\xi) = \sum_{t=0}^{\infty} \left(\frac{1-\delta_f}{1+r}\right)^t \pi(z,\xi) = \frac{1+r}{r+\delta_f}\pi(z,\xi) \tag{8}$$

where $\pi(z,\xi)$ is defined in equation 6. In equilibrium with a positive measure of firms, there exists pairs of productivity and costs $(z^*,\xi^*)$ such that $\Pi(z^*,\xi^*) = c^e$, i.e. such that the marginal entrant is indifferent between entering or not. This defines a region in the space of $(z,\xi)$ for firms that decide to enter. A solution to this problem is a policy function for entry, $\mathbf{1}^e(z,\xi)$ defined as:

$$\mathbf{1}^e(z,\xi) = \begin{cases} 1 & \text{if} \quad \Pi(z,\xi) \geq c^e \\ 0 & \text{otherwise} \end{cases}$$

Note that even in absence of additional overhead cost after entry, workers' outside option in the bargaining protocol implies that some potential firms might decide stay out of the industry if the total wage bill is too high relative to the revenues generated by its matches.

## 3.7 The Surplus Function

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

$$S(z,\xi,a) = J^e(z,\xi,a) + V(z,\xi,a) - J^u(a) = \max\{0, S^h(z,\xi,a)\}, \tag{9}$$

where $S^h(z,\xi,a)$ is equal to

$$S^h(z,\xi,a) = g(z,a) + \frac{(1-\delta_w)}{1+r}(1-(1-\delta_f)(1-\delta_s))J^{u,h}(a) - J^{u,h}(a) \tag{10}$$
$$+ \frac{(1-\delta_w)}{1+r}(1-\delta_f)(1-\delta_s)\left[-\mathbf{1}^t(z,\xi,a)\xi + (1-p^h(z,\xi,a))M(z,\xi,a) + p^h(z,\xi,a)M(z,\xi,a+1)\right].$$

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.,

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

We report the full derivation of the surplus function 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.

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

where $S^h(z, \xi, a)$ is defined in equation (10). The amount of new hires for a firm-$(z, \xi)$ are then given by $v(z, \xi)\phi_f \sum_{a \in \mathcal{A}} \mathbf{1}^h(z, \xi, a)\psi_a^u(a)$.

### 3.8 Training

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

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

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

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

### 3.9 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,

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

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 $\beta$ 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.

Table 2: Parameters directly calibrated

| Parameters | Description | Value | Sources/Targets |
|---|---|---|---|
| $\zeta$ | Correlated distortion | 0 | Assumption |
| $r$ | Interest rate | 0.0033 | Annual return of 4% |
| $\delta_w$ | Workers retirement | 0.0099 | Life-span of 40 years, ages 22-62 |
| $\delta_f$ | Firm exit | 0.0253 | Annual exit rate of 10.50% (ONS) |
| $\eta$ | 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.

## 4.1 Parameters Set Without Solving the Model

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.

We take the UK as a distortion-free economy and set $\zeta$ to zero, which should be interpreted as a normalization against which the counterfactual economies will be compared. The interest rate, $r$, is 0.0033 to match an annual return of 4%. Workers stay in the labor force on average for forty years, corresponding to ages 22 to 62, so $\delta_w$ is 0.0099. The firm destruction rate $\delta_f$ is chosen to match an annual firm exit rate of 10.5%.[5] Finally, the elasticity of matching function, $\eta$, is estimated with generalized method of moments (GMM), by minimizing the distance between new matches formed according to model's matching function (given data on vacancies and unemployed workers) and the number of new hires in the data. Details of data and estimation are reported in Appendix C.1. The estimated value for $\eta$ is 0.5416, with a standard error 0.0134.

## 4.2 Estimated Parameters

The initial human capital of workers and productivity of firms are draws from mean-zero log-normal distributions, i.e., $a \sim \log\mathcal{N}(0, \sigma_a)$, 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 $\xi \sim \mathcal{U}(\underline{\xi}, \overline{\xi})$, with $\underline{\xi}, \overline{\xi} > 0$. Given these parametric assumption, there are 13 parameters to be estimated, denoted by

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

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

Table 3: Estimated parameters

| Parameters | Description | Estimates | St.Dev. | 95% C.I. | |
|---|---|---|---|---|---|
| $c_e$ | Entry cost | 39.26 | 3.665 | 33.19 | 47.61 |
| $\sigma_z$ | Firm-productivity dispersion | 1.204 | 0.106 | 1.018 | 1.470 |
| $\underline{\xi}$ | Training cost (lower bound) | 1.735 | 0.157 | 1.455 | 2.110 |
| $\overline{\xi}$ | Training cost (upper bound) | 26.69 | 2.304 | 22.12 | 31.58 |
| $\lambda_1$ | Hiring costs, convexity | 2.525 | 0.166 | 2.063 | 2.746 |
| $M_e$ | Measure of potential entrants | 0.013 | 0.044 | 0.001 | 0.149 |
| $\beta$ | Workers' bargaining power | 0.457 | 0.042 | 0.379 | 0.550 |
| $\sigma_a$ | Initial human capital dispersion | 1.195 | 0.111 | 0.977 | 1.425 |
| $p^e$ | Experience jump | 0.223 | 0.019 | 0.184 | 0.271 |
| $p^t$ | Training jump | 0.028 | 0.003 | 0.023 | 0.035 |
| $p^d$ | Depreciation jump | 0.432 | 0.040 | 0.346 | 0.514 |
| $b$ | Home production | 20.94 | 1.824 | 17.59 | 25.06 |
| $\delta_s$ | Match separation, % | 1.235 | 0.120 | 1.007 | 1.486 |

Notes: The entries show the parameters estimated by the method of simulated moments. The standard errors with 95% confidence intervals are computed using the Monte Carlos Markov Chain approach.

These 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 $\overline{d}(\theta)$ be a vector of $g \geq \dim[\theta]$ moment conditions (deviations between model and the data), defined as

$$\overline{d}(\theta) = \overline{m} - m(\theta),$$

where $\overline{m}$ is a vector of data targets and $m(\theta)$ is their model counterparts. The vector of parameter values, $\hat{\theta}$, is given by

$$\hat{\theta} = arg \min_{\theta \in \Theta} \quad \overline{d}(\theta)' \overline{d}(\theta). \tag{15}$$

Table 3 reports parameter estimates, and their standard errors with 95% confidence intervals, computed using the Monte Carlos Markov Chain (MCMC) approach, as in Chernozhukov and Hong (2003). Confidence intervals are constructed taking the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the posterior distribution.

The estimated values imply significant heterogeneity in training costs across firms; the maximum ($\overline{\xi}$) 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 of worker and firm 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 comparable value for training cost, 11% of forgone worker-firm output. Large dispersion in training costs across firms is consistent with evidence provided by Arellano-Bover and Saltiel (2021).

Consistent with available estimates (see, among others, Cooper et al. (2007) and Merz and Yashiv (2007)), the hiring costs are highly convex, with $\lambda_1 = 2.525$. We estimate firm-productivity dispersion, $\sigma_z$, 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 for the same period of time (Black et al., 2019).

We estimate worker's 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. The estimated measure of potential firms, $M_e$, ensures non-employed workers receive a job offers with probability $\phi_w = 0.26$. Workers with a median level of skills accept 80% of offers, implying an average non-employment spell of about 3 quarters. 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 while 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%. 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

## 4.3 Model Fit

The estimation uses 40 moments, reported in Table 4, Figure 4, and Figure 5. Overall, the model does remarkably well in fitting the data with an average log-deviation of 0.086 (see Figure 27 in Appendix C.2).

### 4.3.1 Firm-Level Moments

The first column of Table 4 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. These targets are calculated from the Employer Skill Survey (ESS), which is a repeated cross-sectional firm-level bi-annual survey that measures the skills position and skills needs of UK employers. Each wave of the survey has a "Core" component, covering establishments demographics, strategy,

Table 4: Selected Targeted Moments

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

Notes: The entries on the left and right panels show a set of firm-level and worker-level empirical moments used in the method-of-simulated-moments estimation, respectively, together with their model counterparts.

recruitment and number of employees. The 2011 and 2013 waves, which are used here, also have a "Investment in training" component that contains questions on 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 employee receive training. The model does an excellent job matching these targets. The model also generates a firm size distribution that is in line with the data (Figure 4).

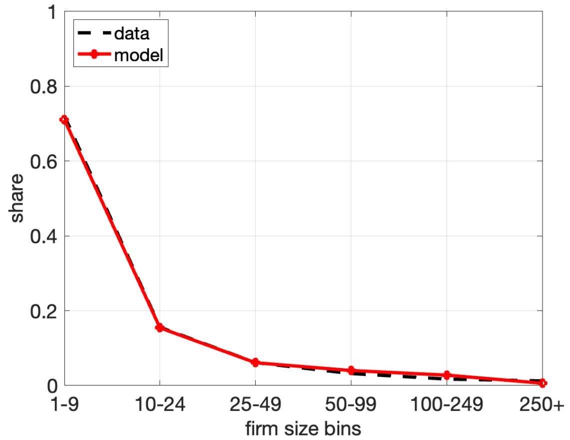
While the model does not provide with a one-to-one map between parameters and target, specific targets have more impact on specific parameter in $\theta$. In particular, the entry cost $c_e$ determines the average and percentiles of firm size distribution, while firm-productivity dispersion, $\sigma_z$, maps into dispersion in firm size. The convexity of the hiring costs, $\lambda_1$, is identified by the distribution of firms across their size. Finally, the boundaries in the support of training costs, $\underline{\xi}$ and $\overline{\xi}$, are identified by the share of firms providing training and number of workers trained within the firm for different firm size.

#### 4.3.2 Worker-Level Moments

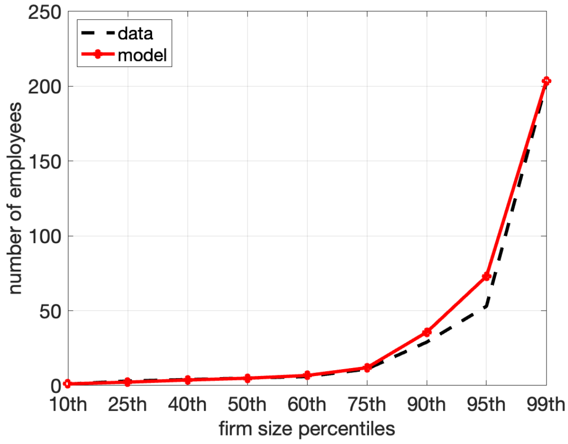
The second column of Table 4 presents a set of worker-level moments: i) earnings level and dispersion, conditional on labor market experience, and at re-employment after an unem-

Figure 4: Firm size distribution

(a) Firm size distribution



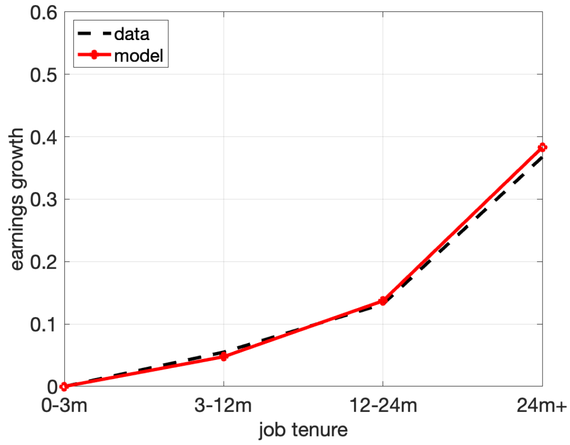

(b) Firm size percentiles



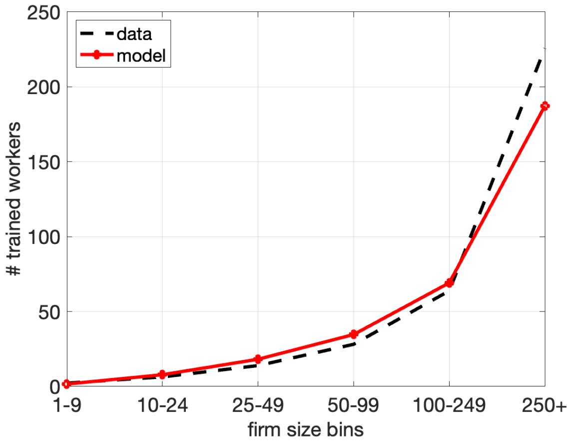

Notes: Left and right panels show selected empirical moments used in the estimation together with their model counterparts.

Figure 5: Earnings profile and training provision

(a) Earnings profiles over job tenure




(b) Trained workers over firm size




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

ployment 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).

Employment rate and average job duration are based on the ONS estimates for the period 2011-2018 and the year 2017, respectively. The other worker-level moments are based on the Five-Quarter Longitudinal (LFS), a longitudinal household survey conducted at a quarterly frequency. Each surveyed household is retained for five consecutive quarters, and a fifth of the sample is replaced each quarter. The survey records information on a wide range of demographic and labor force characteristics; among the others, the survey allows us to track workers age, employment status, job tenure, hourly pay, hours worked, and whether the surveyed worker has received OTJ training or not in the previous quarter. For the calibration, we restrict our focus only to employed individual (both women and men), aged between 22 and 62 y.o. We report descriptives of the sample in the Appendix C.2.

The model also 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 market, their earnings grow by about 10%. After an 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% higher than those of the entrants (Figure 5).

As far as the identification of different parameters is concerned, the exogenous separation rate, $\delta_s$, 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, $\phi_w$. 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. Finally, the distribution of the initial human capital, $\sigma_a$, and bargaining power for workers, $\beta$ are identified, given all other parameters, by the dispersion of earnings at entry and along workers' life cycle.

#### 4.3.3 Non-targeted Moments

Table 5 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 dispersion in log earnings and generate a value for the mean to median earnings

Table 5: Non-targeted Moments

| | Data | Model | | Data | Model |
|---|---|---|---|---|---|
| *Earnings-size regression* | | | *Earnings inequality* | | |
| <10 employees | 0 | 0 | Log-earnings dispersion, 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.

ratio remarkably close to the data, even though we abstract from the on-the-job search - see Hornstein et al. (2011) for a discussion. On the other hand, while the model correctly captures the magnitude of the dispersion in the upper tail of the earnings distribution, it generates a more left-skewed earnings distribution.

# 5 Cross-Country Facts, Revisited

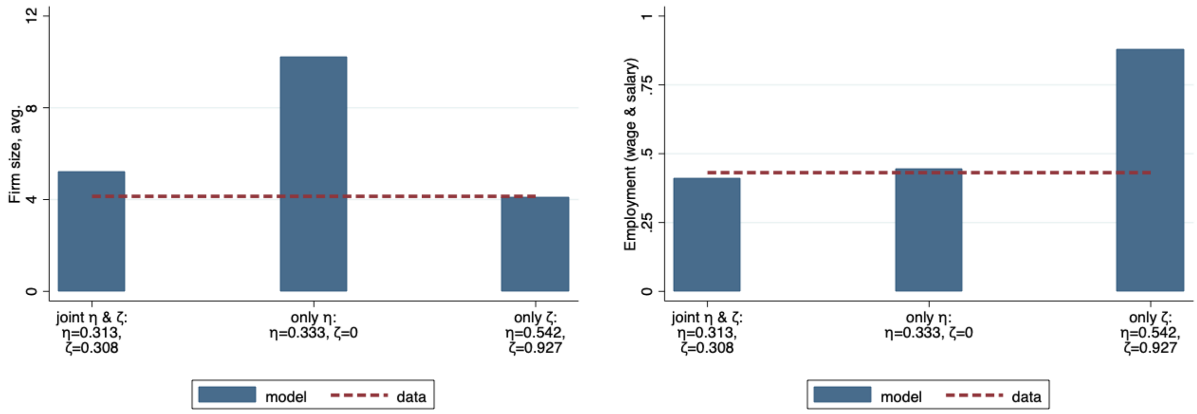
We are now ready to interpret the cross-country facts that we have documented in Section 2 through the lens of the model. To this end, we construct counterfactual economies that differ from the benchmark along two key features: size-dependent distortions, captured by the parameter $\zeta$, and matching frictions captured by the parameter, $\eta$.[6]

In the benchmark economy, $\zeta$ was set to zero, i.e., there were no size-dependent distortions, while $\eta$ was calibrated as 0.542. Keeping all other parameters fixed at their benchmark values, we search for values of $\zeta$ and $\eta$ 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 do this for 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 $b$,

[6]There can clearly be other cross-country differences. In Appendix E we consider differences in worker separation rates, $\delta_s$, and firm exit rates, $\delta_f$. They fail to generate the cross-country patterns observed in the data.

Figure 6: Identification of counterfactual parameters

(a) Average firm size (b) Wage and salary employment




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 $\eta$ and $\zeta$. The red lines refer to the empirical targets.

so that it is about 22.5% of the average earnings, the value estimated in the baseline economy.

## 5.1 Identification of $\zeta$ and $\eta$

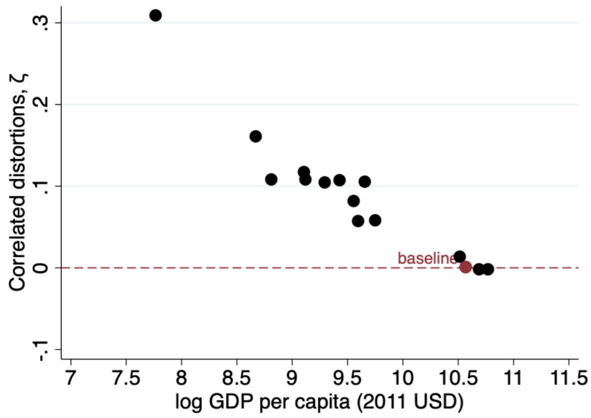
To illustrate this process, consider Indonesia, a country with one-tenth of UK's GDP per capita (4,095 USD vs. 39,000 USD). The average firm size in Indonesia is just 4.14 employees, roughly 12 employees less than the average firm in the UK.[7] The share of wage and salary earners, based on our calculation 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 $\zeta$ around 0.308 (in contrast to 0 for the UK), while $\eta$ is about 0.312 (in contrast to 0.542 for the UK).

The identification of $\zeta$ and $\eta$ 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 $\zeta$, and the employment rate for $\eta$, while still re-calibrating $b$. The calibrated values of $\zeta$ and $\eta$ would be 0.927 and 0.333, respectively. Targeted and non-targeted moments for these three approaches are shown in Figure 6. Panel (a) reports the 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.

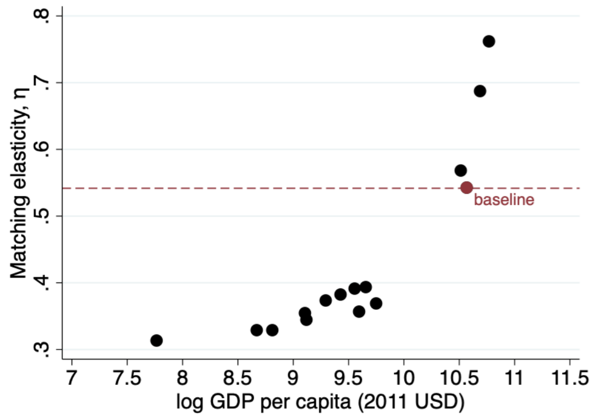
[7]Data for average firm size across countries is from Bento and Restuccia (2017).

Figure 7: Distortions and matching frictions across countries

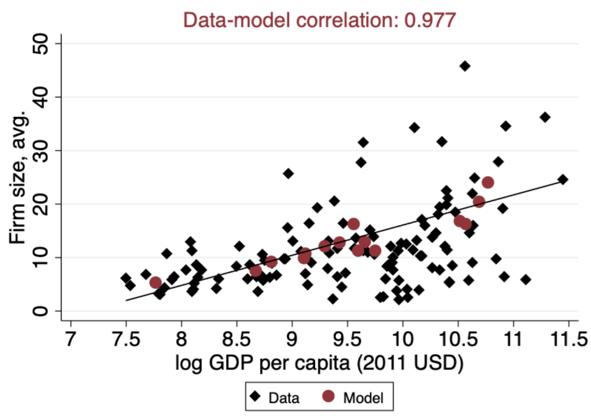
(a) Estimated counterfactual distortions



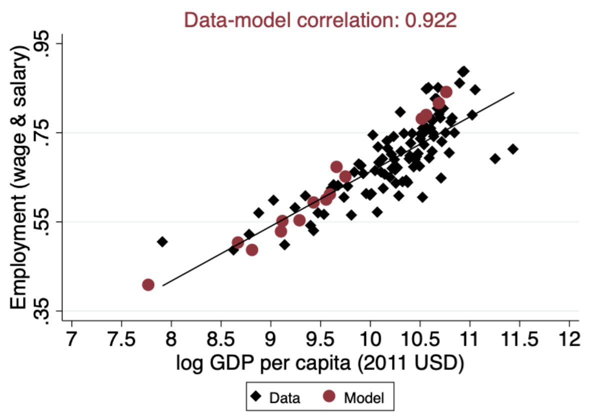

(b) Estimated counterfactual frictions




(c) Average firm size



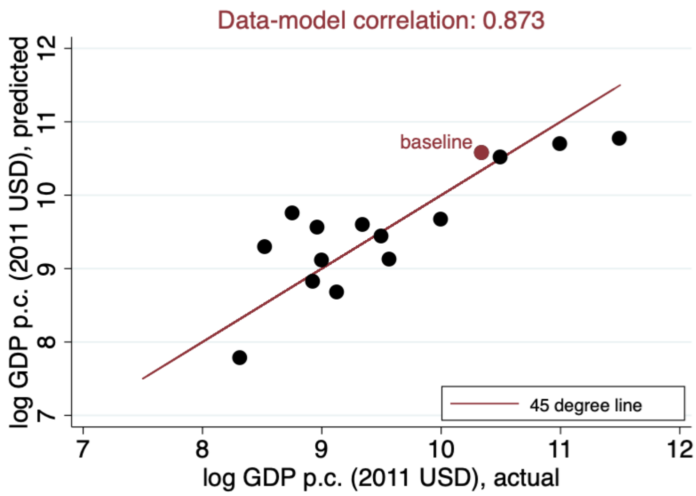

(d) Wage and salary employment




Notes: Panels (a) and (b) report the estimated counterfactual parameters, correlated distortions $\zeta$, and matching elasticity, $\eta$, 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.

Consider only finding a value for $\eta$ to generate the right employment rate in Indonesia. 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 $\zeta$ to generate a smaller average firm size. 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

Figure 8: GDP p.c. across countries: actual vs. predicted




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

was as efficient as it is in the UK.

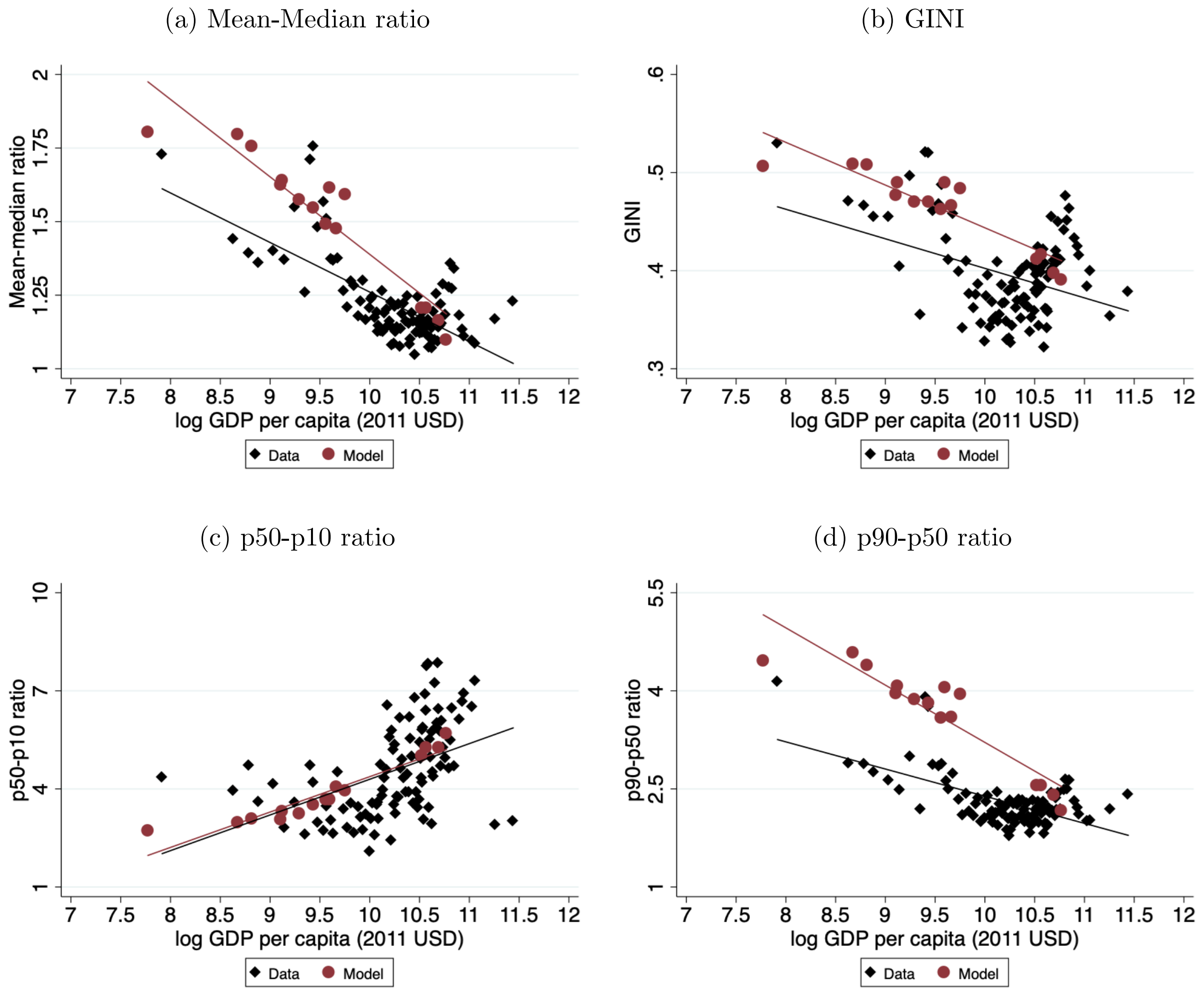
Figure 7 shows $\zeta$ and $\eta$ values for each counterfactual economy (panels a and b), together with the targeted moments, average firm size and employment (panels c and d). The range of calibrated values is quite wide, which is necessary to match the data. While $\zeta$ is zero for the UK, it is as high as 0.3 for the poorest counterfactual economy (Indonesia). As panel (a) in Figure 7 shows, $\zeta$ 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, $\eta$ is as low as 0.3 in poorest countries and increase sharply for countries that have larger wage and salary employment than the UK.

We do not directly target the GDP per capita in counterfactual economies; only the average firm size and average employment are targeted. Yet, as illustrated in Figure 8, 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.

## 5.2 Workers and Firms around the World

We now have several economies that differ in average firm size, employment rate, and GDP per capita. Next, we zoom into these economies and calculate measures of earnings inequality, firm size distribution, training, and life-cycle earnings profiles. Then we compare these outcomes, which are not targeted in the counterfactual exercises, with the data.

Figure 9: Earnings inequality across countries: Model vs. Data

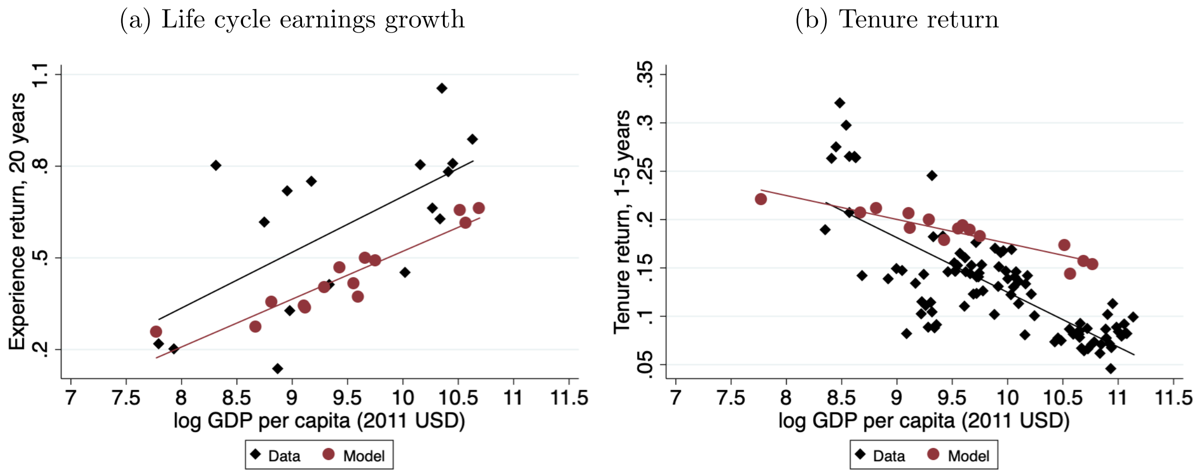



Notes: Each panel shows how a particular measure of earnings inequality changes with the log-GDP per capita across countries. The black diamonds represent the data and red dots the model.

Let's start with inequality. As Figure 9 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 ($\eta$ and $\zeta$), model is able to generate changes in earnings inequality that we observe in the data.

Next, we focus on how earnings change along the life cycle. Panel (a) in Figure 10 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 model and the data,

Figure 10: 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 blue diamonds represent the data and red dots the model.

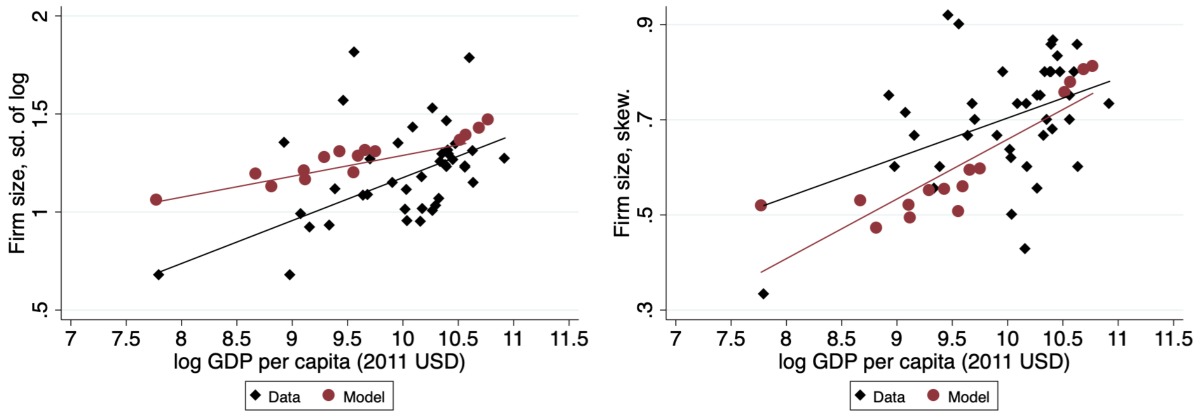
age-earnings profiles are much steeper in richer countries. Earnings grow much faster, however, during initial years of employment in poorer countries. This is documented in panel (b) of Figure 10. The data, from Donovan et al. (2020), shows the average earnings for those between 1 and 5 years of tenure relative to those with less than 6 months of tenure. We calculate the same statistics in the model economy. 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. However, 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 more likely to stay employed in richer countries, and despite their returns to tenure are lower, they experience higher earnings growth along the life cycle.

Finally, we focus on firm distribution and on-the-job training. Figure 11 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, 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 12 show that the model can also account for these cross-country patterns the data.

Figure 11: Firm distribution across countries

(a) Firm size dispersion

(b) Firm size skewness

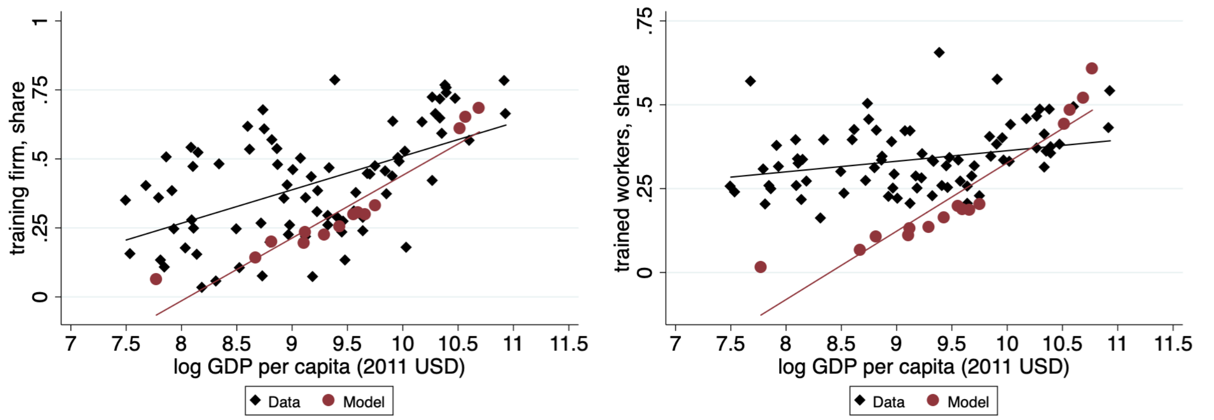



Notes: Panels (a) and (b) show how the dispersion (measured as the standard deviation of log size) and the skewness of firm-size distribution change with the log-GDP per capita. The blue diamonds represent the data and red dots the model.

Figure 12: Training provision across countries

(a) Share of training firms

(b) Share of trained workers




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 the log-GDP per capita. The blue diamonds represent the data and red dots the model.

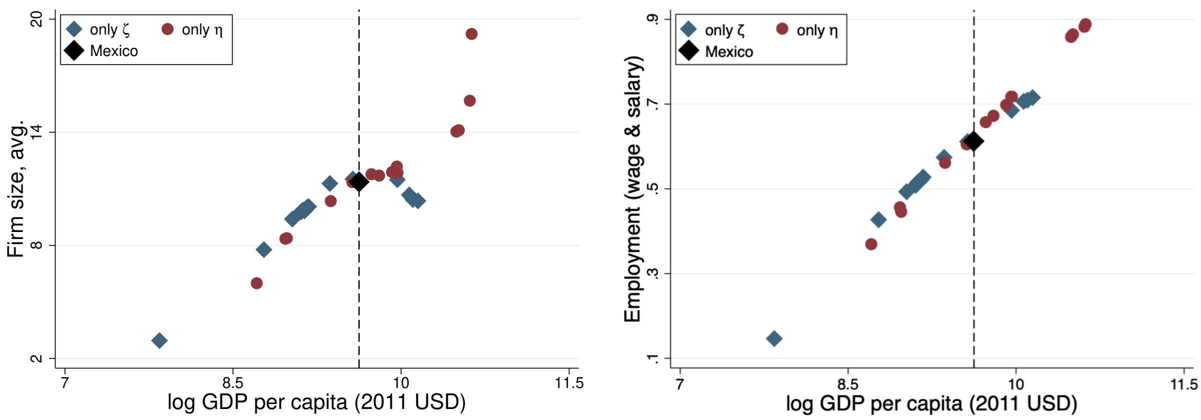
# 6 Frictions vs Distortions

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, and

4,095 USD of Indonesia). Then, for each counterfactual economy in Section 5, we only impose the calibrated values of distortions, $\eta$, *or* frictions, $\zeta$, from Figure 7. If we impose the calibrated value of $\zeta$ for a country, we keep the value of non-employment, $b$, and $\eta$ at their calibrated values for Mexico. If, instead, the calibrated value of $\eta$ is imposed, $b$ and $\zeta$ 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 $\eta$ or $\zeta$ alone.

Figure 13: Firm size and employment across countries: distortions vs. frictions

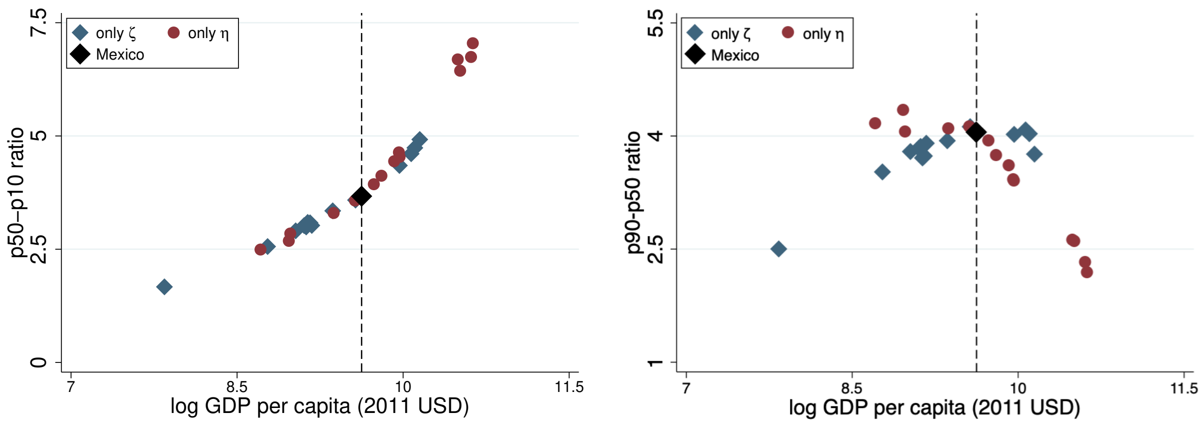
(a) Average firm size (b) Wage and salary employment




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

Figure 14: Earnings inequality across countries: distortions vs. frictions

(a) p50-p10 ratio (b) p90-p50 ratio




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

Figure 13 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, $\zeta$, or matching frictions, $\eta$, alone can generate a positive correlation between GDP per capita on the one hand and the average firm size or wage and salary employment on the other. Yet, their impact differs markedly depending on whether we look at countries poorer or richer than Mexico.

While search frictions play a more important role in richer countries, correlated distortions are more crucial for generating observed patterns in poorer ones. Consider, for example, panel (a) of Figure 13. 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 $\zeta$. In contrast, lower search frictions generate a larger average firm size for richer countries, but they alone can not push the firm size to levels observed for 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 in panel (b) of Figure 13.[8]

What about earnings inequality? Start with the role of correlated distortions. Changes in $\zeta$ 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 14). 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 Figure 14).

## 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 again by comparing the UK versus Indonesia.

**Distribution of firm-level revenues.** The first channel operates through changes in the distribution of firm-level revenues. Panel (a) in Figure 15 shows the average distortions, $\tau$, faced by firms of different productivity levels $z$ in Indonesia (red dots). Recall

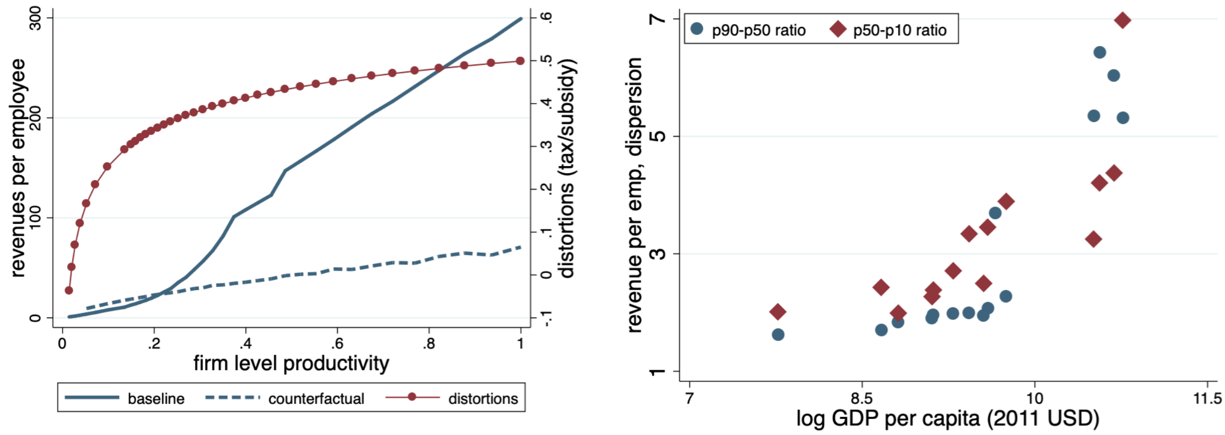
[8] Note that for high levels GDP per capita, the relation between GDP per capita and firm size becomes flat or can even be slightly negative. This is due to general equilibrium effects. A low $\zeta$ encourages entry and, because of matching function congestion, firm growth can slow down.

that the implicit tax rate, $\tau$, is zero in the baseline economy. In Indonesia, with $\zeta$ 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.

Figure 15: Firm-level revenues per employee

(a) UK versus Indonesia

(b) Cross-country

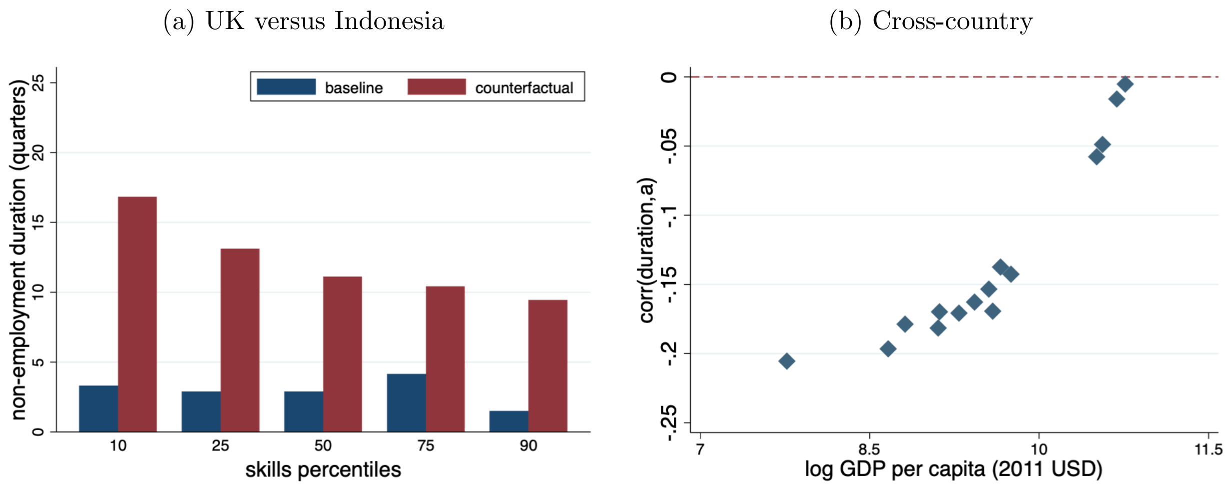



Notes: The red dots in Panel (a) shows the correlated distortions in Indonesia (left axis). The solid and dashed-blue lines show revenue per worker in the benchmark (UK) and and counterfactual (Indonesia), respectively. Panel (b) shows p90-p50 (blue dots) and p50-p10 (red diamonds) ratios for the revenue-per-worker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

Panel (a) of Figure 15 also reports the average revenues per employee in the UK and Indonesia. A higher value of $\zeta$ 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 progressively vanish. Panel (b) in Figure 15 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 14).

**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 16 shows the average non-employment duration for workers of different skills in the benchmark (UK) and the counterfactual (Indonesia). In the UK, it takes on average 3 quarters for a worker to find a job, and durations are relatively similar across workers with different skill levels. Non-employment spells are much longer in Indonesia, about 11 quarters for the average worker. Furthermore, the average spells decline significantly by worker' skills. Hence, as a country gets richer, non-employment duration shrinks and becomes more uniform across workers with different skills. The decline is most significant for workers with low skills.

Figure 16: 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.

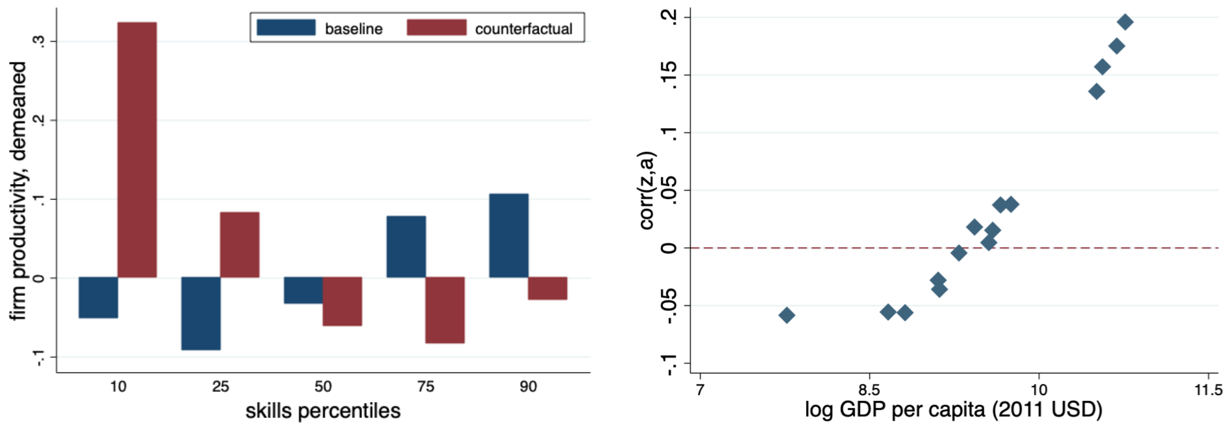
Panel (b) in Figure 16 shows the correlation between workers' skills, $h$, and their non-employment duration across countries. The correlation is always negative, i.e., more skilled workers get out of non-employment faster. But the correlation is stronger in the poorest countries, it increases monotonically 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 17 describes the sorting between firms and workers. In panel (a), the horizontal axis ranks workers again by their skills while the vertical axis 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.

Figure 17: Sorting

(a) UK versus Indonesia

(b) Cross-country




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.

Stronger labor market frictions make it costly for high productivity firms to wait for better workers, and they end up hiring any worker they can find.[9] It takes about 3 quarters to fill a vacancy in the UK. In Indonesia, a firm has to wait on average 5.6 quarters to encounter a potential hire. Panel (b) in Figure 17 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

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

from matching with high productivity firms. The latter is relatively stronger for less skilled workers, who progressively populate employment and accumulate human capital as GPD 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 14).

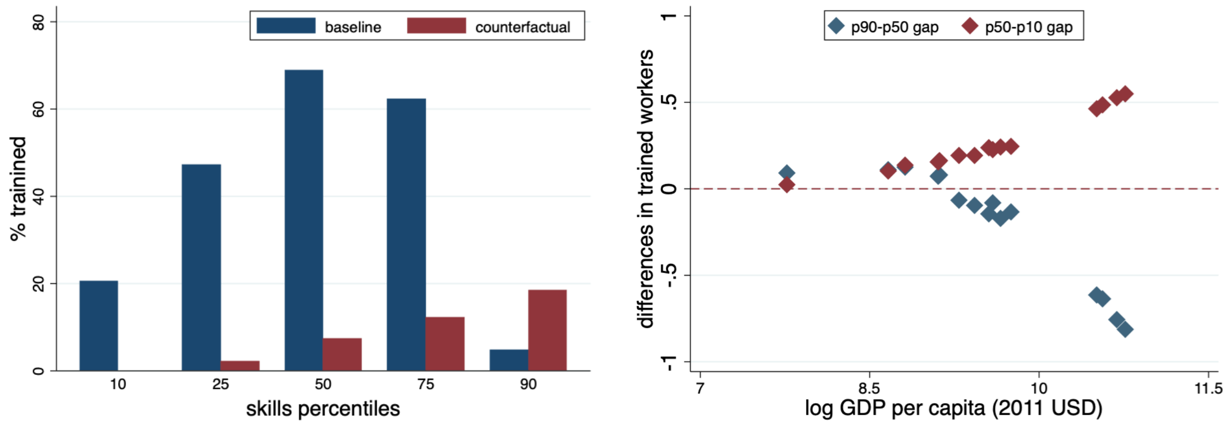
## 6.2 Role of Training

What is the role of OTJ training in the model? Panel (a) in Figure 18 shows the fraction of workers who receive training in the model economy, conditional on their skills.

Figure 18: Training

(a) UK versus Indonesia

(b) Cross-country




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.

In the benchmark (the 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

Table 6: A World without OJT Training

| | Baseline | Counterfactual | Baseline | Counterfactual | Explained |
|---|---|---|---|---|---|
| | with OTJ training | | w/o OTJ training | | |
| | (1) | (2) | (3) | (4) | |
| Elasticity of matching function: $\eta$ | 0.542 | 0.313 | 0.542 | 0.313 | - |
| Distortion correlation: $\zeta$ | 0 | 0.308 | 0 | 0.308 | - |
| Home production: $b$ | 20.94 | 3.505 | 20.94 | 3.505 | - |
| | | *Aggregates* | | | |
| Non-employment rate | 0.212 | 0.593 | 0.203 | 0.439 | 37.96% |
| Average earnings | 1 | 0.124 | 1 | 0.140 | 1.84% |
| Income per capita | 1 | 0.061 | 1 | 0.086 | 2.69% |
| | | *Earnings profile over experience/tenure* | | | |
| Earnings growth, $\mathrm{E}[\log(w_{25}/\bar{w}_1)]$ | 0.801 | 0.280 | 0.731 | 0.363 | 29.45% |
| | | *Earnings inequality* | | | |
| Mean-median ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 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, $\mathrm{p}^{90}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 2.551 | 4.462 | 2.815 | 3.990 | 33.87% |
| 50-10 pct. ratio, $\mathrm{p}^{50}[w_{it}]/\mathrm{p}^{10}[w_{it}]$ | 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).

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

Panel (b) in Figure 18 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 a baseline economy without training, i.e., we set $p^t = 0$ so that no one receives training. The re-calibration uses the same targets in Tables 4, except the ones on training. The moments and parameter estimates are reported in Appendix F. Columns 1 and 3 in Table 6 show the baseline economy with and without training. We then conduct the same counterfactual exercise and move to Indonesia, i.e., impose $\eta = 0.313$ and $\zeta = 0.308$. Columns 2 and 4 in Table 6 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 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 G, 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-$(z, \xi)$ 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.[10] 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.

[10] Attanasio 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.

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. In particular, 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

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

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

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

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.

Table 7 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.[11] Figure 19 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

[11] Alfonsi 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.

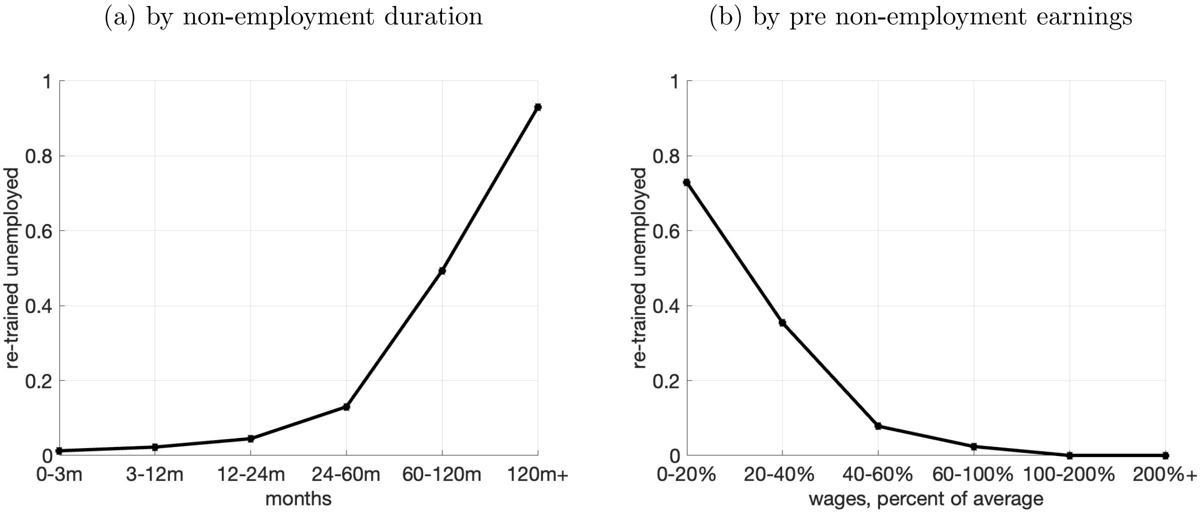
Table 7: Re-training program

| | UK | Indonesia | |
|---|---|---|---|
| | Baseline | Counterfactual | |
| | (1) | (2) | (3) |
| Elasticity of matching function: $\eta$ | 0.542 | 0.313 | 0.313 |
| Distortion correlation: $\zeta$ | 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 |
| *Re-trained workers* | | | |
| $\mathrm{E}\left(\frac{\#\text{re-trained workers}}{\#\text{non-employed workers}}\right)$, % | 0 | 0 | 43.07 |
| *Aggregates* | | | |
| Non-employment rate | 0.212 | 0.593 | 0.471 |
| 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 |
| *Earnings profile over experience* | | | |
| Earnings growth, $\mathrm{E}[\log(w_{25}/\bar{w}_1)]$ | 0.801 | 0.280 | 0.329 |
| *Earnings inequality* | | | |
| Mean-median ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 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.

pre 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 RTC 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

Figure 19: 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.

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; 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 reduce non-employment from 60% to 47% (including both re-trained workers and searchers). 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 distortions are correlated with firms' productivity, the size distribution shifts to the left, resulting 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 the fraction of workers receiving training).

The model also delivers a natural framework to study how the distribution of earnings changes with economic development. The data shows it does change in a particular way: while 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.

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.

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.

Bartelsman, E., Haltiwanger, J., and Scarpetta, S. (2009). Measuring and analyzing cross-country differences in firm dynamics. In *Producer dynamics: New evidence from micro data*, pages 15–76. University of Chicago Press.

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.

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.

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.

Chernozhukov, V. and Hong, H. (2003). An mcmc approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Cooper, R., Haltiwanger, J., and Willis, J. L. (2007). Search frictions: Matching aggregate and establishment observations. *Journal of Monetary Economics*, 54:56–78.

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.

Den Haan, W. J., Ramey, G., and Watson, J. (2000). Job destruction and propagation of shocks. *American economic review*, 90(3):482–498.

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. *New York University*.

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.

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.Louise*.

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.

Hallward-Driemeier, M. and Rijkers, B. (2013). Do crises catalyze creative destruction? firm-level evidence from indonesia. *Review of Economics and Statistics*, 95(1):1788–1810.

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.

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.

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.*

Marinescu, I. and Triyana, M. (2016). The sources of wage growth in a developing country. *IZA Journal of Labor & Development*, 5(1):1–43.

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.

Mortensen, D. T. and Pissarides, C. A. (1999). New developments in models of search in the labor market. *Handbook of labor economics*, 3:2567–2627.

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.*

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.

# A Data Appendix

## A.1 Earnings inequality

To construct earnings inequality we use data from two different sources: the EU Statistics on Income and Living Conditions (SILC) dataset and the IPUMS. We collect information for the 57 countries spanning one or multiple years. Table 8 reports the list of country, year and source.

### A.1.1 IPUMS-International database

IPUMS-International collects cross-country census microdata on individual demographics, labour market outcomes and income among the others. For each country in Table 8 sourced from IPUMS, all the information recorded refer to a representative and stratified samples of the resident population. Sampling and stratification details available here. The surveys allows to identify whether or not the respondent was working over a specified period of time (variable EMPSTAT)[12] When information on employment status is missing, we use information on the average number of hours worked per week overall (variable HRSWORK1) or in the main job (variable HRSMAIN). Hence we define a person to be working if she reports positive number of hours worked (in at least one of the above measure).

These two variables do not distinguish between employees and self-employed workers. To this purpose, we use the variable INCWAGE, which records the respondent's weekly, monthly or annual wage and salary income for employed workers.[13] We annualize weekly or monthly wage and salary income estimates by multiplying them by 52 or 12 respectively. This variables does not include income from self-employment.

In our final sample, we only consider working individuals who report strictly positive wage and salary income. We exclude from the sample working individuals with zero wage and salary income.

### A.1.2 EU Survey of Income and Living Standard (SILC) database

The EU-SILC collects comparable cross-sectional microdata on income and other living conditions. For each country in Table 8 sourced from the EU-SILC, a representative sample of private households is surveyed - and their current members aged 16 and more

[12]See https://international.ipums.org/international-action/variables/EMPSTAT for a description of how employment status is harmonized across countries.

[13]See https://international.ipums.org/international-action/variables/INCWAGE for a description of how wage and salary income is harmonized across countries.

are interviewed. More information about sample size and stratification are reported in https://ec.europa.eu/eurostat/web/income-and-living-conditions/data.

For each interviewed household members, the survey collects information on several demographic characteristics - age, gender, marital status, citizenship and head of households - education attainment, and labor market outcomes. Among the labor market outcomes, the survey uses self-defined current labor market status to distinguish working from non-working individuals (variable PL040). The self-declared main activity status is determined on the basis of whether the interviewed performs any work for pay or profit during the reference week or if he/she was not working but had a job or business from which he/she was absent during the reference week.

The survey allows us to distinguish employed workers from self-employed and family workers. Employees are defined as persons who work for a public or private employer and who receive compensation in the form of wages, salaries, fees, gratuities, payment by results or payment in kind; non-conscripted members of the armed forces are also included. Apprentices, or trainees receiving remuneration are considered as employees. Self-employed persons are defined as persons who work in their own business, professional practice or farm for the purpose of earning a profit, while family workers are persons who help another member of the family run an agricultural holding or other business, provided they are not considered as employees. We exclude self-employed and family workers from our final sample.

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. This information is recorded by the variable PY010G for the single household members. We do not consider any non-monetary salary income components. The income reference period for most of countries is the calendar year previous to the survey year with two exceptions. In Ireland the income reference period is the last twelve months, whereas in the United Kingdom the current income is annualised and aims to refer the current calendar year, i.e. weekly estimates are multiplied by 52, monthly by 12. Reimbursements for work-related expenses, severance and termination pay, employers' social insurance contributions are excluded from employee income.

### A.1.3 The Luxembourg Income Study (LIS) database

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. This dataset provides information on labour income, capital income, social benefits and private transfers, as well as taxes and contributions, demography, employment, and expenditures. More information about sample and stratification is reported at www.lisdatacenter.org/our-data/lis-database.

Table 8: Data source

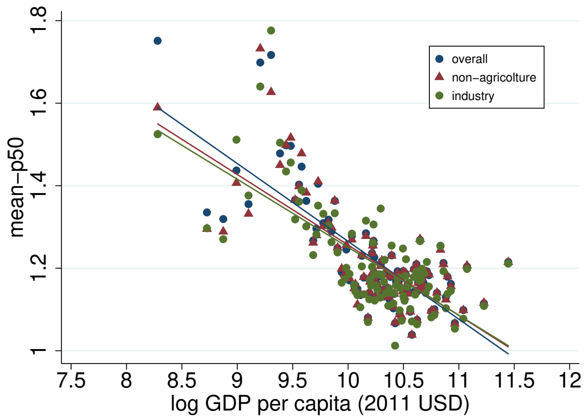
| Country | Year | Source |
|---|---|---|
| Australia | 1981, 1985, 1989, 1995, 2001, 2003, 2004, 2008, 2010, 2014 | LIS |
| Austria | 1994, 1997, 2000, 2004, 2005, 2007, 2010, 2013, 2016 | EU-SILC, LIS |
| Belgium | 1995, 1997, 2000, 2003-216 | EU-SILC, LIS |
| Bulgaria | 2007, 2009 | EU-SILC |
| Brazil | 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.1.4 Further empirical evidence on earnings inequality across countries

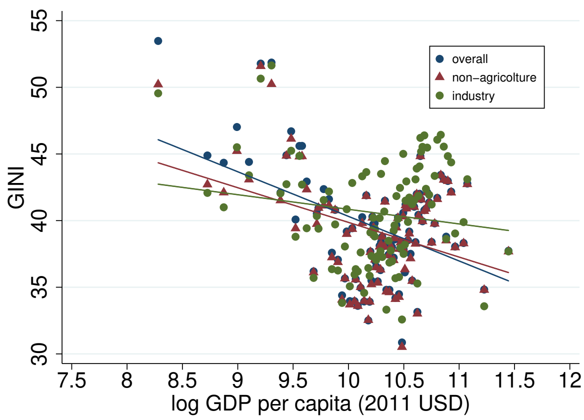
In Figure 20 we report the mean-median earnings ratio, the GINI coefficient, and the p50-p10 and p90-p50 earnings ratios across countries separately for individuals 1) working in any sector (blue dots), ii) working in any non-agriculture sectors (red dots), and iii) working only in the industrial sector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

Figure 20: Earnings inequality across countries, by sectors

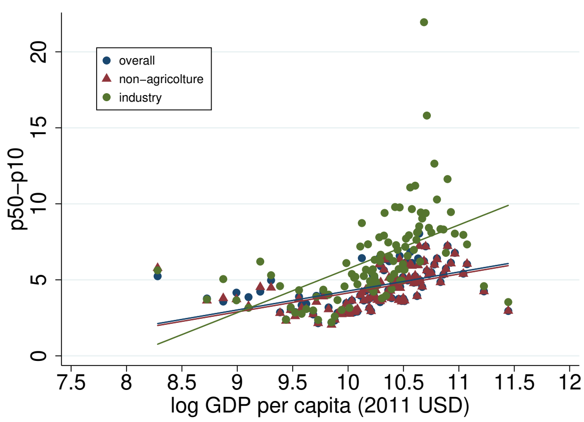
(a) Mean-median ratio



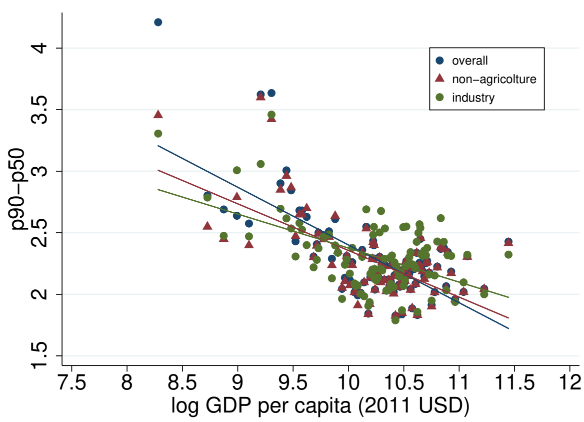

(b) GINI




(c) p50-p10 ratio



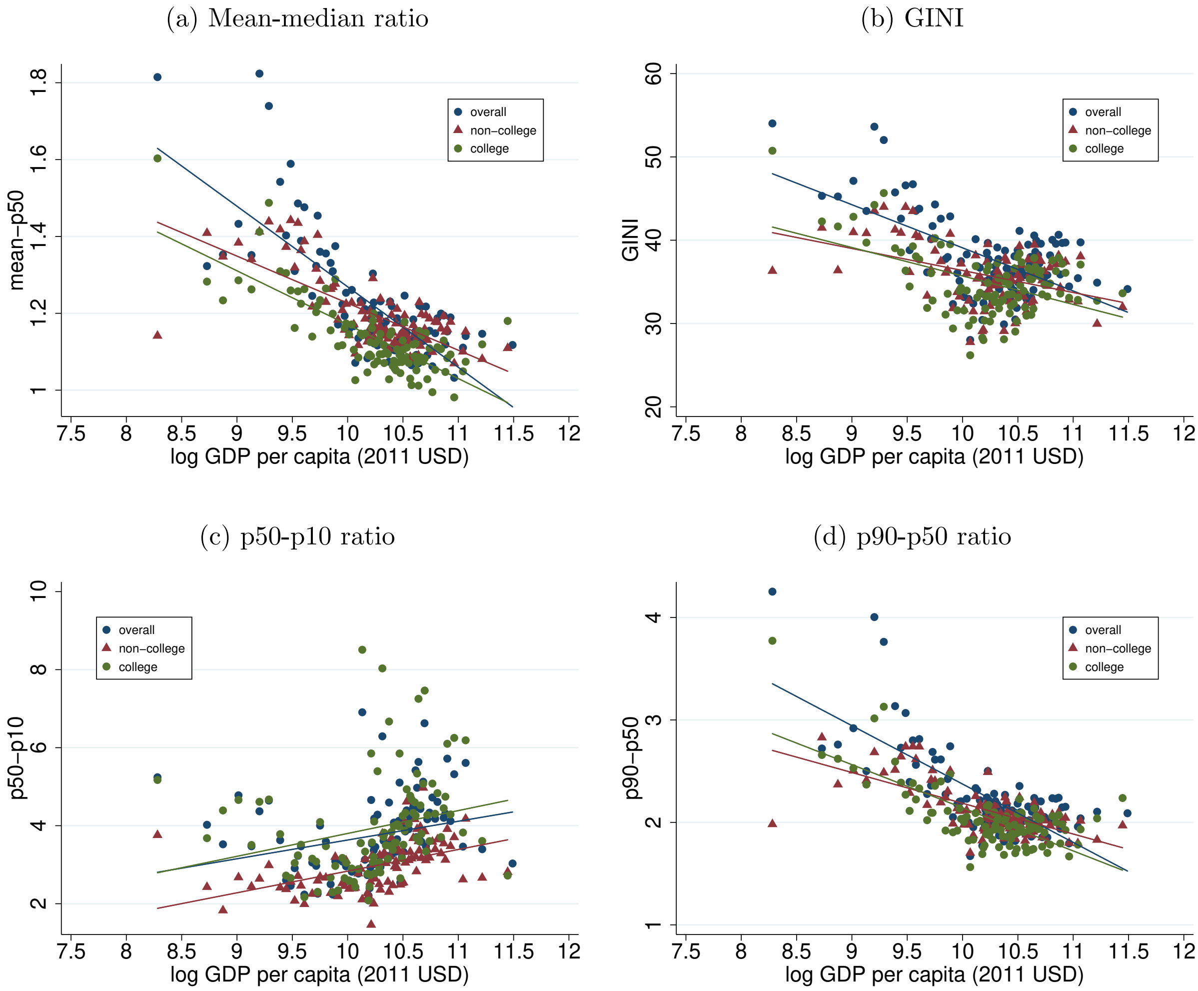

(d) p90-p50 ratio




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.

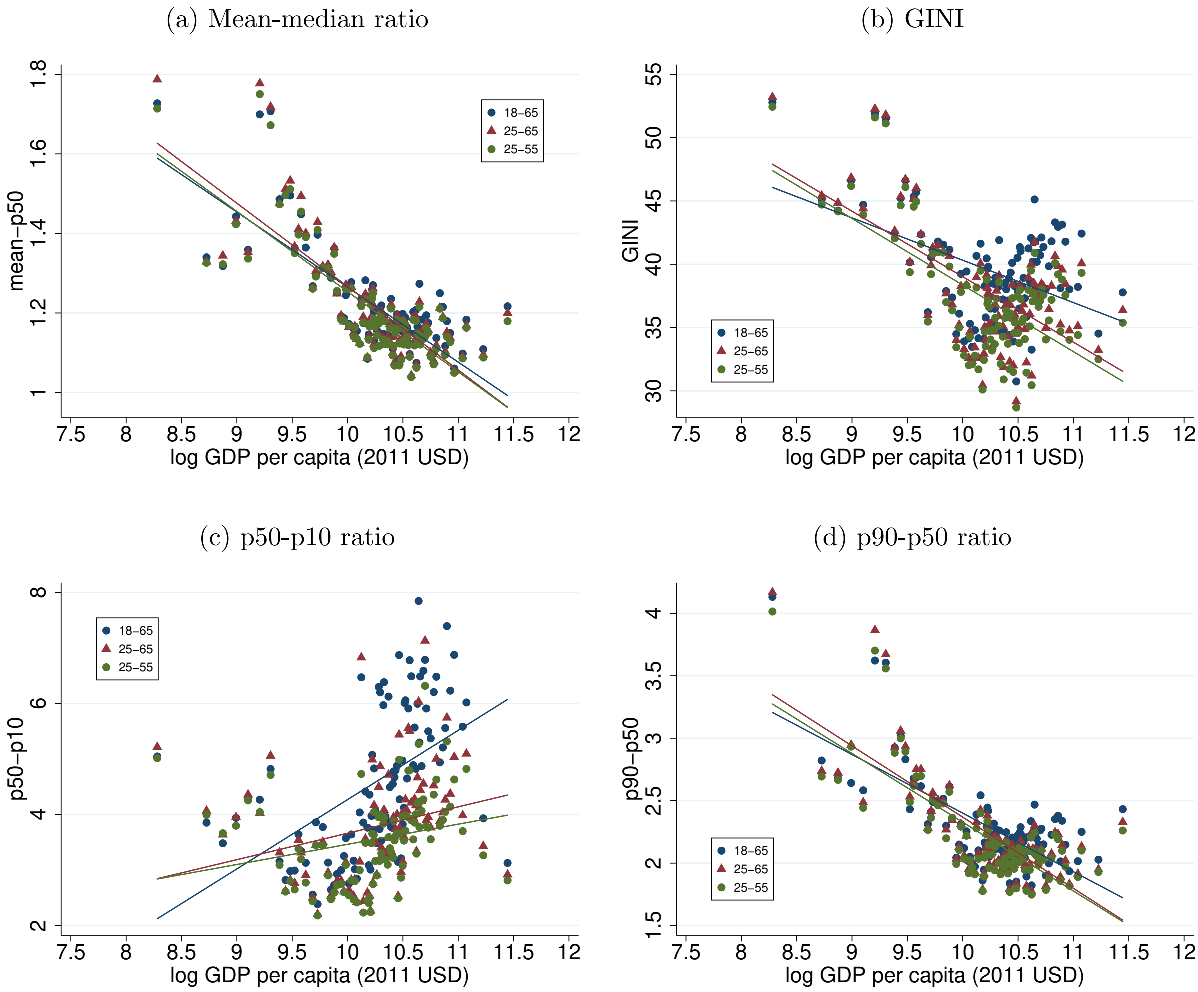
In Figure 21 we report the mean-median earnings ratio, the GINI coefficient, and the p50-p10 and p90-p50 earnings ratios across countries separately for individual in the sample i) without college degree (red dots) and ii) with a college degree (green dots).

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.

Figure 22: Earnings inequality across countries, by age



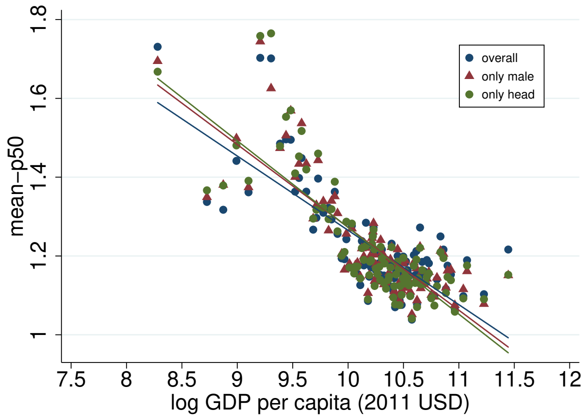

Notes: Each dot corresponds to the average outcome for countries in a given percentile of the GDP per capita distribution. Outcomes are reported as residuals from a regression with year-fixed effects. Source: IPUMS, EU-SILC, LIS and author's calculations.

In Figure 22 we report the mean-median earnings ratio, the GINI coefficient, and the p50-p10 and p90-p50 earnings ratios across countries separately for workers within different age categories, i.e. we consider i) the group of individuals 18-65 years old (blue dots), ii) the group of individuals 25-65 years old (red dots), and iii) the group of prime age individuals, 25-55 years old (green dots)

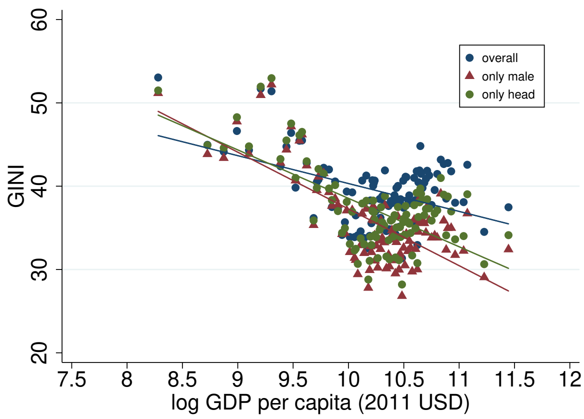
In Figure 23 we report the mean-median earnings ratio, the GINI coefficient, and the p50-p10 and p90-p50 earnings ratios across countries separately for workers belonging to different demographic groups, i.e. we consider i) the group of male and female individuals pooled together (blue dots), ii) the sample of only male individuals (red dots), and iii) the sample of only household heads, regardless their gender (green dots)

Figure 23: Earnings inequality across countries, by demographic groups

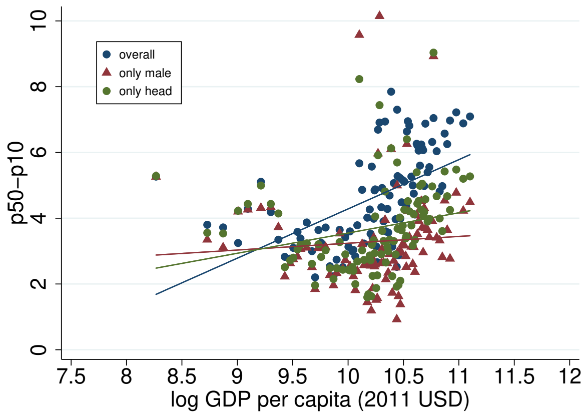
(a) Mean-median ratio



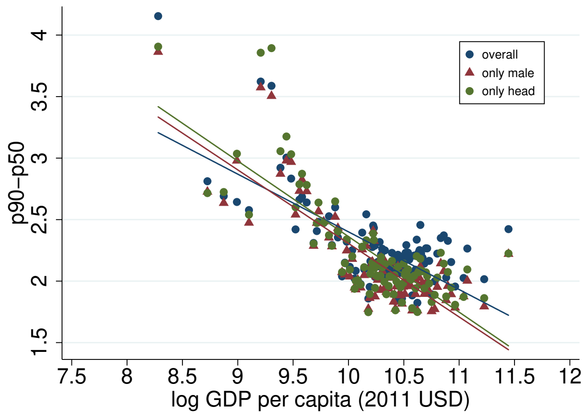

(b) GINI




(c) p50-p10 ratio



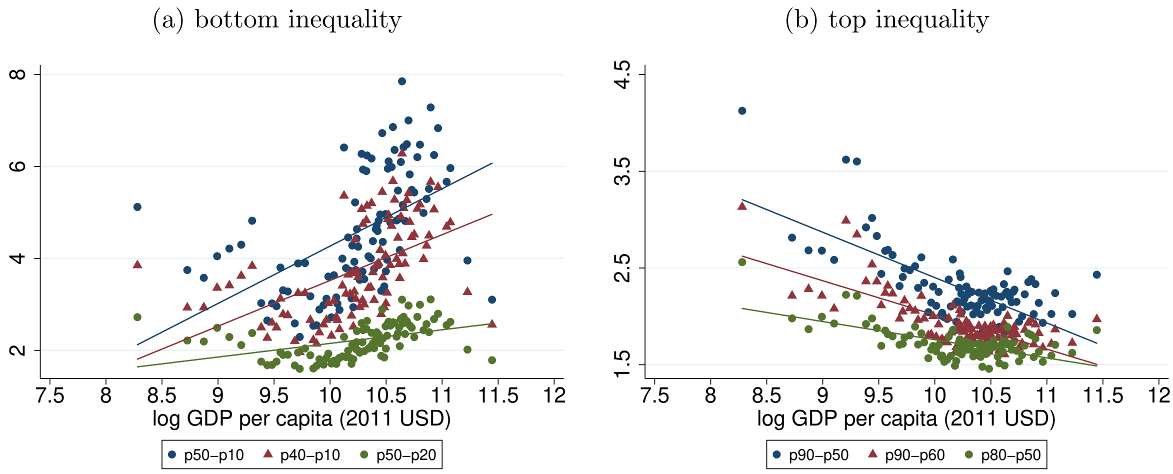

(d) p90-p50 ratio




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 panel (a) of Figure 24 we report alternative measures of bottom earnings inequality, i.e. the p40-p10 and p50-p20 earnings ratios, and we confront them to the p50-p10 earnings ratio. In panel (b) of Figure 24 we report alternative measures of top earnings inequality, i.e. the p90-p60 and p80-p50 earnings ratios, and we confront them to the p90-p50 earnings ratio.

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. In red we report the estimated slope in the regression. Robust standard errors are in parenthesis. Source: IPUMS, EU-SILC, LIS and author's calculations.

## A.2 On-the-job Training

### A.2.1 World-Bank Enterprise Survey (WB-ES)

The World-Bank Enterprise Survey (WB-ES) is a firm-level survey of a representative sample of an economy's private sector. The survey takes the form of repeated cross-section dataset, where in each countries different firms are surveyed across years. The survey only targets formal (registered) companies with 5 or more employees, operating in the the manufacturing and services sectors. This corresponds to economic activities classified with ISIC codes 15-37, 45, 50-52, 55, 60-64, and 72 (ISIC Rev.3.1). Services firms include construction, retail, wholesale, hotels, restaurants, transport, storage, communications, and IT. Firms with 100% government/state ownership are not eligible to participate in the survey. For more details about the sampling methodology, see https://www.enterprisesurveys.org/en/methodology.

The survey includes a large set of information about firm characteristics. For each surveyed firms the dataset records demographic information (age, region of operation, ownerwhip status), number of employees, annual sales, annual wage bills, and different measures of training provision, among the others 1) whether a firm has provided training to all or some of the workforce, and 2) the share of workforce who received training in a given year. Firm-level average wage is constructed using wage bill divided by the number of employees.

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.[14] We remove also Sweden (which is instead included in the Eurostat CV-TS dataset). This leaves us with the following 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 Bissau, Guyana, Honduras, Hungary, India, Indonesia, Iraq, Israel, Jamaica, Jordan, Kazakhstan, Kenya, Kosovo, Kyrgyzistan, Lao P.D.R., Latvia, Lebanon, Liberia, Lithuania, Macedonia, Madagascar, Malaysia, Mali, Mauritius, Mexico, Micronesia, Moldova, Mongolia, 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, WestBank, Yemen, Zambia and Zimbabwe.

### A.2.2 Continuing Vocational Training Survey (CV-TS)

The Continuing Vocational Training Survey (CV-TS) is an firm-level survey belonging to the Eurostat Education and Training Dataset. The survey covers a representative sample of formal enterprises with 10 or more employees in 27 EU countries (Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain and Sweden) plus Norway, North Macedonia and United Kingdom, for the years 2005, 2010 and 2015. The sectors covered are manufacturing and services (mainly). This corresponds to economic activities classified with NACE Rev 1.1 codes C, D (15-16, 17-19, 21-22, 23-26, 27-28, 29-33, 34-35, 20+36+37), E, F, G (50, 51, 52), H, I (60-63, 64), J (65-66, 67), K+O.

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, together with firm-level number of employees. To construct our main empirical evidence,

[14] These countries are Rwanda, Timor-Leste, Togo, Lesotho, Nepal, Senegal, Thailand, Venezuela, Suriname, Sudan, Malawi, Niger, Mauritania, Mozambique, Benin and Tajikistan.

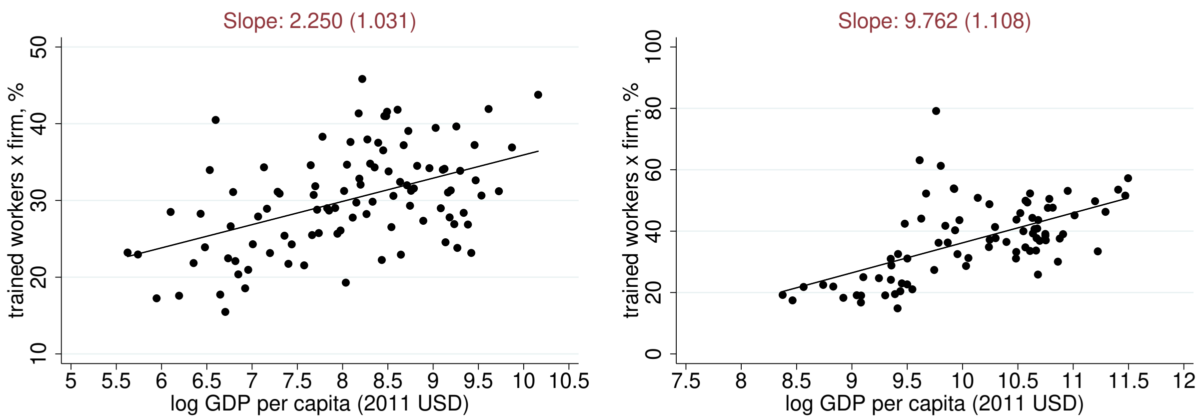
we use the aggregate statistics reported by the Eurostat, available here. Statistics are constructed for each country and year overall, and broken by firm size categories.

### A.2.3 Other datasets

We merge the ES-WB and the CV-TS with information on GDP per capita and population at country-year level. Data on GDP per capita and population are taken from the World Bank Indicator Survey. GDP per capita is expressed in constant 2011 international dollars. 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

(a) World Bank Enterprise Survey (b) Continuing Vocational Training Survey




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.2.4 Further empirical evidence on job training

Using the World-Bank Enterprise Survey we can measure the share of workers trained within each firm as follows:

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

The Eurostat reports this variable constructed using data from the CV-TS.

Figures 25 (a) and (b) report the average share of workers (percent) within each firm receiving formal job training programs. In both figures, the measure of training provision is scattered over the country-average real GDP per capita. Firms in more developed countries provide training to a larger share of their workers.

Table 9: Job training across firm size

| | Trained workers within firms, % | | | | | | |
|---|---|---|---|---|---|---|---|
| | WB-ES | | | | CVTS | | |
| | LAC | ME+AFR | ASIA | others | | EU15 | non-EU15 |
| Firm size (# employees) | | | | | Firm size (# 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).

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. This coefficient implies that one log point higher GDP per capita is associated with 10% percent more workers within firms receiving formal training.

Table 9 reports the share of trained employees within the workforce in firms with different size for different groups of countries. Larger firms provide OTJ training to a larger set of their workforce, consistently in each group of countries.

# B Model Appendix

## B.1 The surplus function

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

$$S(z, \xi, a) = M(z, \xi, a) - J^u(a), \tag{16}$$

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

$$M(z, \xi, a) = \mathbf{1}^h(z, \xi, a)M^h(z, \xi, a) + (1 - \mathbf{1}^h(z, \xi, a))J^{u,h}(a) \tag{17}$$

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

$$\begin{aligned} M^h(z, \xi, a) &= r(z, a) + \frac{(1-\delta_w)}{1+r}(1 - (1-\delta_f)(1-\delta_s))J^{u,h}(a) \\ &+ \frac{(1-\delta_w)}{1+r}(1-\delta_f)(1-\delta_s)\left[-\mathbf{1}^t(z, \xi, a)\xi + (1 - p^h(z, \xi, a))M(z, \xi, a) + p^h(z, \xi, a)M(z, \xi, a+1)\right] \end{aligned} \tag{18}$$

Combining equations (18) and (3), we can write the surplus function as

$$S(z, \xi, a) = \max\{0, S^h(z, \xi, a)\} \tag{19}$$

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.

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

or equivalently

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

## 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, firms' entry and vacancy 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 (7).

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 grid points for firm productivity, 20 grid points for firm-specific training costs, and 60 grid points for workers human capital. We fix minimum and maximum (log) productivity and (log) human capital to -4 and 4 respectively, covering 99.9% of both calibrated distributions. We directly calibrate the boundaries for training costs. To find an equilibrium for this economy, we employ the following algorithm:

1. Formulate a guess for the workers' job contact rate, $\phi_w^0$, and use the definition of matching function to compute the job contact rate for firms, $\phi_f^0$ as follows

$$\phi_f^0 = (1 - (\phi_w^0)^\eta)^{\frac{1}{\eta}}$$

2. Formulate a guess for the distribution of vacancies over firm-level states $(z, \xi)$, $\psi_v^0(z,\xi)$

   2.1. Given $\phi_w^0$ and $\psi_v^0(z,\xi)$, solve for the surplus function, $S^h(z,\xi,a)$. To solve for it, we use value function iteration. We measure convergence using the euclidean distance and stop when tolerance is lower or equal to 1e-04.

   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 $\phi_w^0$, $\psi_v^0(z,\xi)$, $\mathbf{1}^h(z,\xi,a)$ 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 $\phi_f^0$, $\mathbf{1}^h(z,\xi,a)$, $\psi_a^u(a)$, and the bargaining splitting rule, solve the vacancy posting problem of the firm and obtain the optimal policy for vacancy $v(z,\xi)$.

2.5. Compute the firm value at entry, $\Pi(z,\xi)$, and obtain a solution to the entry decision of the firm $\mathbf{1}^e(z,\xi)$

2.6. Given $v(z,\xi)$ and $\mathbf{1}^e(z,\xi)$, construct a new guess for the distribution of vacancy over firm states, $\psi_v^1(z,\xi)$

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 22.1.

In the algorithm, we use a tolerance level of 1e-03.

2.8. Iterate till convergence

3. Compute the measure of entrant firms

$$M = M_e \int_{z\in\mathcal{Z}} \int_{\xi\in\mathcal{E}} \mathbf{1}^e(z,\xi)\psi_z(z)\psi_\xi(\xi)dzd\xi \tag{20}$$

and use stationarity condition to compute total number of firms

$$N = \frac{M}{\delta_f} \tag{21}$$

4. Construct the aggregate measure of vacancy posted

$$v = N\bar{v} \tag{22}$$

where $\bar{v}$ is the average number of vacancy posted, equal to

$$\bar{v} = \int_{z\in\mathcal{Z}} \int_{\xi\in\mathcal{E}} \mathbf{1}^e(z,\xi)v(z,\xi)\psi_z(z)\psi_\xi(\xi)dzd\xi \tag{23}$$

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

6. Check for convergence:

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

In the algorithm, we use a tolerance level of 1e-03.

7. Iterate till convergence

Use $\phi_w^*$, $\psi_v^*(z,\xi)$, 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

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

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.

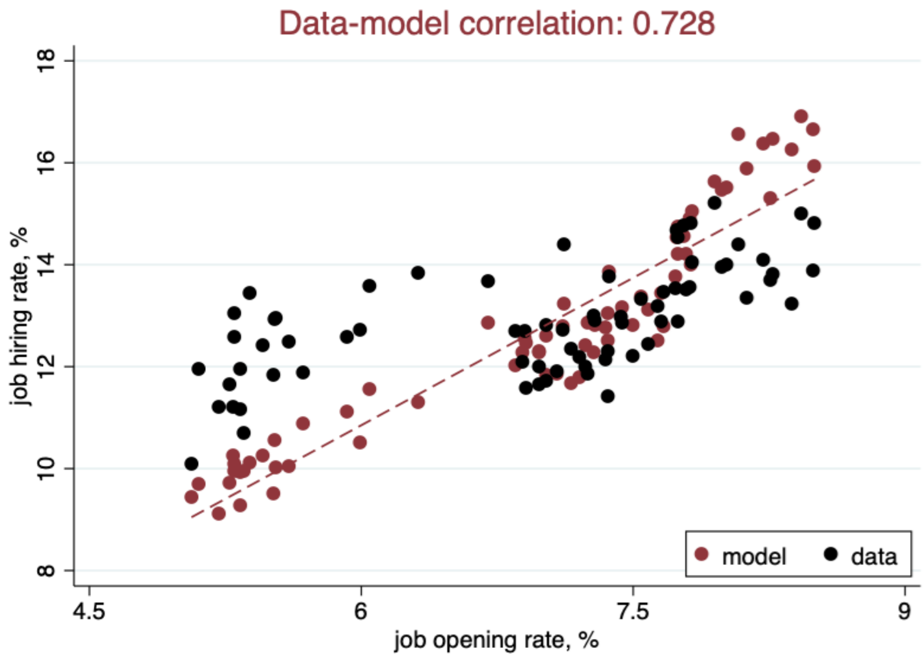
Table 10: Matching elasticity estimation

| Parameters | Description | Estimates | St.Error |
|---|---|---|---|
| $\eta$ | Matching function | 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 |
| $\mathbf{1}_t^{q=2}$ | 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 $\eta$ 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. Figure 26 shows the observed hiring rates against the corresponding predicted values obtained from the estimated matching function.

Figure 26: Matching function estimation fit




## C.2 Model estimation

### C.2.1 Data

Table 11 reports descriptive statistics for the sample of households in the Five-Quarter Longitudinal LFS. We restrict our focus to women and men of age 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

Table 11: Descriptive Statistics

| | Mean | SD | Min | Max | N |
|---|---|---|---|---|---|
| *Employed workers* | | | | | |
| 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 |

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 $\geq 24$ 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 the 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.

$$w^h_{it} = \delta_{y(t)} + \delta_{q(t)} + \epsilon_{it}$$

where $w^h_{it}$ 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 Algorithm

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

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

to treat the equilibrium job contact rate, $\phi_w$ as a parameter to estimate, and let the measure of potential entrants, $M_e$, as an equilibrium object, equal to the solution of the following equilibrium equation:

$$\phi_w - \frac{U(\bar{v}\delta_f M)}{(U^\eta + (\bar{v}\delta_f M)^\eta)^{\frac{1}{\eta}}} = 0 \tag{24}$$

where $M$ is defined in equation (20). To calibrate the model, we follow this algorithm:

1. Guess the following set of parameters:

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

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

2. Given $\phi_w^0$, compute job contact rate for firms, $\phi_f^0$ as follows

   $$\phi_f^0 = (1 - (\phi_w^0)^\eta)^{\frac{1}{\eta}}$$

3. Proceed as in the solution algorithm, step 2

4. Obtain the equilibrium measure of potential entrants $M$ solving equation (24)

5. Use parameter guesses, $M$, $\psi_v^*(z,\xi)$, and relevant policy functions to simulate a large panel of firms and workers

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

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

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

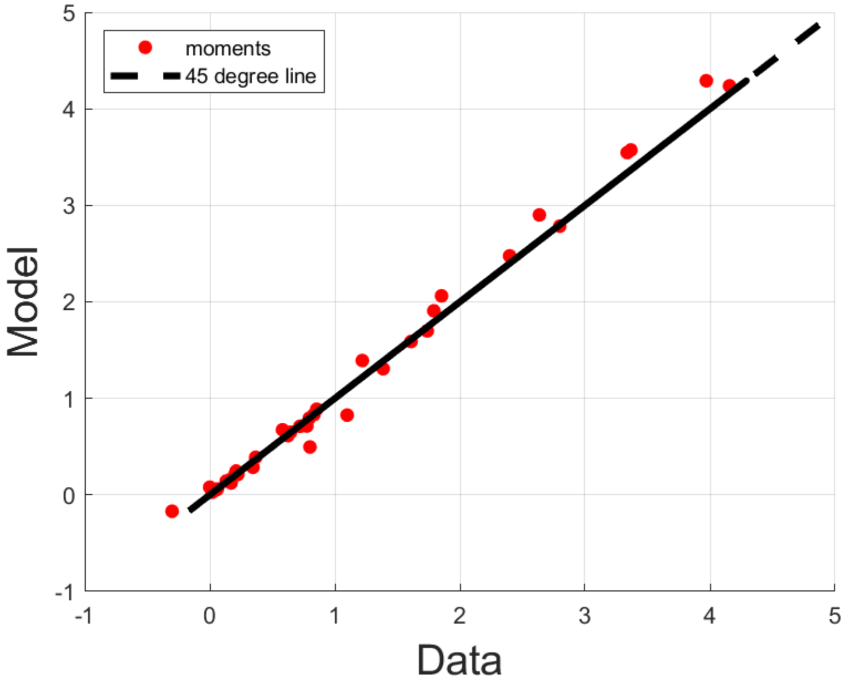
7. Evaluate the distance function:

   $$D(\vartheta^0) = \bar{d}(\vartheta^0)'\bar{d}(\vartheta^0) \tag{25}$$

8. Update guesses and iterate to minimize the distance function

We follow a genetic algorithm to update the vector of guesses. At the found minimum, the log deviation between empirical and simulated moments is 0.086. Figure 27 displays the estimation fit by scattering model-generated statistics against their data counterpart.

Figure 27: Parameters estimation fit




### C.2.3 Standard errors

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

$$f(\vartheta) = \frac{e^{D(\vartheta)}p(\vartheta)}{\int e^{D(\vartheta)}p(\vartheta)d\theta}$$

where $D(\vartheta)$ is defined in equation (25) 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 follow Lise et al. (2016) and use the Metropolis–Hastings algorithm. This algorithm generates a chain of parameters $\vartheta^0, \vartheta^1, \vartheta^2, ...$ as follows. First, we choose a starting value $\vartheta^j$. Next, we impose the proposal density to be a uniform and we extract a new guess $\vartheta^p$ from it. Finally, we update from $\vartheta^{j+1}$ from $\vartheta^j$ for $j = 1, 2, ...$, using the following rule:

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

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.

# D UK versus Indonesia

In Table 12 we report selected outcomes for the benchmark economy, the UK, and for a selected counterfactual economy, Indonesia. To obtain the latter, elasticity of matching function $\eta$ and correlated distortions are re-calibrated to values equal to 0.312 and 0.308, respectively. To do so, we target average firm size and wage and salary employment, and re-adjust home-production to keep its ratio with the average wage equal to the benchmark value. The fourth column of Table 12 reports the data counterpart for each statistics.

Table 12: Counterfactual outcomes

| | UK<br>Baseline | Indonesia<br>Counterfactual | Data |
|---|---|---|---|
| Elasticity of matching function: $\eta$ | 0.542 | 0.313 | - |
| Distortion correlation: $\zeta$ | 0 | 0.308 | - |
| Home production: $b$ | 20.94 | 3.505 | - |
| *Firm-level moments* | | | |
| Average firm size, $\mathrm{E}(\ell_t)$ | 16.185 | 5.179 | 4.141 |
| *Firm training provision* | | | |
| $\mathrm{E}\left(\frac{\#\text{training firms}}{\#\text{firms}}\right)$, % | 65.02 | 6.210 | 6.291 |
| *Aggregates* | | | |
| Non-employment rate | 0.212 | 0.593 | 0.569 |
| Average earnings | 1 | 0.124 | 0.112 |
| Income per capita | 1 | 0.061 | 0.100 |
| Job finding rate, $\phi_w$ | 0.259 | 0.064 | - |
| Job filling rate, $\phi_f$ | 0.298 | 0.172 | - |
| *Earnings Inequality* | | | |
| Mean-median ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 1.207 | 1.805 | 1.687 |
| GINI | 0.416 | 0.506 | 0.502 |
| 90-50 pct. ratio, $\mathrm{p}^{90}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 2.551 | 4.462 | 3.182 |
| 50-10 pct. ratio, $\mathrm{p}^{50}[w_{it}]/\mathrm{p}^{10}[w_{it}]$ | 5.262 | 2.729 | 1.934 |

# E Alternative mechanisms

We explore whether two alternative mechanisms operating through the labor market can account for the cross-country pattern of development. Again, we take the UK as a baseline economy and compare it against Indonesia (our counterfactual economy).

Table 13: Alternative mechanisms

| | UK | Indonesia | | | |
|---|---|---|---|---|---|
| | Baseline | Counterfactual | | | |
| | | Joint $(\eta, \zeta)$ | Only $\delta_s$ | Only $\delta_f$ | Joint $(\delta_s, \zeta)$ |
| | (1) | (2) | (3) | (4) | (5) |
| Elasticity of matching function: $\eta$ | 0.542 | 0.313 | 0.542 | 0.542 | 0.542 |
| Distortion correlation: $\zeta$ | 0 | 0.308 | 0 | 0 | 0.659 |
| Separation rate: $\delta_s$, % | 1.235 | 1.235 | 5.179 | 1.235 | 5.179 |
| Firm exit rate: $\delta_f$, % | 2.526 | 2.526 | 2.526 | 3.253 | 2.526 |
| Home production: $b$ | 20.94 | 3.505 | 15.943 | 19.45 | 1.400 |
| Average firm size, $\mathrm{E}[\ell_t]$ | 16.19 | 5.177 | 10.27 | 15.55 | 4.421 |
| Non-employment rate | 0.212 | 0.592 | 0.378 | 0.259 | 0.334 |
| Income per capita | 1 | 0.061 | 0.563 | 0.890 | 0.051 |
| Training provision, overall % | 65.02 | 6.210 | 50.08 | 60.31 | 0 |
| Earnings growth, $\mathrm{E}[\log(w_{25}/w_1)]$ | 0.801 | 0.280 | 0.568 | 0.756 | 0.614 |
| Mean-median ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 1.207 | 1.805 | 1.482 | 1.269 | 1.327 |
| GINI | 0.416 | 0.506 | 0.461 | 0.430 | 0.427 |

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

*Worker separation rate.* Donovan et al. (2020) documents a sharp reduction of worker separation rate over development. We study this channel in the context of our framework. To do so, we calibrate the counterfactual separation rate to match an average job tenure in Indonesia equal to 3.3 years, as documented by Marinescu and Triyana (2016), see Table 1. We compare baseline and counterfactual outcomes in Table 13, column 3. While a reduction in separation rate can qualitatively account for labor market and inequality patters observed over development, it fails to match almost all the the evidence quantitatively.

*Firm turnover rate.* Bartelsman et al. (2009) documents larger firm turnover in less developed countries. We study this channel by evaluating a counterfactual economy with larger firm entry and exit. We do this by matching a counterfactual yearly firm exit rate for the entire business industry in Indonesia equal to 13.66% as reported by Hallward-Driemeier and Rijkers (2013), Table 1, year 2001. We compare baseline and counterfactual outcomes in Table 13, column 4. Like for worker separation rate, reduction in firm exit rate can only qualitatively explain labor market features observed over development.

# F 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. In this version of the model, human capital accumulation when employed only happens through on-the-job learning. In this framework, 10 parameters are estimated to match 30 moments. Parameters estimates are reported in Table 14.

Table 14: Parameters calibrated through indirect inference

| Parameters | Description | Value |
|---|---|---|
| $c_e$ | Entry cost | 44.75 |
| $\sigma_z$ | Firm-productivity dispersion | 1.221 |
| $\lambda_1$ | Hiring costs, convexity | 2.532 |
| $M_e$ | Measure of potential entrants | 0.031 |
| $\beta$ | Bargaining power | 0.427 |
| $\sigma_a$ | Initial human capital dispersion | 1.035 |
| $p^e$ | Experience jump | 0.209 |
| $p^d$ | Depreciation jump | 0.430 |
| $b$ | Home production | 22.26 |
| $\delta_s$ | Match separation, % | 1.226 |

Targeted moments are reported in Tables 15. Compared to the baseline calibration with OTJ training, we target the exact set of moments except from those related to training provision. Counterfactual outcomes reported in column 3 and 4 of Table 6 are based on this calibration. The left panel of Table 6 reports i) firm-level moments, i.e. average firm size, average and dispersion of log firm size, ii) the share of firms in different firm-size bins, iii) selected firm size percentiles. The right panel of Table 6 reports worker-level moments related to i) earnings distribution, ii) job tenure returns and iii) aggregate moments, i.e. job duration rate and wage and salary employment rate.

Table 15: Targeted Moments (without OTJ training)

| | Data | Model | | Data | Model |
|---|---|---|---|---|---|
| *Firm-level moments* | | | *Earnings distribution* | | |
| 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 |
| *Firm size distribution* | | | 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 | *Job tenure return* | | |
| 50-99 employees | 3.21 | 4.621 | tenure<3 months | 1 | 1 |
| 100-249 employees | 1.73 | 3.080 | tenure∈[3,12) months | 1.055 | 1.053 |
| 250+ employees | 0.88 | 0.210 | tenure∈[12,24) months | 1.132 | 1.136 |
| | | | tenure≥24 months | 1.368 | 1.369 |
| *Firm size percentiles* | | | | | |
| 10th percentile | 1 | 1.181 | *Aggregate moments* | | |
| 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 model counterparts.

# G The role of on-the-job training

How much does OTJ training account for the cross-country patterns of inequality? To answer this question, we impose training decisions from the counterfactual economy (Indonesia) on UK firms. Hence, if a match between a type-$a$ worker and type-$(z, \xi)$ 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 $\zeta$ and $\eta$. Hence, this experiment isolates the impact of correlated distortions and higher labor market frictions on training decisions. Column 1 in Table 16 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.

Table 16: The Benchmark with Counterfactual Training Policies

| | Baseline with baseline training (1) | Baseline with counterfactual training (2) | Counterfactual (3) | Explained |
|---|---|---|---|---|
| Elasticity of matching function: $\eta$ | 0.542 | 0.542 | 0.313 | - |
| Distortion correlation: $\zeta$ | 0 | 0 | 0.308 | - |
| Home production: $b$ | 20.94 | 20.94 | 3.505 | - |
| *Aggregates* | | | | |
| Non-employment rate | 0.212 | 0.236 | 0.593 | 6.432% |
| Average earnings | 1 | 0.932 | 0.124 | 7.729% |
| Income per capita | 1 | 0.903 | 0.061 | 10.33% |
| *Earnings profile over experience/tenure* | | | | |
| Earnings growth, $\mathrm{E}[\log(w_{25}/\bar{w}_1)]$ | 0.801 | 0.760 | 0.280 | 7.994% |
| *Earnings inequality* | | | | |
| Mean-median ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 1.207 | 1.269 | 1.805 | 10.37% |
| GINI | 0.416 | 0.426 | 0.506 | 11.11% |
| 90-50 pct. ratio, $\mathrm{p}^{90}[w_{it}]/\mathrm{p}^{50}[w_{it}]$ | 2.551 | 2.876 | 4.462 | 17.06% |
| 50-10 pct. ratio, $\mathrm{p}^{50}[w_{it}]/\mathrm{p}^{10}[w_{it}]$ | 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).

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.