Misallocation and Inequality*

Nezih Guner
CEMFI

Alessandro Ruggieri
University of Nottingham

July 28, 2021

Abstract

For a large set of countries, we document how the distribution of labor earnings varies by GDP per capita. Changes in earnings distribution are not straightforward: while the standard deviation of log earnings increases with GDP per capita, the mean-to-median ratio declines. We interpret this fact within a model economy with heterogeneous workers and firms, featuring industry dynamics, search and matching frictions, skill accumulation of workers with learning-by-doing and on-the-job training, and earnings inequality both within and across firms. The benchmark economy is calibrated to the UK. We study how the earnings distribution changes as we introduce two distortions in the benchmark economy: wedges on firms' output that are correlated with firm productivity and limited visibility of unemployed workers to open vacancies. These distortions lead to resource misallocation and reduce employment, average firm size, and GDP per capita. They also affect how much firms are willing to pay to workers, how well high-skill workers are matched with high-productivity firms, and how much training workers receive. The model is consistent with a host of facts on changes in firm size distribution, firms' training decisions, and workers' life-cycle earnings profiles with development. It also delivers changes in earnings distribution in line with the data.

Keywords: labor market frictions, correlated distortions, productivity, establishment size, human capital accumulation, job training, life-cycle wage profile, inequality, development

JEL Classification: E23, E24, O11, O47

*Guner: CEMFI, Calle Casado de Alisal 5, 28014, Madrid, Spain, ngunermail@gmail.com. Ruggieri: School of Economics, Sir Clive Granger building, University of Nottingham, University Park, Nottingham, NG7 2RD, aruggierimail@gmail.com. The usual disclaimers apply.

1 Introduction

How does the distribution of labor earnings among workers change with development? We answer this question using household surveys from a large set of countries. We find that the distribution of earnings changes in a particular way as we move from poorer to richer countries. On the one hand, the mean of the earnings distribution increases as a country gets richer. On the other hand, the median increases even more, and, as a result, the mean-median ratio falls. At the same time, the standard deviation of earnings increases with GDP per capita. Hence, while income cut-offs for all percentiles increase with development, the percentiles below the median increase less.

We complement this novel fact on cross-country differences in earnings distribution with a set of facts on cross-country differences in on-the-job training. First, the share of establishments investing in workers' training increases with GDP per capita. Second, an establishment's likelihood of providing on-the-job training increases with establishment size. Finally, the wage premium paid by firms providing on-the-job training declines with development.

We interpret these facts through the lens of a model economy with heterogeneous workers and firms that can generate the observed cross-country differences in the labor earnings distribution. The model economy has three key ingredients. First, not all firms pay the same wage 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, as we document in the next section, firms also differ in how much on-the-job training they provide. The larger firms are much more likely to train their workers. Furthermore, even after controlling for size, firms that offer training pay higher wages. Finally, due to labor market frictions, matching between high-skilled workers and high-productivity-firms is not instantaneous (Lise et al., 2016).

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 lucky worker, who matches with high productivity firm that has a low cost of training, enjoys high wages and high wage growth. Other workers will be less fortunate and will work for firms with lower productivity. Of course, the higher a worker's human capital is, the higher her chances of being employed in a high productivity firm. An unemployed worker's skills depreciate. As workers and firms are matched and separated, and as firms' productivity and workers' skills evolve, 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 carefully replicates the observed firm size distribution, worker's wage profiles and training provision across different firms. It also produces the correct firm-size wage premium, despite not being part of the targeted moments.

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 so far focused on how misallocation affects cross-country differences in firm-size distribution and aggregate productivity.[1] We focus on how misallocation affects the inequality of earnings. Distortions have a direct impact on wages that firms offer. Firms that face distortions shrink and pay lower wages. Furthermore, with size-dependent or correlated distortions, smaller firms, which also pay lower wages, can expand, benefiting from lower overall labor demand. While the link between distortions and inequality is intuitive, it has not been studied in the existing literature.

Second, we assume that countries also differ in the extent of labor market frictions. Some have a more fluid labor market, and workers and firms match easily, while in others it takes a longer time to fill a vacancy or find a job. A less fluid labor market results in lower employment. Search frictions also affect the equilibrium wage distribution. In particular, longer time to fill a vacancy makes firms less willing to wait for the right workers, reducing assortative matching between firms and workers.

Frictions in the model are amplified by the endogenous training decision. On the one hand, frictions directly reduce the surplus in a given match, making firms less willing to incur training costs. On the other hand, frictions distort the sorting of workers to firms, further reducing firms' incentives to incur costly training to improve their workers' skills.

We choose the extent of correlated distortions and labor market frictions to match the average firm size and employment to population ratio in Mexico, a country with about one fourth of the UK's GDP per capita. In the benchmark economy, the average firm size is about 16 workers and about 77% of working age population is employed. Firms are much smaller in Mexico since on average they are made of only 11 workers. Furthermore, just 38% of working age population has a formal employment. We show that the UK-Mexico differences generated by the model economy fits a large set of cross-country facts on firms size distributions, life-cycle wage profile and training. In particular: i) Together with average firm size, the dispersion and skewness of the firms size distribution increase with GDP per capita (Hopenhayn, 2016; Bento and Restuccia, 2017; Poschke, 2018). ii) Wage-experience profile becomes steeper with GDP per capita (Lagakos et al., 2018). iii) On the other hand, wage-tenure profile becomes flatter with GDP per capita (Donovan et al., 2020). iv) Firm-size wage premium declines with GDP per capita in developing countries (Reed and Tran, 2019), in more developed countries (Lallemand et al., 2007) and it has been declining in U.S. since the 80's (Bloom et al., 2018). v) Formal employment increases with development (La Porta and Shleifer, 2014).

The model is able to generate cross-country differences in earnings inequality that are also in line with available evidence documented above: the standard deviation of log earnings increases but the mean-to-median earnings ratio declines with GDP per capita. In 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.

model economy, frictions and distortions affect workers who are in the middle of the skill distribution the most. These are the workers who are trained in the absence of distortions but not otherwise. These are also the ones who benefit most from better sorting. With higher training and better sorting (i.e., with higher GDP per capita), wages of these workers increase much more relative to workers at the bottom of the skill distribution who are not trained anyway and are only matched with low-productivity firms. Hence the p50-p10 wage ratio increases. What about the p90-p50 ratio? The mechanisms above operate similarly to workers at the top of the distribution, but the effect is more muted, making this ratio decline instead.

Finally, we find that job training provision explains around 12% of the differences in income per capita across countries, it accounts for between 12 and 18% of the differences in wage growth after 25 years of experience across countries, and up to 40% of the decline in the mean to median wage ratio observed over development.

While our focus on the interaction between misallocation and earnings inequality is novel, different elements of the model has 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 Lucas (1978) span-of-control model where managers can invest in their skills. Hence, distortions not only affect average firms size, but also the accumulation of managerial skills.

The link between labor market frictions and incentives of workers to invest in their skills has been studied by Engbom (2020). He shows that in countries where job-to-job mobility is more common, wages grow more over the life-cycle. He then builds a life-cycle model of on-the-job training and job-to-job transitions where fluid labor allocates workers to firms more efficiently and provide larger incentives for skill accumulation. Along similar lines, Ma et al. (2020) explore the role on 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.

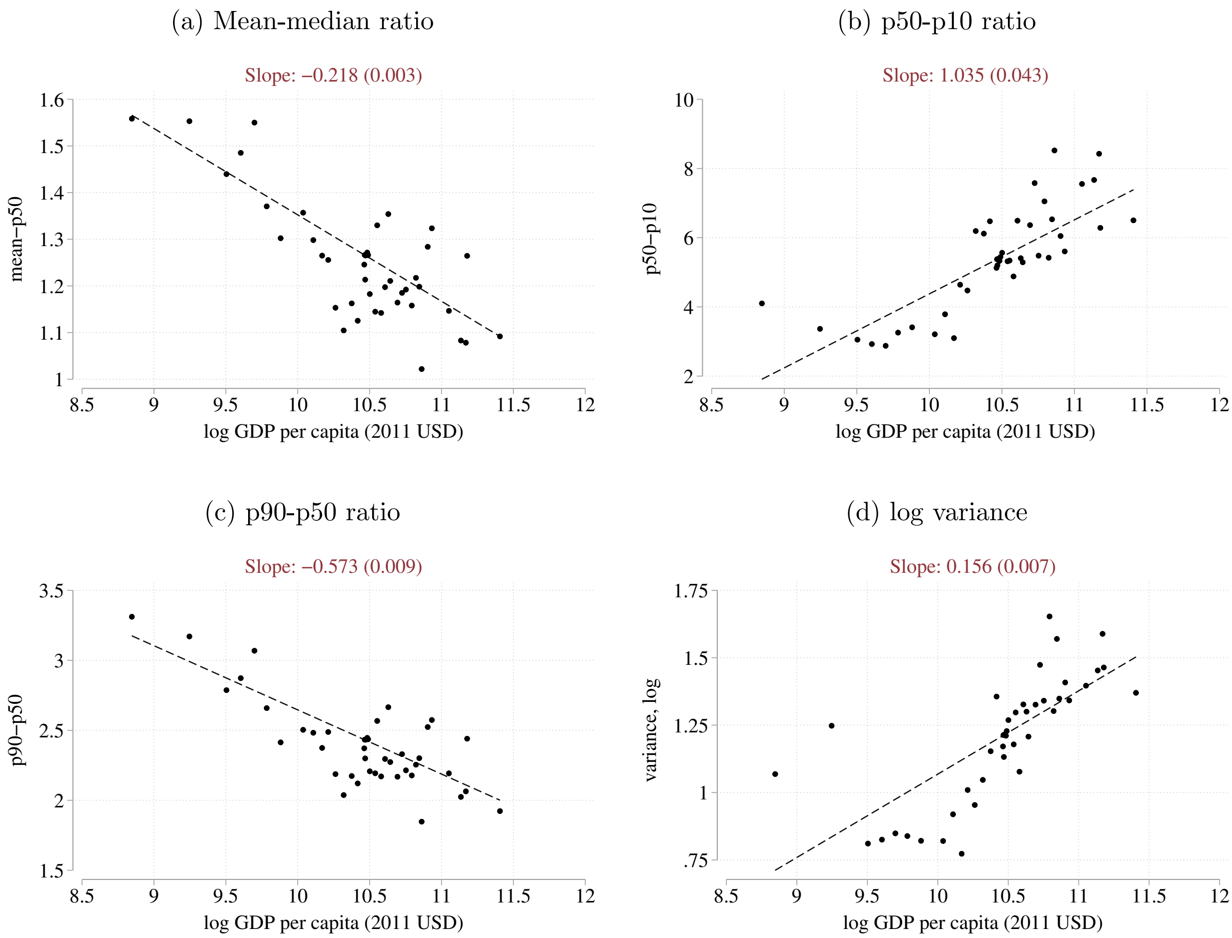
2 Cross-Country Facts

2.1 Earnings Distribution

In this section, we document how the distribution of wage and salary income varies with GDP per capita across countries. To this end, we use household surveys for 55 countries from 1981 to 2016. The poorest country in the sample is India in 1993 with a GDP per capita of 1845 in 2011 USD, while the richest is Noway in 2007 with a GDP per capita of 65000 USD. The primary sources for household surveys are IPUMS International, European Union Survey on

Income and Living Conditions (EU-SILC), and Luxembourg Income Study Database (LIS). We provide details in Appendix A.1 . In a nutshell, for each household survey, we restrict the sample to all individuals between ages 18 and 64 who are not students and report positive wage and salary income. We calculate total gross wages and salary income for wage and salary earners, which includes extra pay, tips, commissions, bonuses, piece-rate payments, occasional earnings, and study how its distribution changes as countries get richer. The share of wage and salary earners ranges from 0.43 (India in 1993) to 0.90 (Norway in 2007).

Figure 1: **Earnings inequality across countries**



Source: IPUMS, EU-SILC, LIS and author's calculations

Figure 1 shows our main findings. Each dot corresponds to the average values of the dependent variable (different inequality measures) for countries in a specific bin of GDP per capita, after conditioning on year fixed effects. Panel (a) shows that the mean-to-median ratio declines significantly as countries get richer. The mean-to-median ratio drops from around 1.5 for the poorest countries in the sample to about 1.1 for the richest ones. Hence, as countries get richer and the mean of the income distribution increases, the median workers gain even more. Panel (b) shows that the lower tail of the earnings distribution cannot catch up with the median—the 50-to-10 earnings ratio increases from around 3 in poor countries to around 8 in the richest ones. The same is also true for the upper tail, as it fails to grow as fast and the 90-to-50 ratio declines as countries get richer. However, the changes in the 90–to-10 ratio are more muted compared to changes in the 50-to-10 ratio. As the lower and upper tails of the earnings distribution get more spread with economic development; there is a significant increase in the variance of log earnings, as illustrated in Panel (d)

In Appendix A.1.3, we show that these findings' results are robust. They are observed for workers in different sectors and workers with and without a college degree. We also find these patterns when we restrict the samples to males, household heads, or workers in prime working ages (25 to 55). Finally, we look at different points of the income distribution, such as 90-to-60 and 40-to-10 or 80-to-50 and 50-to-20 ratios. We show that incomes in lower, and to a certain extent upper tail grow much slower than the incomes in the center of the distribution.

2.2 On-the-job Training

In this section we present a number of empirical facts linking firm-level training provision and country economy development. To this purpose, we exploit information from the World-Bank Enterprise Survey (WB-ES, henceforth) data set.[2] WB-ES is an ongoing project run by the World Bank to collect establishment-level data from a wide range of developing countries through face-to-face surveys. The dataset contains standardized variables for establishments in over 100 countries for at least one year since 2002, and it is representative of the population of establishments with at least five employees. Most importantly for the purpose of this paper, it contains information on firms' demographics (industry, age and number of employees) and training provision, defined as a dummy variable taking value one if firm i is providing a formal job training to her employees at time t, i.e.

$$\mathbf{1}_{it}^{\text{training}} = \begin{cases} 1 & \text{if formal training is provided by firm } i \text{ in year } t \text{ to} \\ & \text{permanent, full-time employees} \\ 0 & \text{otherwise} \end{cases}$$

We complement this data with information from the Eurostat Database on Education and Training. This dataset provides information on the participation of individuals in education and training activities, as well as on outcomes of education, for a sample of 30 middle- and high-income countries. Within this database, the Continuing Vocational Training Survey (CVTS, henceforth) collects information on enterprises' investment in the continuing vocational training of their staff. In particular, the dataset reports the share of firms providing job training, overall and broken by firm size, for each country. As for the WB-ES, the reference period for the provision of continuing vocational training is the calendar year. However, data are collected every five years and available for the period 2005-2015.

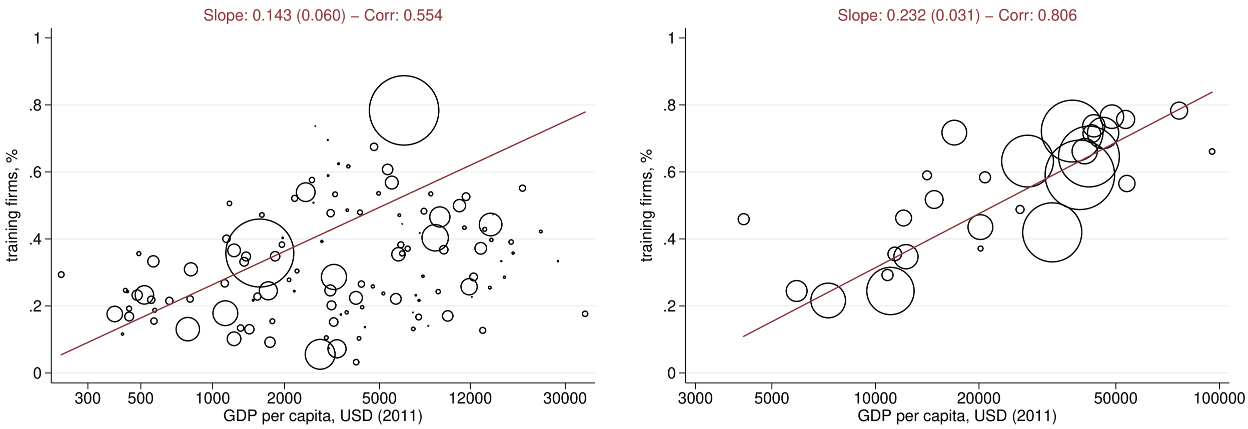
[2] Recent works using the same data set include, among the others, La Porta and Shleifer (2014) and Bento and Restuccia (2017).

Fact 1. Job training provision increases with GDP per capita.

Figures 2(a) and 2(b) report the average share of firms offering formal job training programs to their employees in each countries. In both figures, the measure of training provision is scattered over the country-average real GDP per capita. Each circle represents a country, with larger circles denoting larger population share of the country in the sample.

Figure 2: **Training provision across countries**

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



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

More developed countries have a larger share of firms investing in job training of their own employees. The correlation between the share of firms offering job training and the country log GDP per capita is equal to 0.55 in the WB-ES data. The slope coefficient from a regression of the country-specific share of training firms and log GDP per capita is 0.14 and is statistically significant at the one percent level. In terms of economic significance, the slope tells us that one log point higher GDP per capita is associated with a 14% percent more firms providing formal training to their employees. This pattern is robust across datasets. The correlation in the CV-TS data is larger and equal to 0.806 and the slope coefficients suggests that one log point higher GDP per capita is associated with 23% percent more firms offering training.

Fact 2. Job training provision increases with firm-size

How does firm training provision vary within each country? Table 1 reports the fraction of firms providing job training by different firm size categories. Training provision is significantly correlated with firm size. The share of firms investing in job training doubles as we move from a firm with less than 20 employees to a firm with more than a 500 employees. This pattern is consistent even when we split the sample into different macro regions (LAC, ME-AFR, ASIA, EU15, non-EU15), and across datasets.

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	44.79	29.18
20-49	54.31	31.99	33.63	38.48	20-49	56.00	39.36
50-249	66.94	41.31	47.02	46.47	50-249	71.67	52.82
250-449	81.13	56.86	47.32	56.65	250-449	86.29	67.64
≥500	92.12	68.45	52.28	68.88	500-999	88.00	78.45
					≥1000	96.36	88.73

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

Fact 3. Firms providing on-the-job training pay a wage premium, but the premium declines with development

To study the wage premium of firms providing on-the-job training and how it evolve over development, we consider the following cross-country firm-level regression:

$$\log w_{it} = \alpha \mathbf{1}_{it}^{\text{training}} + \beta \mathbf{1}_{it}^{\text{training}} \times \log \text{GDP}_{c(it)} + \mu_{c(i)} + \mu_t + \mu_{s(i)} + \gamma X_{it} + \epsilon_{it} \tag{1}$$

where w_{it} is the average wage paid by firm i at time t, $\text{GDP}_{c(it)}$ denotes the GDP per capita in country c where firm i operates at time t, $\mu_{c(i)}, \mu_t$ and $\mu_{s(i)}$ are country c, time t and sector s fixed effects, X_{it} includes various firm-level observables while ϵ_{it} is an error term. Nominal variables are deflates using the World Bank PPP index and expressed in 2011 USD. We estimate equation (1) by OLS including controls one by one.[3] Table 2 reports the estimates. Robust standard errors are reported in parenthesis. The data suggests two main correlations. First, firms providing OTJ training pay higher wages, that is about 50% more than those who don't. This is true also conditional on firm size and firm age (column 4, Table 2). Second, this premium declines significantly in richer countries. Doubling GDP per capita lowers the premium by about 4%.

In the Appendix we provide a number of additional evidence on job training over development. In particular, we show that within-firm training provision increase with development and within each country is higher in larger firms. Finally, we discuss few robustness checks to the results reported in Table 2 and show that the training pay premium is lower in richer countries even when we estimate it separately for countries with different income levels.

[3]Among the controls we consider 6 dummies for firm size (1-19,20-49,50-99,100-199,200-499,500+ employees), 8 dummies for firm age (1-2, 3-4, 5-9, 10-14, 15-19, 20-29, 30-49, 50+ y.o.), a dummy for the export status (exporter, non-exporter) and for ownership status (private, public, foreign)

Table 2: Firm level wage premium from training

	$\log w_{it}$				
	(1)	(2)	(3)	(4)	(5)
$\mathbf{1}_{it}^{\text{training}}$	0.611***	0.574***	0.575***	0.569***	0.554***
	(0.156)	(0.156)	(0.155)	(0.155)	(0.154)
$\mathbf{1}_{it}^{\text{training}} \times \log \text{GDP}_{c(it)}$	-0.0388**	-0.0382**	-0.0384**	-0.0384**	-0.0376**
	(0.0187)	(0.0187)	(0.0186)	(0.0185)	(0.0185)
Observations	88267	88267	88267	88267	88267
R^2	0.487	0.488	0.489	0.490	0.493
Country FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓
Firm size		✓	✓	✓	✓
Firm age			✓	✓	✓
Export status				✓	✓
Ownwership					✓

Note: * p<0.10, ** p<0.05, *** p<0.01. Source: World-Bank Enterprise Survey and author′s calculation.

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 they face a constant probability of death. Each worker enters the economy with a given level of human capital (skill). Each period workers can be employed or non-employed. Labor market frictions are represented by a matching function that maps the masses of non-employed workers and open vacancies into new matches. Once a match with a firm is formed, a worker's skills can grow due to learning-by-doing and on-the-job-training. In contrast, non-employment makes workers' skills diminish. Firms differ along three dimensions: productivity, cost of training, and the number of workers. Firms face size-dependent output distortions (wedges) that are correlated with their productivity levels.

3.1 Workers

Workers maximize the expected present value of their utility stream along their stochastic life cycle in the labor market:

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

where $r > 0$ is a discount rate while $\delta_w > 0$ is an exogenous probability of death/quit.

Workers are ex-ante heterogeneous in their initial level of human capital, $h_0 \in \mathcal{H} = \{h_0, h_1, ..., h_H\}$. Initial skills are distributed according to an exogenous probability density function, $\psi_h(h)$. Upon matching with a firm, workers can improve their abilities accumulating job experience and/or receiving on-the-job training, which cause human capital to jump one step in $\mathcal{H}$ with probabilities p^e and p^t, respectively.

Training is costly: training costs, ξ, is firm-specific and randomly distributed across firms with a probability density function ψ_ξ, defined on $\mathcal{E} \subset \mathcal{R}^+$. Human capital is fully portable between jobs. When job destruction occurs, workers retain fully their human capital, although each period of non-employment induces one-step skill depreciation 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 $z \in \mathcal{Z} \subset \mathcal{R}^+$, which is drawn upon entry and is distributed according to a probability density function, $\psi_z(z)$.

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

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

where A is a measure of aggregate productivity, while $g(z,i)$ is the amount produced by a match between a firm z and a worker i with human capital $h(i)$, defined as:

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

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

$$y = Ag(z,\bar{h})\ell \tag{2}$$

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

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

with

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

Linearity of the aggregate production function with respect to ℓ implies that each firm, independent of their productivity z would like to hire as many workers as possible, and, as it will become clear below, are only constrained in their hiring by matching frictions and adjustment costs. This make the problem tractable since a firm treats each of its workers separately. 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 loose a particular worker with probability δ_s or lose all workers and exit with probability δ_f.

3.3 Distortions

Firms are subject to output distortion. Distortions are modelled as Guner et al. (2018) and Bento and Restuccia (2017): each firm retains a fraction $1-\tau$ of its output, where τ is defined in the unit interval and assumed to depend on firm-level productivity z as follows

$$\tau(z) = 1 - z^{-\zeta} \quad \zeta > 0 \tag{3}$$

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

$$r(z,\ell,\bar{h}) = (1-\tau)y(z,\ell,\bar{h}) = (1-\tau)Ag(z,\bar{h})\ell = Az^{1-\zeta}\bar{h}\ell$$

or equivalently

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

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

3.4 The labor market

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. 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, $m(U,v)$ which implies a probability of filling a vacancy for firms, ϕ_f, equal to

$$\phi_f = \frac{m(U,v)}{v},$$

and a probability of finding a job for workers, ϕ_w, equal to

$$\phi_w = \phi_f \frac{v}{U}$$

Workers matched with a firm earn a wage equal to $w(z, \xi, h)$, 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 a share $b < 1$ of aggregate productivity A.

3.5 The problem of the worker

3.5.1 Value of unemployment

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

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

where $\mathbf{1}^h(z, \xi, h)$ is an indicator function for match formation, defined below. Hence, each period an unemployed worker do not match a with a firm with probability (1-ϕ_w) and is unemployed for the period. Unemployment can result in lower skills with probability p^d, while the value of being unemployed at the end of the period, $J^{u,h}(h)$, is given by:

$$J^{u,h}(h) = Ab + \frac{(1-\delta^w)}{1+r} J^u(h). \tag{4}$$

With probability ϕ_w the worker matches a firm and takes a random draw from $\psi_v(z, \xi)$, the distribution of z and ξ across firms that post vacancies. When a worker and firm are matched, if there is a positive surplus, employment takes places, in which case $\mathbf{1}^h(z, \xi, h) = 1$. Otherwise, a match is not formed, and the worker stays unemployed. The function $J^{e,h}(z, \xi, h)$, which is defined below, is the value of being employed at the end of the period in a firm with productivity z and training costs ξ.

3.5.2 Value of employment

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

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

where again $\mathbf{1}^h(z, \xi, h)$ is an indicator function for a positive surplus for a match between a type-(z, ξ) firm and type-h worker. If the surplus is positive, the value of employment is given by

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

where $w(z, \xi, h)$ is the wage rate and $p^h(z, \xi, h) = p^e + \mathbf{1}^t(z, \xi, h)p^t$ and $\mathbf{1}^t(z, \xi, h)$ 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,ξ) firm and a with ability of worker h. The value of this match accruing to the firm at the beginning of the period is equal to

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

with

$$V^h(z,\xi,h) = r(z,h) - w(z,\xi,h) +$$
$$\frac{(1-\delta_w)}{1+r}(1-\delta_f)(1-\delta_s)\left[\mathbf{1}^t(z,\xi,h)\xi + p^h(z,\xi,h)V(z,\xi,h+1) + (1-p^h(z,\xi,h))V(z,\xi,h)\right].$$

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

3.6.2 Vacancy posting decision

Firms choose the amount of vacancies $v(z,\xi)$ to maximize the total value of new hires subject to convex costs, $c(v)$, with $c(0) \geq 0, c'(\cdot) > 0, c'(0) = 0$ and $c''(\cdot) > 0$. Hence, in each period, the problem of a firm reads as

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

where ψ_h^u is the endogenous distribution of ability for unemployed workers. The firm posts $v(z,\xi)$ vacancies and gets in contact with $v(z,\xi)\phi_f$ unemployed workers, distributed across skill levels with a pdf $\psi_h^u(h)$. Each match with a positive surplus is valued as $V^h(z,\xi,h)$. The first order conditions for the optimal amount of vacancies v is then given by

$$c'(v(z,\xi)) = \phi_f \sum_{h\in\mathcal{H}} \mathbf{1}^h(z,\xi,h)V^h(z,\xi,h)\psi_h^u(h)$$

3.6.3 Entry decision

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 ξ from distributions ψ_z and ψ_ξ. Upon learning their type, firms decide to enter if they can cover the entry cost c^e. The entry condition reads as follows:

$$\Pi(z,\xi) \geq c^e, \tag{8}$$

where $\Pi(z, \xi)$ denotes the discounted sum of per-period aggregate profits of a firm with productivity z and training costs ξ, defined as follows

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

In equilibrium with a positive measure of firms, there exists a pair of productivity and costs (z^*, ξ^*) 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, ξ) for firms that decide to enter. A solution to this problem is a policy function for entry, $\mathbf{1}^e(z, \xi)$ defined as:

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

Notice 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 the its matches.

3.7 The surplus function

Because of search and matching frictions, there is a non-trivial gain in welfare, called match surplus, accruing to workers and firm when they became matched. The surplus for a given period is defined to be the difference in the payoffs of the worker and the firm depending on whether or not the match stays alive; that is, the threat points of the worker and the firm are those associated with separation. The value of the surplus $S(z, \xi, h)$, is equal to

$$S(z, \xi, h) = M(z, \xi, h) - J^{u,h}(h) \tag{10}$$

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

$$M(z, \xi, h) = J^e(z, \xi, h) + V(z, \xi, h) = \mathbf{1}^h(z, \xi, h)[J^{e,h}(z, \xi, h) + V^h(z, \xi, h)] + (1 - \mathbf{1}^h(z, \xi, h))J^{u,h}(h)$$

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

$$M(z, \xi, h) = \mathbf{1}^h(z, \xi, h)M^h(z, \xi, h) + (1 - \mathbf{1}^h(z, \xi, h))J^{u,h}(h) \tag{11}$$

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

$$\begin{aligned} M^h(z, \xi, h) = {} & g(z, h) + \frac{(1 - \delta_w)}{1 + r}(1 - (1 - \delta_f)(1 - \delta_s))J^{u,h}(h) \\ & + \frac{(1 - \delta_w)}{1 + r}(1 - \delta_f)(1 - \delta_s) \left[-\mathbf{1}^t(z, \xi, h)\xi + (1 - p^h(z, \xi, h))M(z, \xi, h) + p^h(z, \xi, h)M(z, \xi, h + 1) \right] \end{aligned} \tag{12}$$

Combining equations (12) and (4), we can write the surplus function as follows

$$S(z, \xi, h) = \max\{0, S^h(z, \xi, h)\} \tag{13}$$

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

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

or equivalently

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

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

$$\mathbf{1}^h(z,\xi,h) = \begin{cases} 1 & \text{if} \quad S^h(z,\xi,h) > 0 \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

where $S^h(z,\xi,h)$ is defined in equation (13).

3.8 Training decision

Each worker-firm pair decides to invest in training to maximize the joint value of the match. Workers and firms solve the following problem:

$$\mathbf{1}^t(z,\xi,h) = \arg\max_{\mathbf{1}^t\in\{0,1\}} \quad \mathbf{1}^t p^t[M(z,\xi,h+1) - M(z,\xi,h+1)] - \mathbf{1}^t\xi \tag{15}$$

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

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

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,

$$\max_{w(z,\xi,h)} \quad \left[J^{e,h}(z,\xi,h) - J^{u,h}(h)\right]^\beta V^h(z,\xi,h)^{1-\beta}, \tag{16}$$

where $\beta \in (0,1)$ is the workers' bargaining power. An interior solution to this problem is implicitly defined by the standard Nash splitting rule, i.e.

$$(1-\beta)\left[J^{e,h}(z,\xi,h) - J^{u,h}(h)\right] = \beta V^h(z,\xi,h).$$

This implies wages $w(z,\xi,h)$ are chosen such the worker's surplus equals a β share of the match surplus:

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

3.10 Equilibrium

A stationary recursive competitive equilibrium 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 (16);
3. *training*: training decision is the solution of the problem (15);
4. *market clearing*: goods and labor market are cleared;
5. *measure of entrants*: for all Borel sets $\mathcal{Z} \times \mathcal{E} \subset \mathcal{R}^+ \times \mathcal{R}^+$ it must be that

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

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

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

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

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

4 Bringing the Model to the Data

In this section, we estimate the model parameters to match a host of facts on firms and workers. The benchmark economy is estimated using data from the UK. The choice of UK reflects two considerations: First, it is a high-income economy, which we contrast with poorer economies in our counterfactuals. Second, the availability of data. on firm- and worker-level job training provision allow us to identify key parameters governing human capital accumulation and wage profiles.

4.1 Functional forms

We begin by specifying functional forms for the matching function $m(U, v)$, hiring costs $c(v)$, distribution of productivity, initial human capital and training costs, ψ_z, ψ_h and ψ_ξ.

To model contacts between firms and workers, we use the matching function of Den Haan et al. (2000):

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

The parameter η determines the elasticity of matches with respect to vacancies and unemployment. A larger value of η implies a more fluid labor market, as a give number of vacancies and unemployment map into a larger number of matches. The matching functions implies the a contact rate for workers and for firms 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. Hiring costs are modelled using a convex function following Cooper et al. (2007):

$$c(v) = \frac{\lambda_0}{\lambda_1} v^{\lambda_1}, \quad \lambda_0 > 0, \lambda_1 > 1,$$

where λ_0 is cost shifter, while λ_1 governs the cost convexity. This cost function implies the following amount of vacancy posted by a firm (z,ξ):

$$v(z,\xi) = \left(\frac{\phi_f}{\lambda_0}(1-\beta) \sum_{h \in \mathcal{H}} S(z,\xi,h) \mathbf{1}^h(z,\xi,h) \psi_h^u(h) \right)^{\frac{1}{\lambda_1 - 1}}$$

and the amount of new hires equal to $v(z,\xi)\phi_f \sum_{h\in\mathcal{H}} \mathbf{1}^h(z,\xi,h)\psi_h^u(h)$. We choose a mean-zero log-normal distribution to model initial human capital and firm-level productivity, i.e.

$$h \sim \log\mathcal{N}(0,\sigma_h), \quad \sigma_h > 0,$$

and

$$z \sim \log\mathcal{N}(0,\sigma_z), \quad \sigma_z > 0,$$

Finally, we choose a uniform distribution to model training costs:

$$\xi \sim \mathcal{U}(\underline{\xi}, \overline{\xi}), \quad \underline{\xi}, \overline{\xi} > 0.$$

The model is solved at quarterly frequency and the population is normalized to one.

4.2 Data

We estimate the model to the UK economy in the period 2010-2016. We combine information from two different datasets: the Five-Quarter Longitudinal Labour Force Survey (LFS, henceforth) dataset and the Employer Skills Survey (ESS, henceforth).

Table 3: Parameters directly calibrated

Parameters	Description	Value	Source/Targets
ζ	Correlated distortion	0	assumption
A	Productivity shifter	1	normalization
λ_0	Hiring costs, scalar	1	normalization
r	Interest rate	0.0033	annual return of 4%
δ_w	Workers retirement	0.0099	Life-span of 40 years, ages 22-62
δ_f	Firm exit	0.0253	annual exit rate of 10.50% (ONS)
η	Matching function	0.5416	estimated using GMM

4.2.1 The Labour Force Survey

The Five-Quarter Longitudinal LFS is a stratified 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 (expressed in 2010 sterlings), hours worked, and whether the surveyed worker has received OTJ training or not. 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.

4.2.2 The Employer Skills Survey

The Employer Skills Survey is a repeated cross-sectional firm-level biannual survey aimed at measuring the skills position and skills needs of UK employers. Each wave of the survey has a "Core" component, covering establishments demographics, strategy, recruitment and number of employees. The 2011 and 2013 waves have also a second facet, the "Investment in training" follow-up component, which covers the investment establishments make in training their staff. For the calibration, we restrict our focus to these two waves.

4.3 Parameters Set Without Solving the Model

Some parameters can be determined based on available evidence or set to their data counterparts a prior, without solving the model. To this end, we take UK as a distortion-free economy and fix the correlated distortion ζ to zero. The productivity shifter A, and the hiring costs λ_0 are also normalized to 1. We set the interest rate r to 0.0033 to match an annual return of 4%. Workers stay in the labor force for forty years, hence we set δ_w to

0.0099. We calibrate firm destruction rate δ_f to match an annual firm exit rate of 10.5%.[4] Finally, we estimate the efficiency of matching function, η through GMM. In particular, we minimize the following objective function:

$$\hat{x} = \arg \max_{\{x_0, x_1, x_2, x_3\}} \left[\left(\frac{1}{T} \sum_{t=1}^{T} Z_t' \epsilon_t(x) \right)' W_T \left(\frac{1}{T} \sum_{t=1}^{T} Z_t' \epsilon_t(x) \right) \right]$$

where $\epsilon_t(x)$ denotes the moment conditions, i.e.

$$\epsilon_t(x) = \left[h_t - \frac{u_t v_t}{(u_t^{x_0} + v_t^{x_0})^{\frac{1}{x_0}}} - \sum_{i=1}^{4} x_i \mathbf{1}_t^{\mathrm{q}=i} \right]$$

with h_t equal to the number of new hirings at time t, v_t the number of open vacancy and u_t the number of non-employed workers. We also remove seasonal effects by including dummies for quarters. $Z_t' = [u_t, v_t, u_{t-4}, v_{t-4}, \mathbf{1}_t^{\mathrm{q}=1}, \mathbf{1}_t^{\mathrm{q}=2}, \mathbf{1}_t^{\mathrm{q}=3}, \mathbf{1}_t^{\mathrm{q}=4}]$ is the vector of instruments, where we included the fourth lag for non-employment and active vacancies. We estimate η using a two-step GMM and obtain an estimate of η =0.5417 (s.e.=0.0134), which is significant at one percent level. [5]

4.4 Estimated Parameters

The remaining 13 parameters

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

are estimated by method of simulated moments to minimize the sum of square residuals between the model-implied values and data for 46 aggregate, worker- and firm- level targets. Let $\overline{d}(\vartheta)$ be a vector of $g \geq \dim[\vartheta]$ moment conditions (deviations between model and the data), defined as

$$\overline{d}(\vartheta) = \overline{m} - m(\vartheta)$$

where $\overline{m}$ is a vector of sample statistics while $m(\vartheta)$ is a vector of simulation-based statistics. The vector of parameters' values, $\hat{\vartheta}$ is the argument that minimize the following objective function,

$$\hat{\vartheta} = arg \min_{\vartheta \in \Theta} \quad \overline{d}(\vartheta)' \overline{d}(\vartheta) \tag{17}$$

While the model does not provide with a one-to-one map between parameters and target, there are a few parameters in ϑ that are guided by a specific targets. In particular, the exogenous separation rate, δ_s, determines the average job duration in the model. In the data jobs last on average about 5.36 years (Mumford and Smith, 2004). Moreover, the measure of potential entrants M_e maps, given all other parameters, maps into a rate of wage and salary employment equal to 76.58% (ONS).

[4] Data on firm exit rate come from the ONS Business Demographic Statistics for the period 2011-2018.

[5] Details of data and estimation are reported in Appendix C.

Figure 3: **Estimation fit**

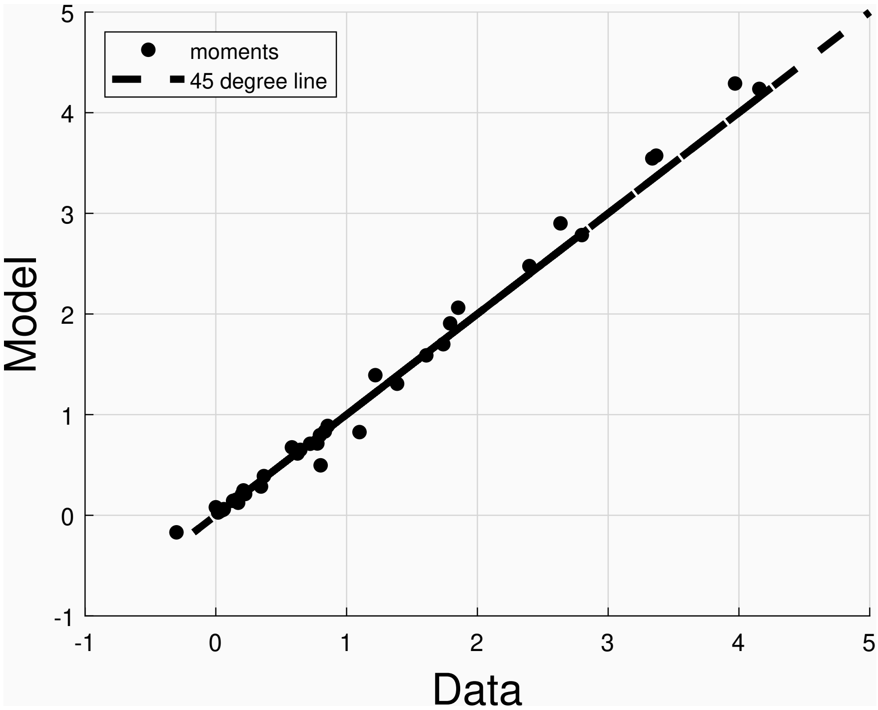


Figure 3 displays the estimation fit by scattering model-generated statistics against their data counterpart. The model does remarkably well in fitting the data with an average log-deviation of 0.087. Tables 4 lists all the moments used to obtain $\hat{\vartheta}$. The first column of Table 4 pertain to firm-level targets: i) number of firms and average firm size, ii) mean and standard deviation of log employment, iii) firm size distribution, and iv) firm-level training. The model does an excellent job generating a firm size distribution that is very much aligned with the data (Figure 4). Furthermore, it captures the fraction of firms that train their workers, about 65%, and the fact that larger firms are much more likely to train (among firms with more than 250 employees, almost 90% of firms train their workers). The second column of Table 4 presents the set of worker-level moments: i) wage levels and dispersions conditional on labor market experience and re-employment, ii) fraction of workers that receive training, iii) number of workers who receive training across different firm-size, iv) returns to training, and v) returns to tenure. The model again does a great job. Workers enter the labor market at an average wage that is about 50% below the mean. After 20 years in the labor market, their wages grow by about 10%. After unemployment, re-employed workers' wages are lower than the mean, both in the data and in the model. The dispersion of wages 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. Around 20% of workers get training, both in the data and the model. The returns to training 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).

Table 5 reports parameter estimates, $\hat{\vartheta}$, and their standard errors and 95% confidence

Table 4: Targeted Moments

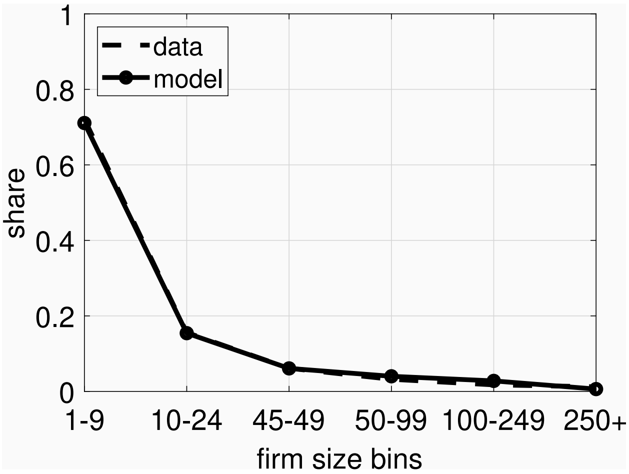
	Data	Model		Data	Model
Firm-level moments			*Worker wage distribution*		
Number of firms (over population)	0.171	0.158	Wage at entry, E[log($w_1/\bar{w}$)]	-0.5176	-0.50483
E(ℓ_t)	16.423	16.185	Wage after 20 y.o., E[log($w_{20}/\bar{w}$)]	0.1071	0.10928
E(log ℓ_t)	1.7393	1.6996	Wage at re-emp, E[log($w_R/\bar{w}$)]	-0.3010	-0.16948
std(log ℓ_t)	1.2198	1.3922	Dispersion at entry, sd[log w_1]	0.5818	0.6749
			Dispersion after 20 y.o., sd[log w_{20}]	0.7959	0.7954
Firm-size distribution			Dispersion at re-emp, sd[log w_R]	0.8335	0.8329
1-9 employees	72.12	71.08			
10-24 employees	15.95	15.43	*Trained workers*		
25-49 employees	6.12	6.09	$E\left(\frac{\#\text{trained workers}}{\#\text{workers}}\right)$	0.2114	0.24715
50-99 employees	3.21	4.00			
100-249 employees	1.73	2.78	*Worker-level training return*		
250+ employees	0.88	0.62	$\log w_{it} = \beta_1 \mathbf{1}^t_{it} + \epsilon_{it}$	0.1991	0.20773
Firm-size percentiles			*Job tenure return*		
10th percentile	1	1.083	tenure<3 months	1	1
25th percentile	3	2.285	tenure∈[3,12) months	1.0551	1.0539
40th percentile	4	3.696	tenure∈[12,24) months	1.1320	1.1434
50th percentile	5	4.900	tenure≥24 months	1.3675	1.3893
60th percentile	6	6.732			
75th percentile	11	11.893	*Workers trained within the firm*		
90th percentile	29	35.631	overall	9.121	7.953
95th percentile	53	72.979	1-9 employees	2.229	1.625
99th percentile	202	203.50	10-24 employees	6.381	7.850
			25-49 employees	13.951	18.054
Firm training provision			50-99 employees	28.150	34.395
$E\left(\frac{\#\textit{training firms}}{\#\textit{firms}}\right)$			100-249 employees	63.816	69.194
overall	0.646	0.650	250+ employees	225.70	186.17
1-49 employees	0.611	0.644			
20-249 employees	0.776	0.714	*Aggregate moments*		
250+ employees	0.855	0.888	Job duration	5.360	5.036
$E\left(\frac{\#\textit{trained employees}}{\#\textit{employees}}\right)$			Employment rate	0.776	0.788
overall	0.4588	0.4843			

intervals, computed using the Monte Carlos Markov Chain (MCMC) approach as in Chernozhukov and Hong (2003) together with 95% confidence intervals.[6] The estimated values imply a significant heterogeneity in training costs across firms; the maximum is about 15 times the minimum. Similar to the literature, the hiring costs are highly convex, as λ_1 is larger than 2. We estimate that worker's bargaining power is close to 0.5 (a value typically assumed in the literature). For each period of employment, there is about 22% chance that

[6] Confidence intervals are constructed taking the 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentiles of the simulated distribution

Figure 4: **Firm-level moments**

(a) Firm size distribution



(b) Firm size percentiles

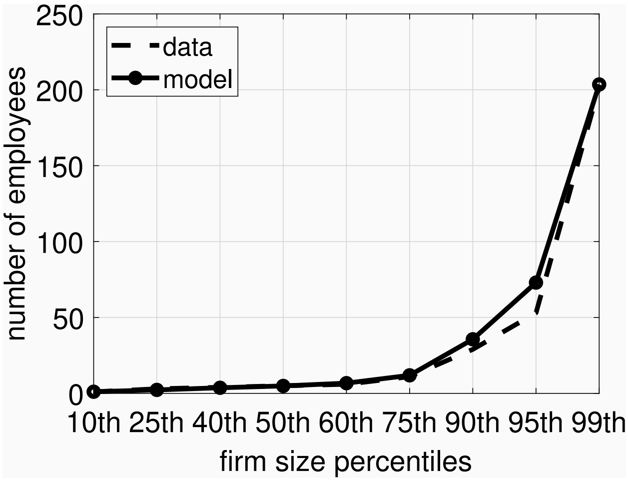
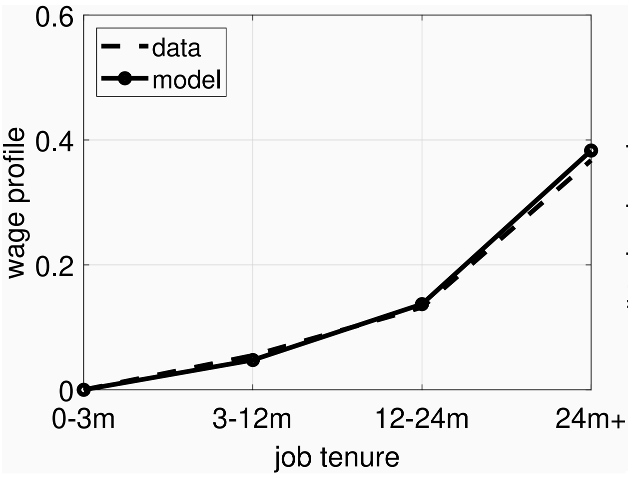
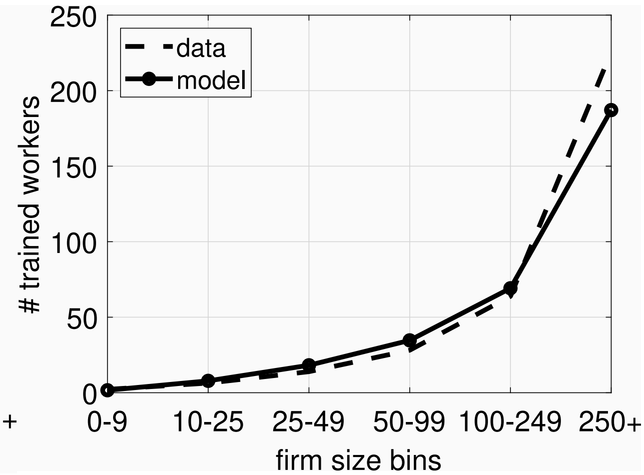


Figure 5: **Worker-level moments**

(a) Wage profiles over job tenure



(b) Trained workers over firm size



workers' skills can jump by one level, while for each period of unemployment they decline by one level with 43% probability. The training jumps skills by one level with a small probability, about 3%. These probabilities determine how wages in the model evolve by experience, tenure and training in the model. Finally, the model replicate quite closely the average number of workers who received training across different firm-size.

4.5 Non-targeted moments

Table 6 reports two sets of non-targeted moments. The first block of Table 6 compares data and model-based average wages for workers employed in firms differing by their number of employees, while the second block reports different measures of wage inequality.

Table 5: Estimated parameters

Parameters	Description	Estimates	St.Dev.	95% C.I.	
δ_s	Match separation	0.01235	0.0012	0.010065	0.014859
b	Home production	20.9430	1.8241	17.589	25.057
M_e	Measure of potential entrants	0.01272	0.0444	0.0008	0.1493
c_e	Entry cost	39.262	3.6646	33.186	47.613
$\underline{\xi}$	Training cost (lower bound)	1.7346	0.1569	1.4546	2.1103
$\overline{\xi}$	Training cost (upper bound)	26.668	2.3036	22.124	31.580
λ_1	Hiring costs, convexity	2.5246	0.1656	2.0633	2.7461
β	Bargaining power	0.4573	0.0416	0.3789	0.5497
σ_h	Initial human capital dispersion	1.1950	0.1110	0.9767	1.4246
σ_z	Firm-productivity dispersion	1.2044	0.1060	1.0178	1.4697
p^e	Experience jump	0.2233	0.0194	0.1836	0.2709
p^t	Training jump	0.0282	0.0030	0.0233	0.0347
p^d	Depreciation jump	0.4318	0.0400	0.3455	0.5142

The model generates a positive wage-size premium and the size premium in wages is of similar order of magnitude observed in the data. The model also replicates well the observed wage inequality in UK. It generates a value for the mean to median wage ratio remarkably close to the data, despite over-predicting log-wage dispersion. This is due to the fact that - while the model correctly captures the magnitude of the dispersion in the upper tail of the wage distribution - it generates a much more left-skewed lower tail.

5 A World with Larger Frictions

In this section, we move from our benchmark economy to an economy with larger frictions. We focus on two key frictions in the model: size-dependent distortions, captured by the parameter ζ and matching frictions captured by parameter η. In the benchmark economy ζ was set to zero, i.e. there were no distortions on firm's choice while η was calibrated as 0.542. How can we select alternative values for these two parameters?

We target a country with a lower GDP per-capita, Mexico, and asks the following question: keeping all other parameters fixed at their benchmark values, can we find values ζ and η that generate an economy that looks like Mexico? In particular, we choose values for ζ and η to match the average firm size and formal employment in Mexico. The average firm size is taken from Bento and Restuccia (2017), who report a value of 10.687 employees, roughly 7 employees less than the average firm in the U.K. The employment rate is taken from the

Table 6: Non-Targeted Moments

	Data	Model
Wage-size regression		
<10 employees	0	0
$\in [10, 25)$ employees	0.151	0.183
$\in [25, 50)$ employees	0.244	0.342
$\in [50, 250)$ employees	0.407	0.680
$\geq$250 employees	0.586	1.039
Wage inequality		
Log-wage dispersion, sd[$\log w_{it}$]	0.7788	0.9317
Mean-median wage ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$	1.2763	1.2067
90-50 pct. wage ratio, $\mathrm{p}^{90}[w_{it}]/\mathrm{p}^{50}[w_{it}]$	2.4100	2.5506
50-10 pct. wage ratio, $\mathrm{p}^{50}[w_{it}]/\mathrm{p}^{10}[w_{it}]$	2.9384	5.2618

ILO-STAT database, that reports a share of workers in the active population (25-65 years old) who is formally employed equal to 0.384, around 50 percentage lower than the U.K.

Table 7: Counterfactual parameters

	Parameters	
	Benchmark	Mexico
Joint Calibration (η, ζ)		
ζ, correlated distortions	0	0.09410
η, elasticity of the matching function	0.54167	0.42019
Only η		
ζ, correlated distortions	0	0
η, elasticity of the matching function	0.54167	0.36820
Only ζ		
ζ, correlated distortions	0	0.23410
η, elasticity of the matching function	0.54167	0.54167

Table 7 shows the results of this calibration exercise. To match the Mexican firm size and the employment to population ratio, the model requires a value of ζ around 0.094 (in contrast to 0 for the UK), while η is about 0.42 (in contrast to 0.54 for the UK). Table 8 shows how these two parameters are identified. The identification of η versus ζ is given by the differential magnitude in the effect exerted by these parameters on the employment rate. Suppose, for instance we select each parameter separately to match only one target, the average firm size, instead of choosing them simultaneously to match both targets. The calibrated values of ζ and η would be 0.37 and 0.23, respectively. While both parameters affect each targets symmetrically (reduction in η and an increase in ζ unambiguously reduce both firm size and employment), they do so with different strength: relative to the baseline in Table 4, for the same drop in average firm size (from 16.4 to 10.7 employees), reduction in labor market efficiency generate much lower drop in employment rate (from 78.84 to 44.97%) compared to increasing distortions (from 78.84 to 31.82%).

Table 8: Counterfactual targets

	Firm Size		Employment	
	Data	Model	Data	Model
Joint Calibration (η, ζ)	10.687	11.019	38.399	38.981
Only η	10.687	10.880	38.399	31.818
Only ζ	10.687	10.497	38.399	44.979

Figure 6 reports of the effects of the calibrated frictions - i.e. higher correlated distortions and lower matching efficiency - on firm productivity distribution (left panel) and labor market fluidity (right panel) respectively. On the one hand, correlated distortions redistribute resources from high-productivity to low-productivity firms while compressing the entire distribution on a lower productivity level relative to the baseline. At the same time, lower matching efficiency reduces in the number of contacts between firms and workers up to one third, relative to the baseline values.

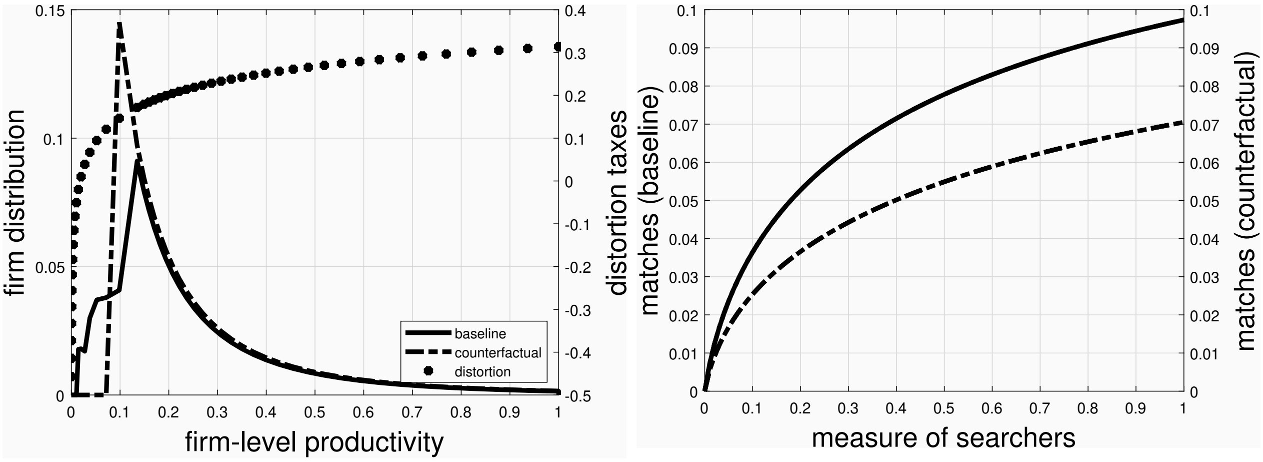
How does the economy with larger frictions looks like? Columns 1 and 2 of Table 9 report baseline and counterfactual outcomes with joint changes in frictions. A comparison between the benchmark economy and counterfactual world shows that - just by re-calibrating two parameters (η and ζ) to generate the right average firm size and the observed level of employment for a low-income economy - the model is able to account for how firms and worker-level outcomes change along development.

Let's start with firms:

1. Dispersion and skewness of the firm size distribution significantly decline, as documented in Poschke (2018)

2. The number of firms that offer training is lower, as documented in Section 2

Figure 6: **Correlated distortions and labor market efficiency**



3. The share of workers within each firm receiving training is lower, as reported in the Appendix A.4

4. The relation between the firms size and average firm-level wage strengthens, as documented by Reed and Tran (2019) and Lallemand et al. (2007).

Turning to the workers:

1. Experience-wage profiles becomes becomes less steep, as documented by Lagakos et al. (2018)

2. Tenure-wage profiles becomes steeper, as documented by Donovan et al. (2020)

3. Firm-level training pay premium increases, as documented in Section 2

While we keep the aggregate productivity A at its benchmark value of 1, the model is also able to generate a decline in average per capita income and average wages, which are very close to the data (per capita income declines by almost 80% while average wages decline by about 60%).

Table 9: Counterfactual outcomes

	Baseline	Counterfactual			Data
		Joint (η, ζ)	Only η	Only ζ	
Elasticity of matching function: η	0.54167	0.42019	0.36820	0.54167	-
Distortion correlation: ζ	0	0.09410	0	0.23410	-
Firm-level moments					
Number of firms (over active pop.)	0.158	0.230	0.260	0.316	0.1902
$\mathrm{E}(\ell_t)$	16.1854	11.0192	10.4974	10.8803	10.6870
$\mathrm{std}(\ell_t)$	37.1581	12.8896	12.3514	11.3534	-
$\mathrm{skew}(\ell_t)$	5.1774	2.6723	2.7730	2.5773	-
Firm training provision					
$\mathrm{E}\left(\frac{\#\text{training firms}}{\#\text{firms}}\right)$, %	65.02	22.20	30.46	12.29	37.70
E (#trained employees)	7.953	1.913	2.527	1.025	-
$\mathrm{E}\left(\frac{\#\text{trained employees}}{\#\text{employees}}\right)$, %	48.43	14.61	20.11	9.08	-
Wage profile over experience					
Wage at entry, $\mathrm{E}[\log(w_1/\bar{w})]$	-0.5048	-0.2600	-0.2536	-0.2449	-
Wage growth, $\mathrm{E}[\log(w_{10}/w_1)]$	0.2732	0.1665	0.1247	0.1528	-
Wage growth, $\mathrm{E}[\log(w_{20}/w_1)]$	0.6141	0.3264	0.3642	0.3278	0.4111
Wage growth, $\mathrm{E}[\log(w_{25}/w_1)]$	0.8013	0.4244	0.3971	0.4010	0.4402
Wage profile over job tenure					
Wage at tenure<3 months	1	1	1	1	-
Wage at tenure∈[3,12) months	0.0539	0.0668	0.0597	0.0520	-
Wage at tenure∈[12,24) months	0.1434	0.1724	0.1786	0.1464	-
Wage at tenure≥24 months	0.3893	0.4371	0.4648	0.3903	-
Worker-level large-firm wage premium					
$\log w_{it} = \beta_1 \mathbf{1}_{it}^{100+} + \epsilon_{it}$	0.4904	0.5481	0.4506	0.5145	-
Worker-level training wage premium					
$\log w_{it} = \beta_1 \mathbf{1}_{it}^{t} + \epsilon_{it}$	0.2077	0.2715	0.4208	0.0989	-
Firm-level training wage premium					
$\log w_{jt} = \beta_1 \mathbf{1}_{jt}^{t} + \epsilon_{jt}$	0.0397	0.0564	0.0720	0.0239	-
Aggregates					
Non-employment rate (25-65 y.o.)	0.2116	0.6102	0.5502	0.6817	0.6160
Income per capita	1	0.2381	0.3221	0.1712	0.2253
Average wage	1	0.4840	0.5885	0.3593	0.4404
Job finding rate, ϕ_w	0.2585	0.0914	0.0901	0.1012	-
Job filling rate, ϕ_f	0.2984	0.3382	0.2362	0.5324	-

Columns 3 and 4 of Table 9 report counterfactual outcomes separately for one-by-one changes in matching elasticity and correlated distortions. While both frictions are needed to match differences in average firm size and employment rate over development, and - taken separately - both frictions are consistent with the majority of evidence, only a reduction in labor market fluidity is able to generate larger tenure-wage profile and larger firm-size pay premium.

5.1 Matching and training sets

Figure 7: **Equilibrium matching sets**

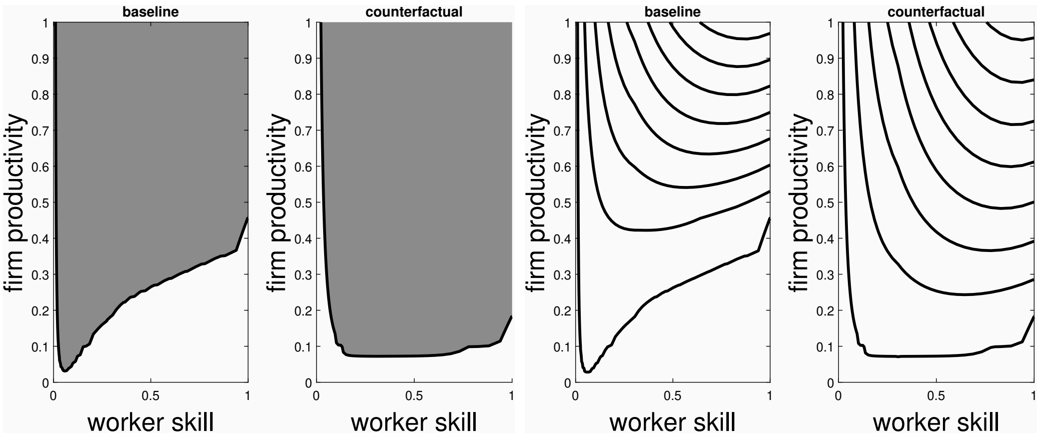
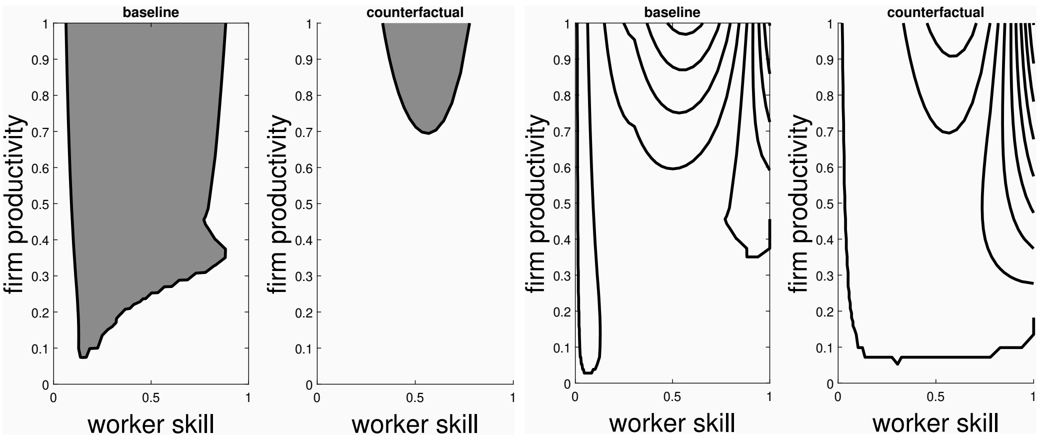


Figure 8: **Equilibrium training sets**



To understand the mechanisms behind these changes, we next ask how employment formation and training provision vary between the benchmark and the counterfactual economy. The two figures on the top row of Figure 7 shows the matching set between workers and firms. For each sub-figure horizontal axis is the productivity of workers and the vertical axis is the productivity of firms. The shaded areas are the matches with positive surplus. In the benchmark economy, only very skilled workers are matched with high productivity

Table 10: Implications for wage inequality

	Baseline	Counterfactual		
		Joint (η, ζ)	Only η	Only ζ
Log-wage dispersion, sd[log w_{it}]	0.9317	0.7075	0.8125	0.5625
Coefficient of variation, sd[w_{it}]/E[w_{it}]	0.7359	0.7629	0.8888	0.6054
Mean-median wage ratio, E[w_{it}]/p^{50}[w_{it}]	1.2067	1.3793	1.5554	1.2511
90-50 pct. wage ratio, p^{90}[w_{it}]/p^{50}[w_{it}]	2.5506	3.0026	3.6385	2.4522
50-10 pct. wage ratio, p^{50}[w_{it}]/p^{10}[w_{it}]	5.2618	2.3813	2.5679	1.9487

firms. Hence, there is positive assortative matching between workers ad firms. This is not the case in the counterfactual where very unproductive firms do not have workers and a higher productivity does not imply a more skilled workforce for the firm.

Why is this happening? On the one hand, correlated distortions reduce wage differential between firms, making them more similar as potential employees for the workers. On the other hand, larger labor market frictions imply that it is costly for high productivity firms to wait to match with better workers. The two figures in Table 8 show the training sets, i.e. the combination of firm- and worker-level productivity at which a particular match results in training for workers. Training provision shrinks dramatically in the counterfactual economy. This reduction happens to realize at low-productivity firms for workers in the middle of the skill distribution.

Figure 9: **Productivity, skills and wages**

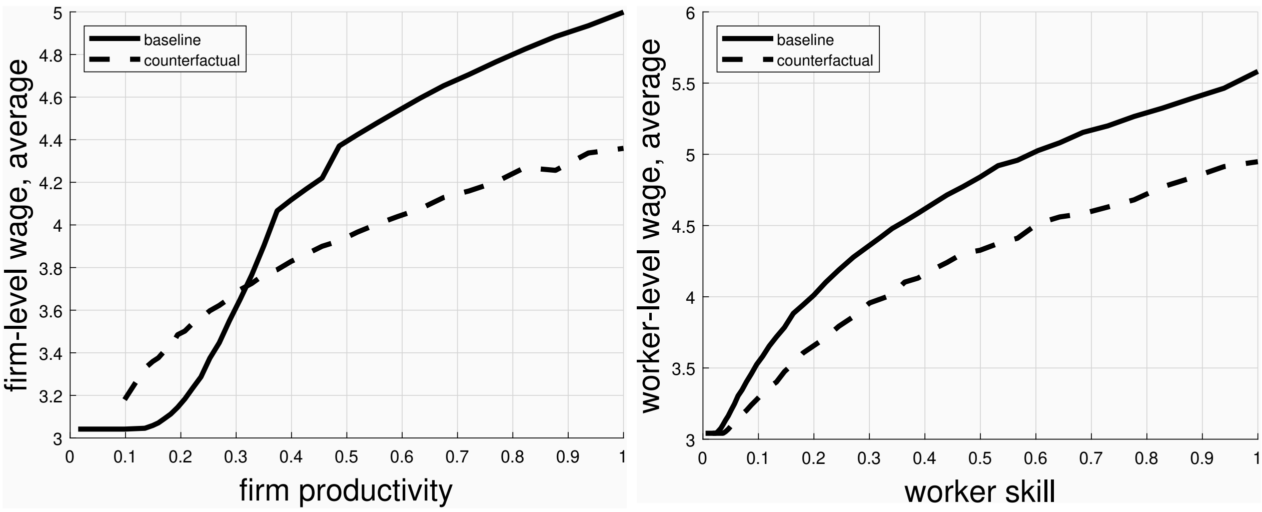


Table 11: Fixed OTJ training policy

	Baseline	Counterfactual		Explained
Elasticity of matching function: η	0.54167	0.54167	0.42019	-
Distortion correlation: ζ	0	0	0.09410	-
Training policy: $\mathbf{1}^t(z, \xi, h)$	baseline	counterfactual	counterfactual	-
	Firm-level moments			
$\mathrm{E}(\ell_t)$	16.1854	21.5297	11.0192	
$\mathrm{std}(\ell_t)$	37.1581	44.4817	12.8896	
$\mathrm{skew}(\ell_t)$	5.1774	4.6435	2.6723	
	Training provision			
$\mathrm{E}\left(\frac{\#\text{training firms}}{\#\text{firms}}\right)$, %	64.0196	23.9257	22.2029	95.88%
	Wage profile over experience			
Wage growth, $\mathrm{E}[\log(w_{20}/\bar{w}_1)]$	0.6141	0.5935	0.3264	7.16%
Wage growth, $\mathrm{E}[\log(w_{25}/\bar{w}_1)]$	0.8013	0.7500	0.4244	13.61%
	Aggregates			
Non-employment rate (25-65 y.o)	0.2116	0.2344	0.6102	5.72%
Income per capita	1	0.9106	0.2381	11.73%
Average wage	1	0.9573	0.4840	8.28%
	Wage inequality			
Log-wage dispersion, $\mathrm{sd}[\log w_{it}]$	0.9317	0.9168	0.7075	6.65%
Mean-median wage ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$	1.2067	1.2254	1.3793	10.83%

5.2 Wage Inequality

How does the wage inequality looks like in the counterfactual economy? Table 10 reports several measure of wage inequality across three different counterfactual economies. The model is able to generate two facts on wage inequality we documented:

1. the dispersion of log-wages increases with development, while
2. the mean-median ratio declines

Figure 9 shows how wages depend on firm and worker productivity. Assortative matching in the baseline economy results in steeper firm-level wage policy, where high- (low-) productivity firms pay relatively higher (lower) wage compared to an economy with higher frictions.

5.3 Role of OTJ training

What is the role of training in the model economy? To answer this question, we conduct two main experiments. In the first one, we revisit the baseline economy but constrain firm's training decisions at their counterfactual policy rules.

Specifically, if a match between a type-h worker and type-(z, ξ) implies training for worker in the counterfactual economy, we impose the same policy in the baseline economy too, even if it is not profitable for the match. The column 2 in Table 11 reports the results.

With the training decision rules fixed at the counterfactual economy, a lower fraction of firms train their workers compared to the baseline (23.92% versus 64.02%). As a result, income per capita and average wages are lower too, part of it because of a lower wage growth after during workers life-cycle. With fixed training policy, change in wage inequality induced by lower frictions is muted. Focusing on mean-median ratio, the endogenous training decisions explain about 11% of changes between the benchmark and the counterfactual.

Table 12: A world without OTJ training

	Baseline	Counterfactual	Baseline	Counterfactual	Explained
	with OTJ training		w/o OTJ training		
Elasticity of matching function: η	0.54167	0.42019	0.54167	0.42019	-
Distortion correlation: ζ	0	0.09410	0	0.09410	-
	Firm-level moments				
$\mathrm{E}(\ell_t)$	16.1854	11.0192	16.1801	7.6523	
$\mathrm{std}(\ell_t)$	37.1581	12.8896	32.1430	9.1213	
$\mathrm{skew}(\ell_t)$	5.1774	2.6723	4.3362	2.7851	
	Training provision				
$\mathrm{E}\left(\frac{\#\text{training firms}}{\#\text{firms}}\right)$, %	64.0196	22.2029	-	-	-
	Wage profile over experience				
Wage growth, $\mathrm{E}[\log(w_{20}/\bar{w}_1)]$	0.6141	0.3264	0.5872	0.3356	12.55%
Wage growth, $\mathrm{E}[\log(w_{25}/\bar{w}_1)]$	0.8013	0.4244	0.7308	0.4230	18.33%
	Aggregates				
Non-employment rate (25-65 y.o.)	0.2116	0.6102	0.2028	0.5093	23.11%
Income per capita	1	0.2381	1	0.3315	12.26%
Average wage	1	0.4840	1	0.5509	12.97%
	Wage inequality				
Log-wage dispersion, $\mathrm{sd}[\log w_{it}]$	0.8317	0.7075	0.8808	0.6962	17.66%
Mean-median wage ratio, $\mathrm{E}[w_{it}]/\mathrm{p}^{50}[w_{it}]$	1.2067	1.3793	1.2795	1.3848	38.99%

As an alternative exercise, we re-calibrate a baseline economy assuming no job-training can be provided. We use the targets in Tables 4, except the ones related to training. We then conduct exactly the same counterfactual exercise, i.e., we impose $\eta = 0.42$ and $\zeta = 0.094$

on benchmark economy that does not have any training. Table 12 reports the results. The changes in inequality are once again more muted. Now the standard deviation of log wages increase only by 18% points (0.88 vs. 0.70), instead of 22% (0.93 vs. 0.71), accounting for 18% of the total change. Similarly, the change in the mean-median ratio is also smaller. For the benchmark calibration the mean-median age ratio increased from 1.21 to 1.38, while the increase is from 1.28 to 1.38. Hence, training account for about 39% of the change in the mean-median wage ratio.

5.4 Alternative mechanisms

Finally, we explore whether two alternative mechanisms operating through the labor market can account for the cross-country pattern of development.

Table 13: Alternative mechanisms

	Baseline	Counterfactual		
	U.K.	Mexico	Mexico	Mexico
Elasticity of matching function: η	0.54167	0.42019	0.54167	0.54167
Distortion correlation: ζ	0	0.09410	0	0
Separation rate: δ_s, %	1.235	1.235	2.537	1.235
Firm exit rate: δ_f, %	2.526	2.526	2.526	2.984
Average firm size	16.423	10.687	13.410	15.936
Employment rate	0.7758	0.3840	0.7068	0.7515
Wage growth, $\mathrm{E}[\log(w_{25}/w_1)]$	0.7753	0.4402	0.7100	0.8000
Training provision, overall %	64.50	37.70	58.79	59.19
Income per capita	1	0.2253	0.8134	0.9165
Log-wage dispersion, $\mathrm{sd}[\log w_{it}]$	0.8317	0.7075	0.8982	0.9127
Wage inequality (mean-median ratio)	1.2067	1.3793	1.2927	1.2325

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 formal jobs) in Mexico equal to 5 years, as documented by de la Parra (2016). 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 counterfacutal yearly firm exit rate for the entire business industry in Mexico equal to 0.125 as reported by Bartelsman et al. (2009). We compare baseline and counterfactual outcomes in Table 13, column 4. A reduction in firm exit rate can qualitatively explain only some features observed over development. Among the others, lower firm exit rate compresses wage growth after 25 years, mainly because of a higher average wage at entry compared to the counterfactual.

6 Accounting for Cross-Country Differences

In this section, we reproduce the counterfactual exercise from the previous section for a larger set of countries. We select four additional countries, Brazil, Peru, Vietnam and India, which cover a wider range of income per capita. For each country, we select η and ζ so that we match the average firm size and formal employment rate for a given country. The average firm size in these countries are even smaller than the one in Mexico. The average firm in India, for example, has only about 3.1 workers. The formal employment to population ratio is also much lower. While 43% of Brazilian population older than 25 y.o. were employed in formal firms, the number is 20% in Vietnam and only 9.1% in India. Hence, the counterfactual economies require larger values for ζ and lower ones for η than the ones implied for Brazil and Mexico.

Table 14: Counterfactual parameters

	Baseline	Counterfactual				
	U.K.	Brazil	Mexico	Peru	Vietnam	India
Elasticity of matching function: η	0.54167	0.43021	0.42019	0.40022	0.37186	0.31919
Correlated distortion: ζ	0	0.08408	0.09410	0.10090	0.10632	0.11410

Table 15: Model prediction

	Baseline	Counterfactual				
	U.K.	Brazil	Mexico	Peru	Vietnam	India
Average firm size	16.1854	15.1820	11.0192	8.8029	5.4271	2.7139
Employment rate (formal)	0.7840	0.4367	0.3898	0.3614	0.2448	0.1188
Wage growth, E[log(w_{25}/w_1)]	0.8013	0.4251	0.4244	0.38082	0.3088	0.3661
Training provision, overall %	65.02	26.129	22.20	18.84	14.77	10.28
Income per capita	1	0.2665	0.2381	0.2131	0.1644	0.1249
Log-wage dispersion, sd[log w_{it}]	0.9317	0.7299	0.7075	0.6809	0.6515	0.6153
Wage inequality (mean-median ratio)	1.2067	1.4235	1.3793	1.4148	1.4111	1.4322

Table 14 shows the results. The model does a very good job generating lower income per capita in these economies resulting from larger distortions. Workers in these poorer countries earn much less and experience much lower wage growth. They are also much less likely to receive training. Furthermore, the model is able to generate changes in inequality that are very much in line with evidence proposed in the literature.

7 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 that distortions, either modelled as explicit policies or implicit taxes on firms' output or input prices, affect firms' decisions on how much to produce. If distortions are correlated with firms' productivity, productive firms end up smaller than they should while low productivity forms expand. Hence, the firm size distribution shifts to the left, which results in smaller firms and lower incomes. However, this literature has been silent on how misallocation might affect earnings distribution since they often are embedded within competitive labor markets. Yet, there is growing evidence that firm-level drivers are fundamental to understand earnings inequality.

On the other hand, 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.

In this paper, we combine these two approaches to study how misallocation affects earnings inequality. The marriage seems to be a happy one. The benchmark economy can speak 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 earnings distribution changes with economic development. The data shows that the distribution of earnings changes with development in a particular way: while the standard deviation of log earnings increases with development, the mean-to-median ratio declines. We show that the model can replicate this pattern in the data if a poor country is identified as one with higher distortions along two dimensions: higher implicit taxes on firms' output that are correlated with firm productivity and lower ability of the labor market to match unemployed workers and open vacancies.

References

Abowd, J. M., Kramarz, F., and Margolis, D. N. (1999). High wage workers and high wage firms. *Econometrica*, 67(2):251–333.

Alpman, A., Balestra, C., and Murtin, F. (2018). Unveiling the monetary value of non-market activities using experienced well-being and time use surveys. In *35th IARIW General conference. http://www. iariw. org/copenhagen/alpman. pdf.*

Andrews, I., Gentzkow, M., and Shapiro, J. M. (2017). Measuring the sensitivity of parameter estimates to estimation moments. *The Quarterly Journal of Economics*, 132(4):1553–1592.

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.

Bloom, N., Guvenen, F., Smith, B. S., Song, J., and von Wachter, T. (2018). The disappearing large-firm wage premium. In *AEA Papers and Proceedings*, volume 108, pages 317–22.

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.

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. and Venkateswaran, V. (2019). The sources of capital misallocation. *American Economic Review*, 109(7):2531–67.

de la Parra, B. S. (2016). Formal firms, informal workers and household labor supply in mexico. *Job Market Paper, University of Chicago.*

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).

Engbom, N. (2020). Labor market fluidity and human capital accumulation. Technical report, working paper. Retrieved from https://www. niklasengbom. com/wp-content ....

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.

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.

Hopenhayn, H. A. (2016). Firm size and development. *Economía*, 17(1):27–49.

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.

Jørgensen, T. (2020). Sensitivity to calibrated parameters.

La Porta, R. and Shleifer, A. (2014). Informality and development. *Journal of Economic Perspectives*, 28(3):109–26.

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.

Lallemand, T., Plasman, R., and Rycx, F. (2007). The establishment-size wage premium: evidence from european countries. *Empirica*, 34(5):427–451.

Lise, J., Meghir, C., and Robin, J.-M. (2016). Matching, sorting and wages. *Review of Economic Dynamics*, 19:63–87.

Lucas, R. E. (1978). On the size distribution of business firms. *The Bell Journal of Economics*, pages 508–523.

Ma, X., Nakab, A., and Vidart, D. (2020). Human capital investment and development: The role of on-the-job training.

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.

Mumford, K. and Smith, P. N. (2004). Job tenure in britain: Employee characteristics versus workplace effects. *Economica*, 71(282):275–297.

Poschke, M. (2018). The firm size distribution across countries and skill-biased change in entrepreneurial technology. *American Economic Journal: Macroeconomics*, 10(3):1–41.

Reed, T. and Tran, T. T. (2019). *The Large-Firm Wage Premium in Developing Economies*. The World Bank.

Restuccia, D. and Rogerson, R. (2008). Policy distortions and aggregate productivity with heterogeneous establishments. *Review of Economic dynamics*, 11(4):707–720.

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 wage 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 41 countries spanning one or multiple years. Table 16 reports the list of country, year and source.

Table 16: Data source

Country	Year	Source	Country	Year	Source
Austria	2005, 2010	EU-SILC	Latvia	2006, 2010	EU-SILC
Belgium	2005, 2009	EU-SILC	Lithuania	2006, 2009	EU-SILC
Bulgaria	2007	EU-SILC	Luxembourg	2005, 2010	EU-SILC
Croatia	2010	EU-SILC	Malta	2007, 2010	EU-SILC
Cyprus	2005, 2010	EU-SILC	Netherlands	2006, 2010	EU-SILC
Czech republic	2006, 2009	EU-SILC	Norway	2005, 2010	EU-SILC
Denmark	2005, 2009	EU-SILC	Panama	1970	IPUMS
Dominican Republic	1981	IPUMS	Poland	2005, 2009	EU-SILC
Estonia	2005, 2010	EU-SILC	Portugal	2005, 2010	EU-SILC
Finland	2005, 2009	EU-SILC	Puerto Rico	1990, 2000, 2005	IPUMS
France	2005, 2010	EU-SILC	Romania	2007, 2009	EU-SILC
Germany	2005, 2009	EU-SILC	Slovakia	2006, 2009	EU-SILC
Greece	2005, 2009	EU-SILC	Slovenia	2006, 2009	EU-SILC
Hungary	2006, 2010	EU-SILC	Spain	2005, 2009	EU-SILC
Iceland	2005, 2010	EU-SILC	Sweden	2005, 2009	EU-SILC
Israel	1995	IPUMS	Switzerland	2007, 2009	EU-SILC
Italy	2005 2009	EU-SILC	Trinidad and Tobago	2000	IPUMS
India	1993, 1999	IPUMS	USA	2000, 2005, 2010	IPUMS
Indonesia	1976, 1995	IPUMS	Uruguay	2006	IPUMS
Ireland	2005, 2009	EU-SILC	United Kingdom	2005, 2009	EU-SILC
Jamaica	1981, 1991, 2001	IPUMS			

A.1.1 IPUMS-International

IPUMS-International collects cross-country census microdata on individual demographics, labour market outcomes and income among the others. For each country in Table 16 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)[7] When information on employment status is missing, we use information on the average number of hours worked per week overall (variable HRSWORK1)

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

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.[8] 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-SILC

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

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

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 Earnings inequality across countries

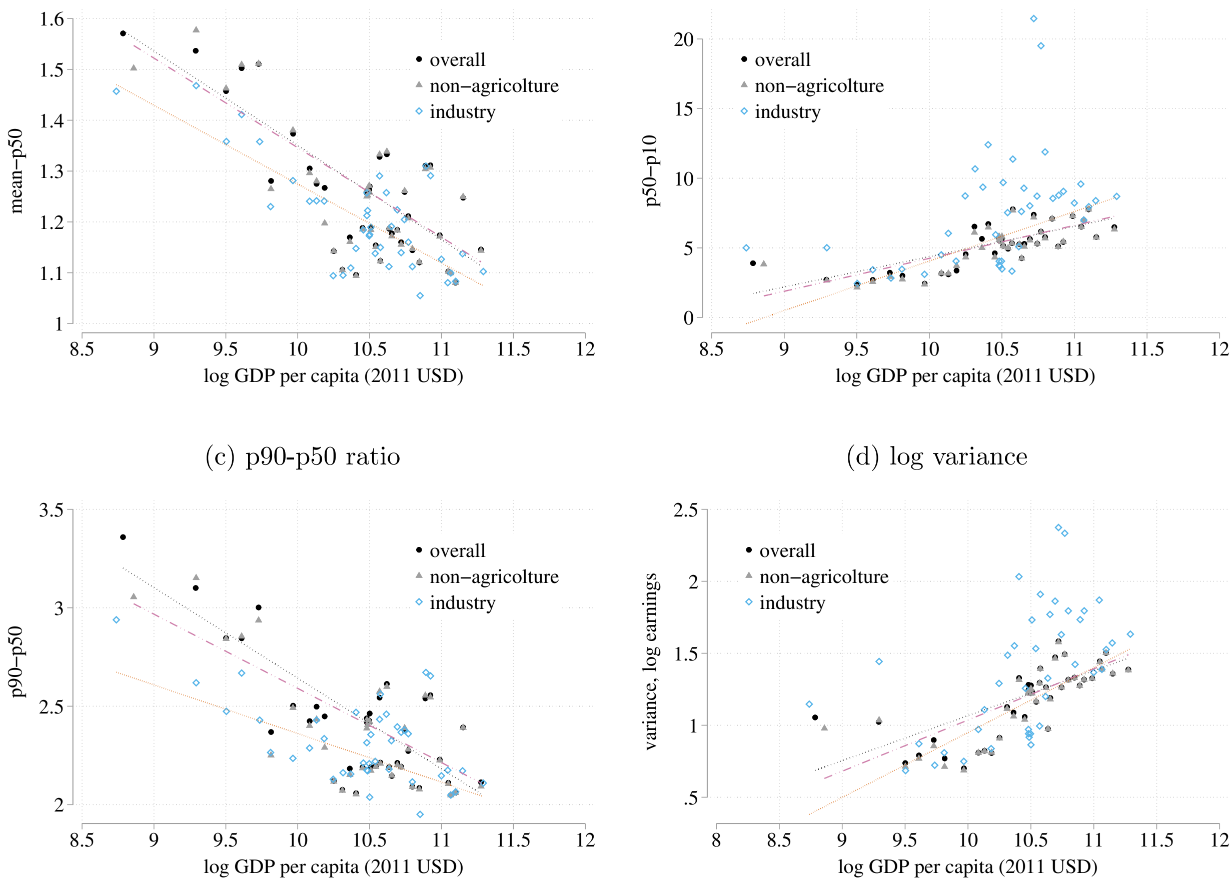
Figure 10: **Earnings inequality across countries, by sectors**

(a) Mean-median ratio

(b) p50-p10 ratio

(c) p90-p50 ratio

(d) log variance



Source: IPUMS, EU-SILC, LIS and author's calculations

Figure 11: **Earnings inequality across countries, by education**

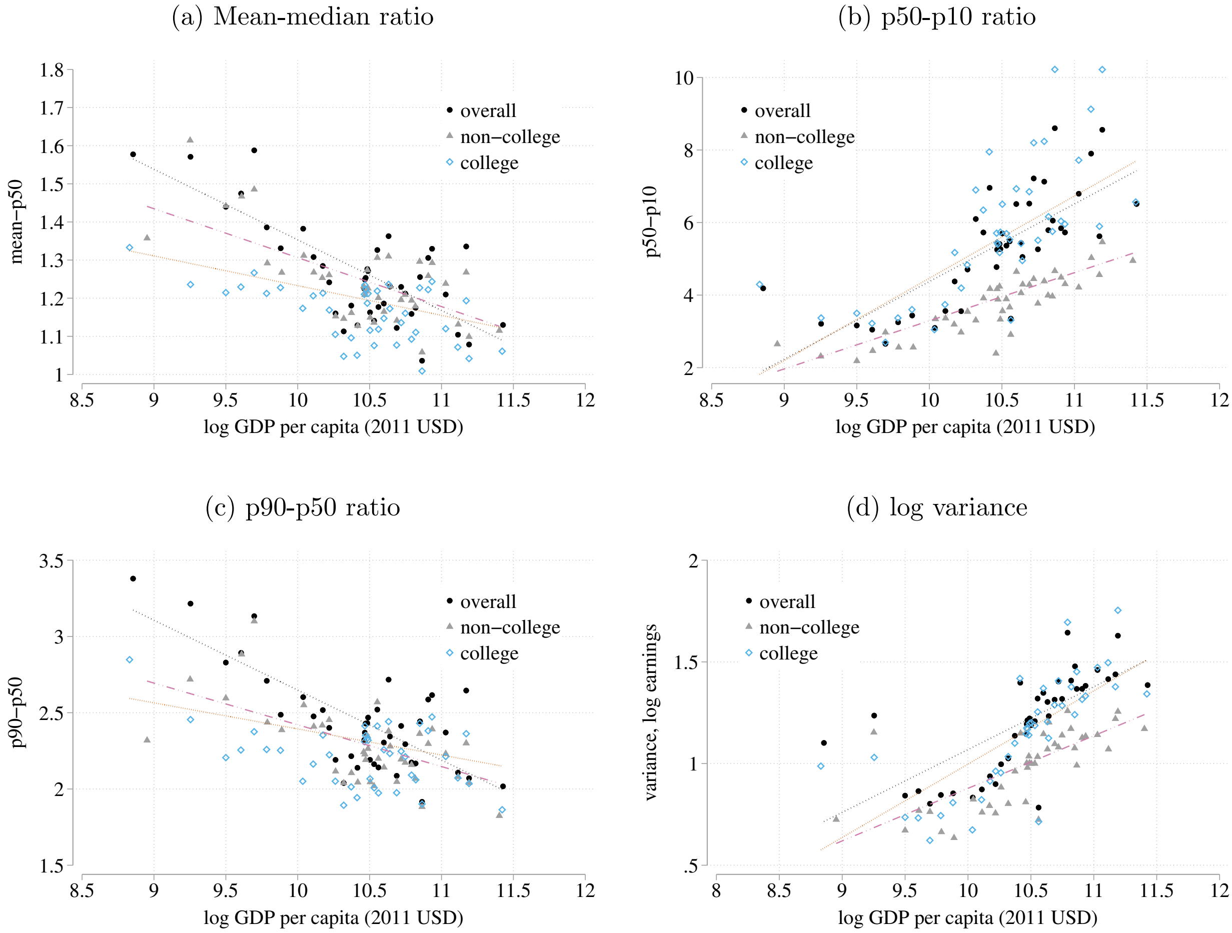
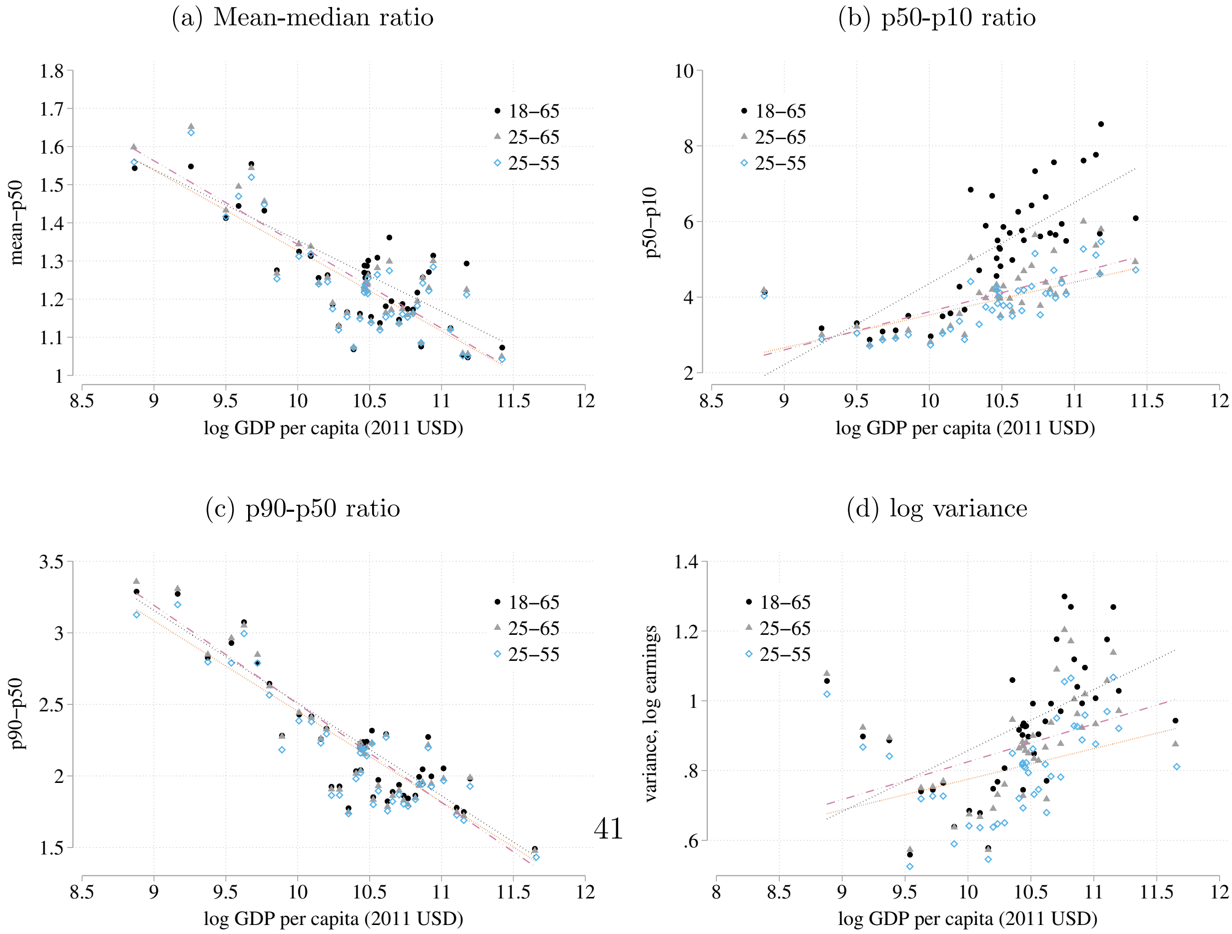
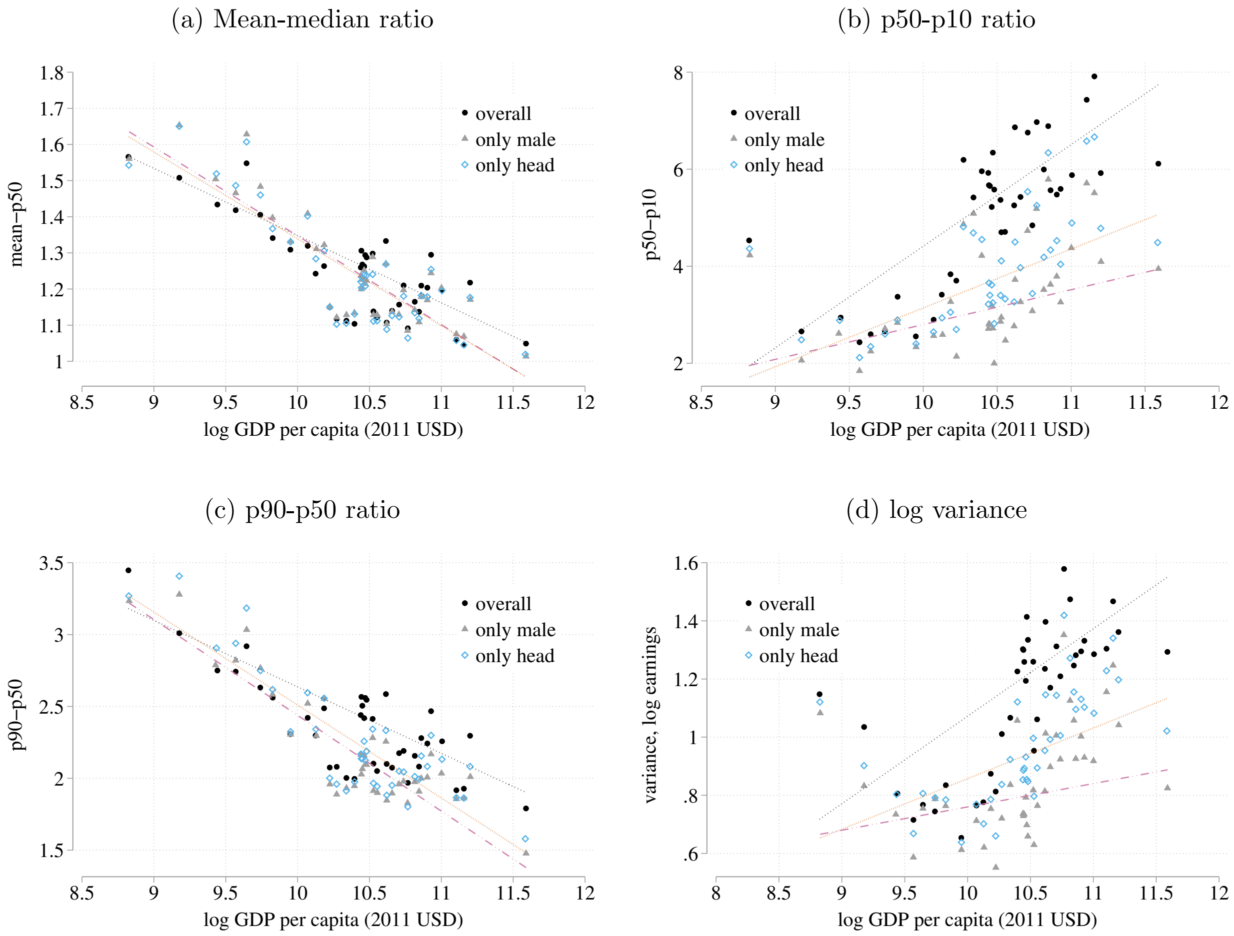


Figure 12: **Earnings inequality across countries, by age**



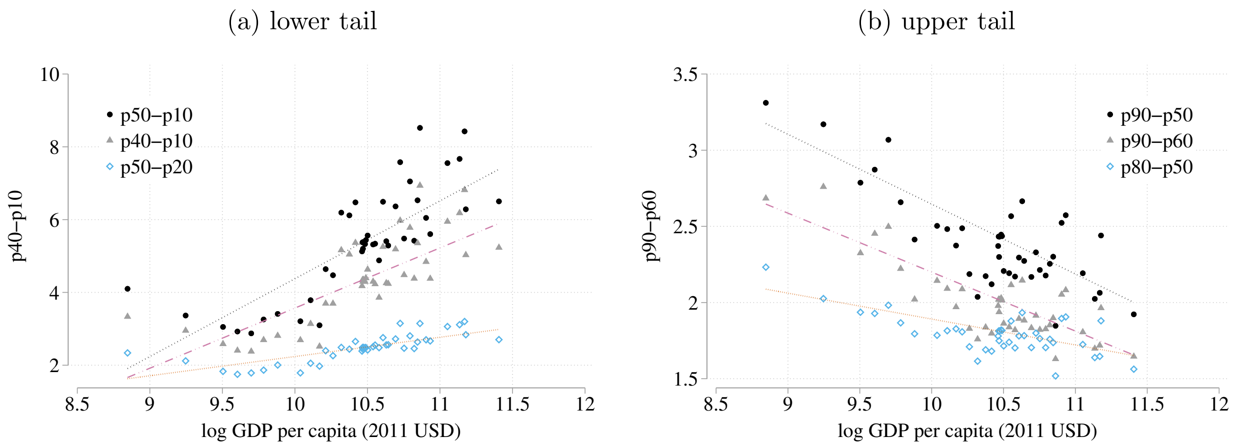
Source: IPUMS, EU-SILC, LIS and author's calculations

Figure 13: **Earnings inequality across countries, by demographic groups**



Source: IPUMS, EU-SILC, LIS and author's calculations

Figure 14: **Earnings inequality across countries, alternative measures**



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.[9] 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.

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

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, 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.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.

A.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:

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

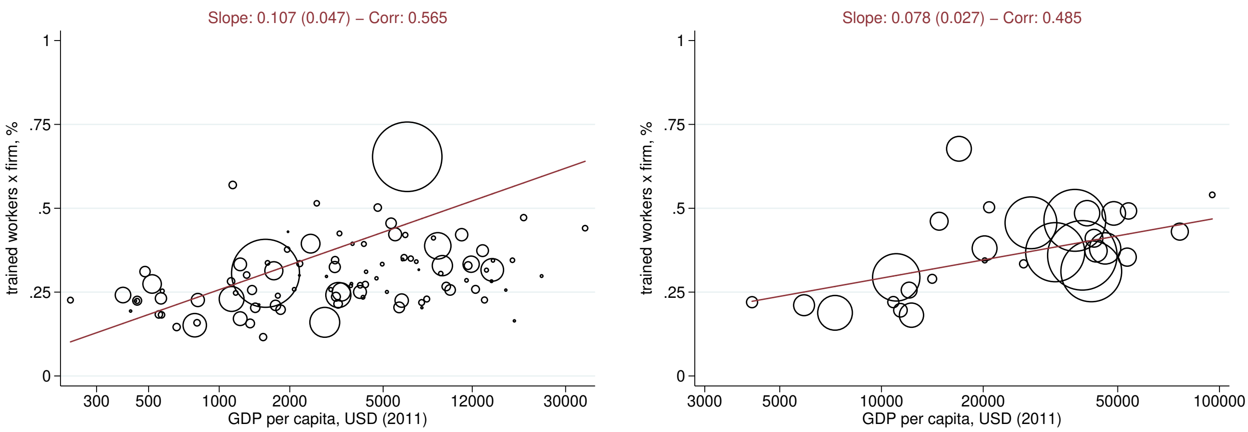
The Eurostat reports this variable constructed using data from the CV-TS.

Fact 1. The share of trained workers within firms increases with GDP per capita. Figures 15(a) and (b) report the average share of workers 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. Each circle represents a country, with larger circles denoting larger population share of the country in the sample.

Firms in more developed countries provide training to a larger share of their workers. 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

Figure 15: **Training provision across countries**

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



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

Table 17: 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	<50	29.31	21.96
20-49	40.06	25.56	29.72	30.18	50-249	37.92	30.13
50-249	44.35	26.68	35.51	30.36	≥500	49.71	46.25
250-449	52.51	30.30	32.22	28.86			
≥500	50.73	32.37	34.34	28.98			

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

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.

Fact 2. Larger firms provide OTJ training to larger share of workers. Table 17 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.

Table 18: Firm level wage premium from training

	$\log w_{it}$				
Quartiles	(1)	(2)	(3)	(4)	
Percentiles	(0-25)	(25-50)	(50-75)	(75-90)	(90-100)
$\mathbf{1}_{it}^{\text{training}}$	0.283***	0.267***	0.189***	0.177***	0.179***
	(0.0404)	(0.0430)	(0.0290)	(0.0485)	(0.0540)
Observations	28453	16870	23260	12604	8578
R^2	0.691	0.533	0.348	0.318	0.469
Country FE	✓	✓	✓	✓	✓
Time FE	✓	✓	✓	✓	✓
Sector FE	✓	✓	✓	✓	✓
Controls	✓	✓	✓	✓	✓
Average GDP p.c. (2011 USD)	881.83	2447.44	5169.667	10060.75	20823.44

Note: * p<0.10, ** p<0.05, *** p<0.01. Source: World-Bank Enterprise Survey and author′s calculation.

Fact 3. Firms providing OTJ training pay a wage premium, but the premium declines with development. As a robustness analysis for the wage return from OTJ training, we re-estimate a version of equation (1) in the main text for different groups of countries separately. In particular, we split the sample in five groups according to the GDP per capita of the country where firm operates and estimate the following equation:

$$\log w_{it} = \alpha \mathbf{1}_{it}^{\text{training}} + \mu_{c(i)} + \mu_t + \mu_{s(i)} + \gamma X_{it} + \epsilon_{it}$$

for each group separately. Table 18 reports the estimates for the sample of firms in the first three quartiles of GDP p.c., and for two subsamples of the last quartile, separating those above and below the 90th percentile. In each regression we include the full set of controls (firm size and firm age dummies, dummies for export and ownership status). Robust standard errors are reported in parenthesis. The estimates confirm the main evidence. Firms providing OTJ training pay higher wages, and the wage premium is estimated to be between 20 and 30%. Second, this premium lowers significantly in richer countries. Moving from the first quartile to the last decile of the sample, the premium decrease of 10 percentage points, from 28.3 to 17.9%.

A.5 Estimation data

Table 19 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 76% of the individuals reports to be full-time employed, and work on average 37 hours in a week. Around 25% of the respondents who are employed reports to have received on-the-job training in the current quarter. The LFS reports information for tenure on the job using indicators for whether an individual has been employed in the same firm for < 3 months, for a period $\in [3, 12)$ months, $\in [12, 24)$ months, and for ≥ 24 months.

Table 19: Descriptive Statistics

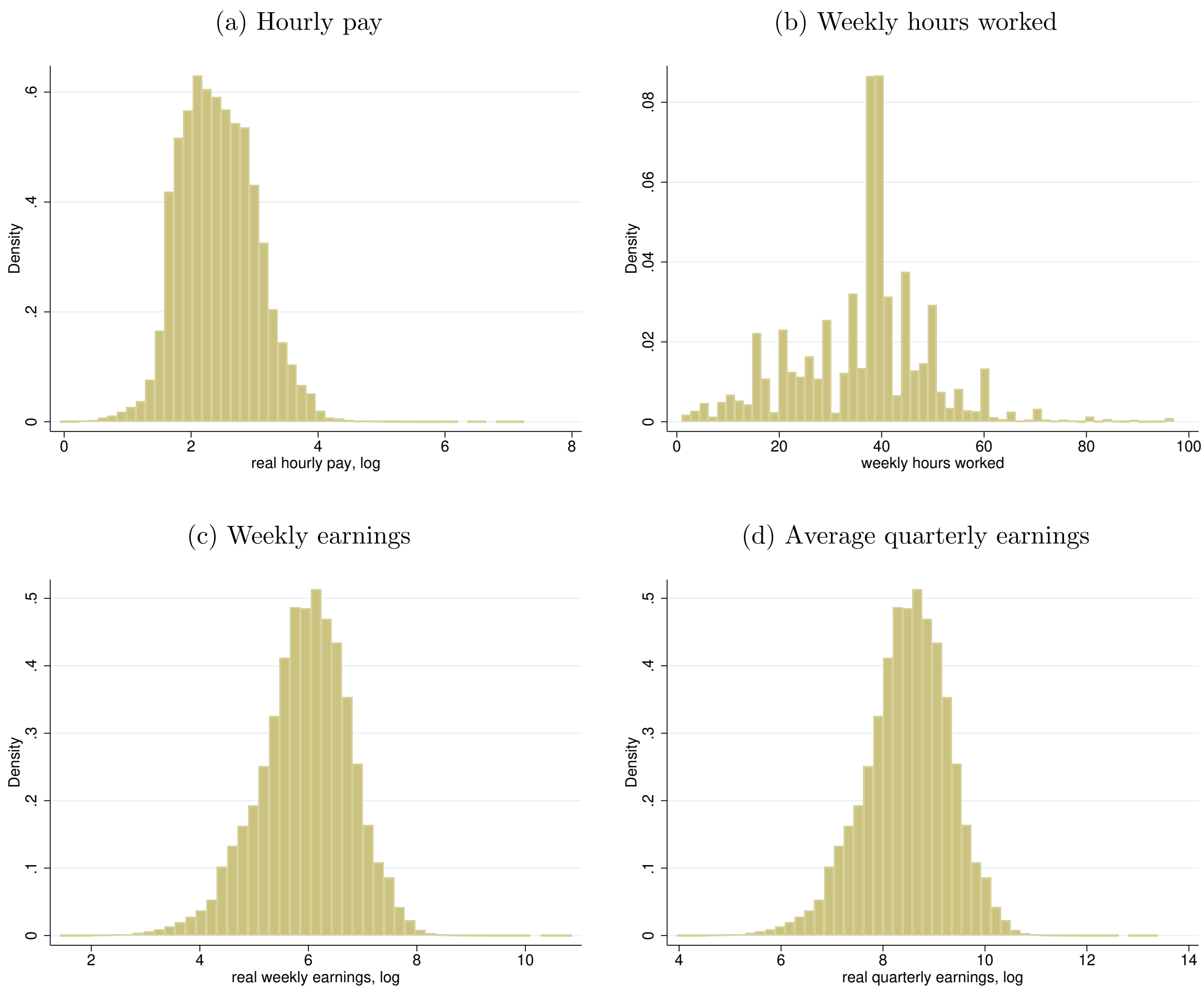
	Mean	SD	Min	Max	N
Employed workers					
Age	41.629950	11.638060	22	62	85,524
Female	0.5054908	0.4999703	0	1	85,524
Full-time	0.7559546	0.4295223	0	1	85,524
Hours worked	37.043440	12.098500	1	97	85,524
Log Hourly pay	2.385007	0.5989295	0.025252	7.247456	85,524
Log Quarterly Earnings	8.456721	0.8237451	3.955738	13.39207	85,524
Training	0.2442638	0.4296524	0	1	85,524
Tenure<3 months	0.0377040	0.1904806	0	1	85,524
Tenure∈[3,12) months	0.0385089	0.1924224	0	1	85,524
Tenure∈[12,24) months	0.1085912	0.3111274	0	1	85,524
Tenure≥24 months	0.8151959	0.3881409	0	1	85,524

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.



(a) Hourly pay

(b) Weekly hours worked

(c) Weekly earnings

(d) Average quarterly earnings

B Model Appendix

B.1 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, ϕ_w^0, and use the definition of matching function to compute the job contact rate for firms, ϕ_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, ξ), $\psi_v^0(z, \xi)$

   2.1. Given ϕ_w^0 and $\psi_v^0(z, \xi)$, solve for the surplus function, $S^h(z, \xi, h)$. 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, $\mathbf{1}^h(z, \xi, h)$ and on-the-job training $\mathbf{1}^t(z, \xi, h)$

   2.3. Use ϕ_w^0, $\psi_v^0(z, \xi)$, $\mathbf{1}^h(z, \xi, h)$ and $\mathbf{1}^t(z, \xi, h)$ to simulate a large panel of workers and construct a distribution of non-employed workers over human capital, $\psi_h^u(h)$, and the aggregate measure of workers who are non-employed, U.

   2.4. Given ϕ_f^0, $\mathbf{1}^h(z, \xi, h)$, $\psi_h^u(h)$, 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

$$M = M_e \int_{z \in \mathcal{Z}} \int_{\xi \in \mathcal{E}} \mathbf{1}^e(z, \xi) \psi_z(z) \psi_\xi(\xi) dz d\xi \tag{18}$$

and use stationarity condition to compute total number of firms

$$N = \frac{M}{\delta_f} \tag{19}$$

4. Construct the aggregate measure of vacancy posted

$$v = N\bar{v} \tag{20}$$

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

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

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

6. Check for convergence:

   - if ϕ_w^1 and ϕ_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 ϕ_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 Appendix

C.1 Estimation of matching elasticity

We estimate the matching elasticity outside of the main estimation algorithm. To compute quarterly new hirings use employment gross inflows from the ONS Labor Force Survey Flows Estimates (dataset X02, available here). From the same source, we obtain data on aggregate active vacancies (dataset AP2Y, available here) and stock of non-employed workers (dataset ANZ6, available here). 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 20 reports estimates and standard errors obtained using the robust GMM weighting matrix in the second step.

Table 20: Matching elasticity estimation

Parameters	Description	Estimates	St.Error
η	Matching function	0.5416	0.0134
$\mathbf{1}_t^{q=1}$	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

C.2 Estimation algorithm

In the calibration 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, ϕ_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:

$$\phi_w - \frac{U(\bar{v}\delta_f M)}{(U^\eta + (\bar{v}\delta_f M)^\eta)^{\frac{1}{\eta}}} = 0 \tag{22}$$

where M is defined in equation (18). To calibrate the model, we follow this algorithm:

1. Guess the following set of parameters:

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

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

2. Given ϕ_w^0, compute job contact rate for firms, ϕ_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 (22)

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:

$$D(\vartheta^0) = \bar{d}(\vartheta^0)'\bar{d}(\vartheta^0) \tag{23}$$

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 percentage deviation between empirical and simulated moments is 8.078117e-02.

C.3 Standard errors

To obtain estimates standard errors and confidence intervals, we follows Chernozhukov and Hong (2003) methodology. The estimation procedure of Chernozhukov and Hong (2003) 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 (23) 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 ϑ^j. Next, we impose the proposal density to be a uniform and we extract a new guess ϑ^p from it. Finally, we update from ϑ^{j+1} from ϑ^j for $j = 1, 2, ...$, 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. Figures 16 and 17 report the posterior density for the estimated parameters in Table 5 of the main text.

Figure 16: **Posterior distributions**

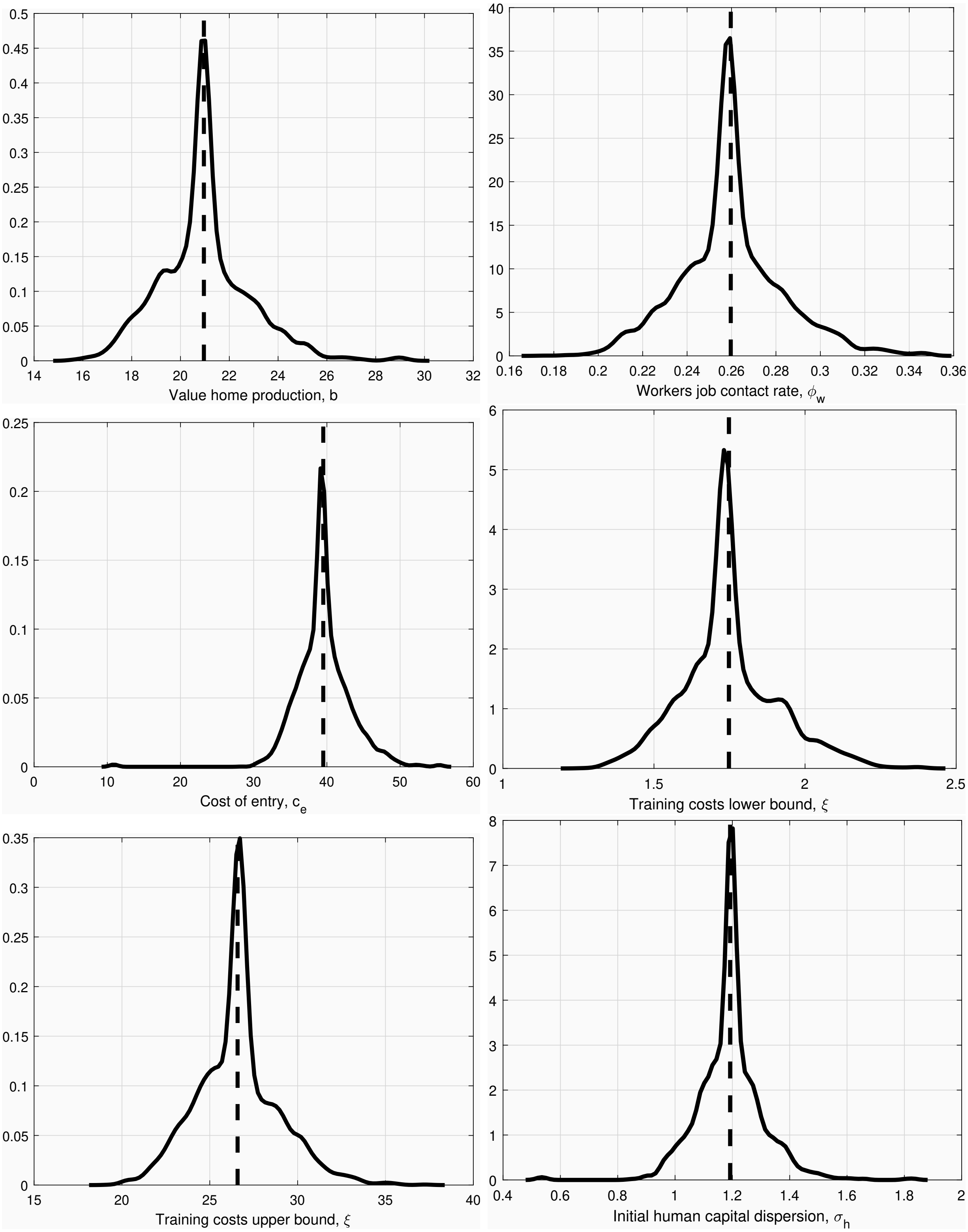
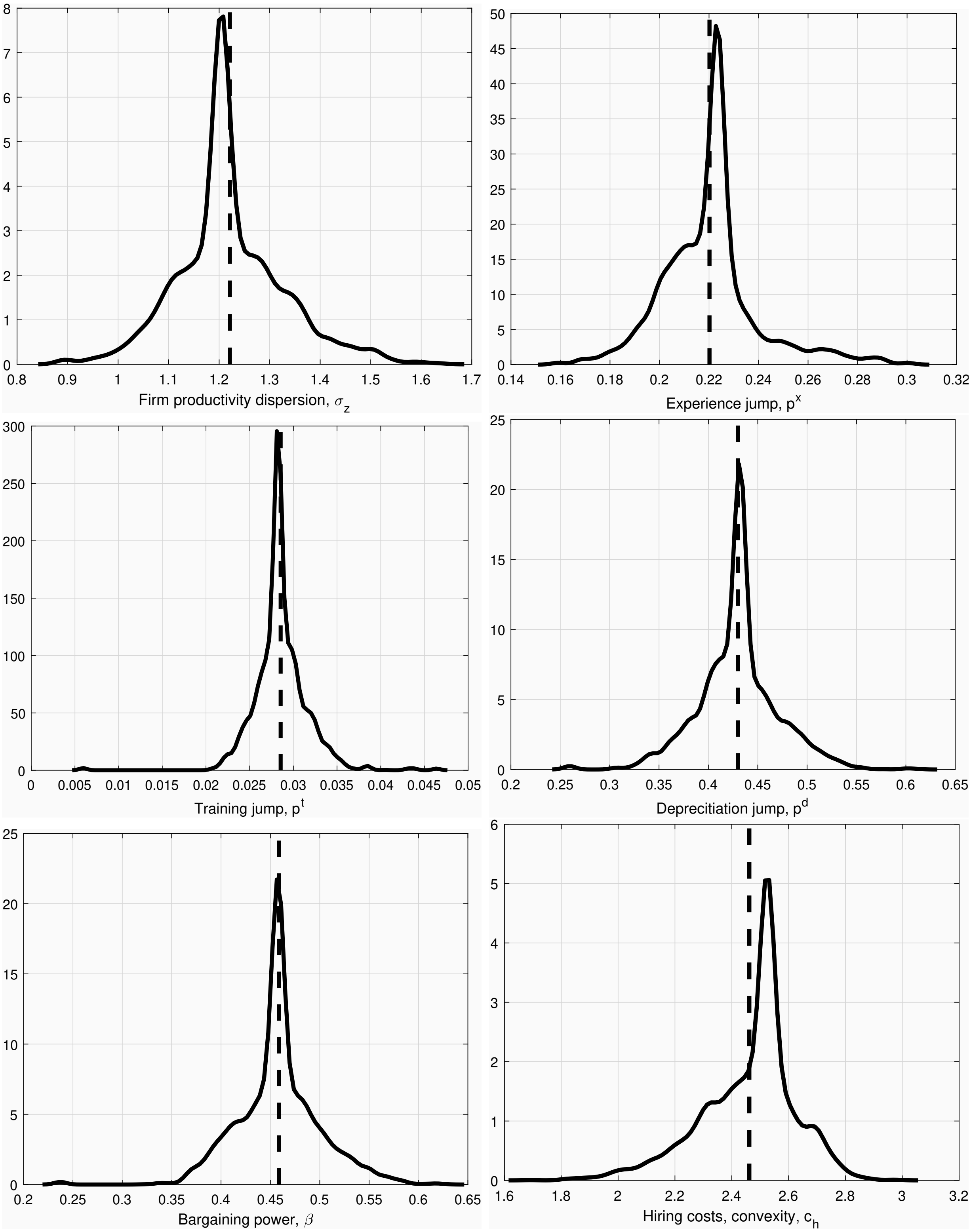


Figure 17: **Posterior distributions**



C.4 Identification sensitivity

To assess identification of parameters we conduct two main sensitivity analysis. First, we discuss how sensitive sensitive are the estimates, $\hat{\theta}$, to the choice of the targets used to calibrate the parameters δ_s, b and M, i.e. employment rate, the value of non-market activity and job duration. To do this, we use the sensitivity measures proposed by Andrews et al. (2017), which is constructed as follows:

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

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

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

Table 21 reports the elasticity of each estimated parameter j to misspecification in moment condition i, i.e.

$$\Lambda_{(i,j)}\frac{\bar{d}_{(i)}(\vartheta)}{\vartheta_{(j)}}$$

This final measure can be interpreted as the percent bias in parameter j estimate for a one percent perturbation in moment condition i.

Table 21: Sensitivity to targeted moments

Parameters	Description	Elasticity to changes in selected targets: Employment rate	Value non-market activity	Job duration
c_e	Entry cost	-1.7418e-06	-0.00093274	6.6288e-06
$\underline{\xi}$	Training cost (lower bound)	0.00084365	-0.11584	0.0026146
$\overline{\xi}$	Training cost (upper bound)	-6.2391e-05	0.0013263	-0.00020921
λ_1	Hiring costs, convexity	7.259e-05	0.016565	0.00016924
β	Bargaining power	0.00033109	0.23608	-0.0028881
σ_h	Initial human capital dispersion	0.0010453	-0.018899	0.0026965
σ_z	Firm-productivity dispersion	8.7147e-05	-0.023688	4.3933e-05
p^e	Experience jump	-0.00048228	-0.12304	-0.0059591
p^t	Training jump	-0.0035956	-0.28828	-0.044325
p^d	Depreciation jump	0.00034966	-0.05707	0.003446

Parameter estimates are sensitive to model misspecification in employment rate and job duration. The bias is lower that 0.1 percent for most of parameters. The probability of a human capital jump due to on-the-job learning, p^e and job training, p^t, and the probability

Table 22: Estimate sensitivity to matching elasticity

Parameters	Description	Elasticity to changes in η
c_e	Entry cost	-0.0037542
$\underline{\xi}$	Training cost (lower bound)	-0.37768
$\overline{\xi}$	Training cost (upper bound)	0.010577
λ_1	Hiring costs, convexity	-0.047913
β	Bargaining power	-0.74726
σ_h	Initial human capital dispersion	0.084774
σ_z	Firm-productivity dispersion	-0.023752
p^e	Experience jump	-0.48782
p^t	Training jump	-9.2998
p^d	Depreciation jump	0.62543

of skill depreciation during non-employment, p^u, seem to be sensitive to misspecification in the value of non-market activity.

Finally, we look at how sensitive are the estimates to changes in the calibrated value of η. To this purpose, we use the sensitivity measure proposed by Jørgensen (2020), who extends Andrews et al. (2017)'s measure as follows:

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

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

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

Table 22 reports the elasticity of each estimated parameter j to the value of η, equal to

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

The probability of a human capital jump due to on-the-job learning, p^e and job training, p^t, and the probability of skill depreciation during non-employment, p^u, are particularly sensitive to changes in matching elasticity. This is the case because η directly affects matching efficiency. Changes in the rate at which new matches are formed are key determinant of wage dynamics.

D 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 need to be calibrated. Three of them - those reported in Table 23 - have a direct mapping to a specific moment and are calibrated to match them.

Table 23: Parameters calibrated solving the model

Parameters	Description	Value	Source/Targets
δ_s	Match separation	0.01229	Job duration=5.36 years (Mumford and Smith, 2004)
b	Home production	22.2638	Value non-market activity=19% (Alpman et al., 2018)
M_e	Measure of potential entrants	0.03063	Employment rate= 77.58% (ONS)

The remaining seven parameters - reported in Table 24 - are calibrated to match 29 moments reported in Tables 25 and 26. 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 12 are based on this calibration.

Table 24: Parameters calibrated through indirect inference

Parameters	Description	Value
c_e	Entry cost	44.752
λ_1	Hiring costs, convexity	2.5322
β	Bargaining power	0.4272
σ_h	Initial human capital dispersion	1.0349
σ_z	Firm-productivity dispersion	1.2210
p^e	Experience jump	0.2089
p^d	Depreciation jump	0.4301

Table 25: Firm-level Targeted Moments

	Data	Model
Firm-level moments		
Number of firms (over population)	0.171	0.248
$E(\ell_t)$	16.423	16.180
$E(\log \ell_t)$	1.7393	1.7892
$\text{std}(\log \ell_t)$	1.2198	1.3707
Firm-size distribution		
1-9 employees	72.12	69.11
10-24 employees	15.95	15.68
25-49 employees	6.12	7.31
50-99 employees	3.21	4.62
100-249 employees	1.73	3.08
250+ employees	0.88	0.21
Firm-size percentiles		
10th percentile	1	1.1808
25th percentile	3	2.6889
40th percentile	4	3.9837
50th percentile	5	5.0975
60th percentile	6	7.1108
75th percentile	11	13.597
90th percentile	29	39.889
95th percentile	53	72.536
99th percentile	202	175.42

Table 26: Worker-level Targeted Moments

	Data	Model
Wage distribution		
Wage at entry, $E[\log(w_1/\bar{w})]$	-0.5176	-0.4788
Wage after 20 y.o., $E[\log(w_{20}/\bar{w})]$	0.1071	0.1084
Wage at re-emp, $E[\log(w_R/\bar{w})]$	-0.3010	-0.1625
Wage dispersion at entry, $\text{sd}[\log w_1]$	0.5818	0.5705
Wage dispersion after 20 y.o., $\text{sd}[\log w_{20}]$	0.7959	0.7380
Wage dispersion at re-emp, $\text{sd}[\log w_R]$	0.8335	0.7348
Job tenure return		
tenure$<$3 months	1	1
tenure$\in$[3,12) months	1.0551	1.0526
tenure$\in$[12,24) months	1.1320	1.1357
tenure$\geq$24 months	1.3675	1.3693