Firms, Training, and Development*

Xiao Ma
Peking University

Alejandro Nakab
Universidad Torcuato Di Tella

Daniela Vidart
University of Connecticut

March 2026

Abstract

We study how firms shape on-the-job training along the development path. Using data from over 100 countries, we show that smaller firms consistently provide fewer training opportunities to their workers. Evidence from administrative firm-level data in China and Mexico indicates that cross-firm differences in productivity and labor shares help explain this pattern, while cross-country evidence on worker mobility highlights the importance of poaching risk for firms' training incentives. We formalize these mechanisms in a general equilibrium wage-posting model in which labor market frictions shape worker mobility and interact with firm characteristics to determine training investments. A quantitative version of the model, calibrated to match economies at different income levels, shows that labor market turnover is the dominant driver of cross-country training gaps. Through its effect on training, reducing labor market turnover in low- and middle-income economies to U.S. levels can raise income per capita by up to 10–20 percent.

Keywords: On-the-job training; Human capital accumulation; Firm heterogeneity; Labor market frictions

JEL codes: E24, O11, J24, D22, M53

*Email: xiaoma@phbs.pku.edu.cn, anakab@utdt.edu, and daniela.vidart@uconn.edu. An earlier version of this paper circulated under the title "Who Pays for Training? Theory and Evidence on Firm-Level Differences in Training Investments." We would like to thank Titan Alon, Francesco Amodio, Niklas Engbom, David Lagakos, Elena Pastorino, Josep Pijoan-Mas, Tommaso Porzio, Valerie Ramey, Natalia Ramondo, and Thomas Sargent for their helpful comments. We are also grateful for the insightful comments of participants at UCSD, UNH - Workshop on Economic Dynamics, SUNY Albany, SEA, Bank of Portugal and Bank of Spain Workshop in Labor Economics, Stony Brook, SED Winter Meeting, PHBS Meeting on Theoretical and Empirical Macroeconomics, and RIDGE Micro to Macro Development Workshop.

1 Introduction

A large body of evidence shows that on-the-job training has large and persistent effects on worker productivity (Leuven and Oosterbeek, 2004; Bassanini et al., 2005). Given the substantial cross-country gaps in on-the-job training documented in the literature, with workers in richer economies receiving considerably more training over their careers than those in poorer ones (Ma et al., 2024), these gaps are an important source of cross-country differences in productivity and income. While recent work has explored some channels driving these cross-country training gaps,[1] much less attention has been paid to the role of firms and their characteristics, such as size, productivity, and organizational structure, which a growing literature shows to be central to training investments (Black et al., 1999; Braga, 2018; Gregory, 2019; Friedrich et al., 2021; Jarosch, 2022). Because these characteristics vary systematically with a country's stage of development, for example with poorer countries having a larger share of small and less productive firms, firms may constitute a key channel shaping investments in on-the-job training and cross-country human capital gaps.

In this paper, we study how firms shape on-the-job training along the development path. First, we provide new empirical evidence on how training patterns vary with firm characteristics across countries at different income levels. We then develop a model in which firm heterogeneity interacts with labor market frictions to determine training provision incentives. The model allows us to interpret the empirical patterns and to assess how differences in the firm distribution, firm-level returns to training, and labor market turnover contribute to cross-country training gaps. Together, our results highlight the central role of firms in on-the-job learning and provide new insights into how human capital accumulation evolves with economic development, with implications for policies aimed at raising productivity through training investments.

To start, in the empirical portion of the paper, we study how training investments vary across firms and countries. We first construct a harmonized and consistent measure of on-the-job training that encompasses all organized or sustained learning activities occurring outside the formal education system. This measure captures several important sources of workers' human capital accumulation, including participation in seminars or workshops as well as task-related learning arising from instruction by coworkers. Using data from over 100 countries, we show that on-the-job training opportunities are systematically lower in

[1] For example, in our earlier work (Ma et al., 2024), we study the role of self-employment, physical capital differences, and aggregate job turnover in explaining cross-country training gaps.

smaller firms across all countries, a pattern that is robust to controlling for year, country, and industry fixed effects.

We then investigate the mechanisms underlying this relationship, focusing on three factors emphasized in the literature as central to firms' training decisions: productivity, the labor share, and job turnover. The focus on productivity is motivated by theoretical work showing that more productive firms face higher returns to training investments, as human capital and firm productivity are complementary (Acemoglu and Pischke, 1998; Bagger et al., 2014). This complementarity is supported empirically by evidence of positive assortative matching between firms and workers (Barth et al., 2016; Abowd et al., 2018; Song et al., 2019). The focus on the labor share reflects the fact that it directly determines how much of the returns to human capital accumulation firms capture relative to workers, thereby shaping their incentives to train. Finally, the focus on job turnover stems from the classic hold-up problem in training: the possibility that workers leave after being trained weakens firms' incentives to invest in training (Becker, 1964; Acemoglu, 1997; Acemoglu and Pischke, 1999; Pischke, 2001; Moen and Rosen, 2004).

We next bring these mechanisms to the data. Using administrative firm-level data from China and Mexico, we first document that cross-firm differences in labor shares and productivity are systematically related to firm-provided training. Firms with higher labor shares are less likely to provide training, whereas firms with higher productivity are more likely to do so. Because larger firms typically exhibit both lower labor shares and higher productivity, these channels help explain the positive relationship between firm size and training provision. However, even after controlling for these factors, firm size remains a strong predictor of training, suggesting that additional mechanisms—such as economies of scale in training—also contribute to this relationship. These findings are robust to alternative measures of productivity and to the inclusion of year, industry, and firm fixed effects. To investigate the role of job turnover, we turn to country-level European data and show that the share of formally trained workers in a country-year declines with job turnover rates, consistent with higher poaching risk weakening firms' incentives to invest in training.

We then develop an analytical model that formalizes how firms' incentives to invest in training vary with firm characteristics and embeds the mechanisms in the empirical analysis. The model features firm heterogeneity and labor market frictions à la Burdett and Mortensen (1998). Firms differ in labor productivity and post vacancies and wages to meet workers through random search. Workers separate from firms either through exogenous job destruc-

tion into unemployment or via job-to-job transitions driven by on-the-job search.

Training investments are determined and financed after the worker and firm match, with equilibrium outcomes shaped by firms' incentives and firm characteristics in ways that mirror the empirical patterns. First, consistent with the positive relationship between productivity and training, production exhibits complementarities between firm productivity and worker human capital, encouraging more productive firms to train more. Second, consistent with the negative relationship between the labor share and training, firms with higher labor shares appropriate a smaller fraction of the returns to training, reducing their incentives to invest. Third, consistent with the positive relationship between firm size and training, training returns depend directly on firm size, reflecting economies of scale in training. Fourth, consistent with the role of job turnover, higher separation risk reduces firms' incentives to invest in training, as wage differences across firms imply differential exposure to poaching.

We then extend the analytical model to a quantitative setting to assess the relative importance of these mechanisms in explaining cross-country training gaps. We first calibrate the model to the United States and then recalibrate it to match representative economies at six income levels. In each recalibration, we focus on parameters that govern the mechanisms highlighted in both the empirical analysis and the analytical model and that shape firms' incentives to invest in training. First, to capture the role of productivity–training complementarities and firms' ability to appropriate training returns, we vary the parameters governing the firm productivity and size distribution across countries. Firms' own productivity determines their returns to training in production, while the shape of the productivity distribution shapes wage-posting incentives and therefore labor shares and the extent to which firms can appropriate training returns. Second, to capture the role of job turnover and poaching risk, we vary the parameters governing labor market turnover, including separation rates and job-to-job mobility, which determine the expected duration of matches across countries. Third, to capture cross-country differences in learning environments and how these vary with firm size, we vary parameters governing the level and scale of training returns.

We begin with a decomposition analysis that changes the parameters associated with each channel one at a time in each economy, to evaluate their relative strength in explaining the training gap. We find that differences in labor market turnover are the dominant driver of training gaps along the entire development path. Reducing labor market turnover to U.S. levels increases training by at least 15% in all economies, with particularly large ef-

fects in poorer countries, where high worker turnover severely limits firms' incentives to invest in training. In contrast, differences in training returns and in the productivity and size distribution of firms play a more limited role overall and are especially less important in low-income economies, where high labor market turnover substantially dampen training incentives across all firms. Nevertheless, the effects of reducing labor market turnover are amplified when combined with improvements in the other two channels, and especially training returns: lower frictions concentrate employment in larger and more productive firms, precisely where training returns are highest in the U.S. Taken together, these findings underscore the central importance of the hold-up problem in the underprovision of training in the developing world, and the potential gains from coordinated policy interventions that promote sustained job attachment together with economies of scale in training.

Finally, we quantify how each channel contributes to cross-country differences in human capital and income through its effects on training. Consistent with our prior results, we find that reducing labor market turnover to U.S. levels generates the largest gains among the three channels, especially in low- and middle-income economies, raising human capital and income by 19.29% and 17.23%, respectively, in the poorest country, for instance. Aligning the productivity and firm size distribution yields smaller but still positive gains of 5.91% and 5.93% in both human capital and income in this setting. These findings highlight that reforms improving labor market efficiency and addressing the "missing middle firms" problem, whereby poorer countries have relatively few medium-sized firms, can generate substantial additional income gains through increased training.

The remainder of the paper proceeds as follows. Section 2 reviews the related literature and positions our contribution. Section 3 describes the data and presents the empirical evidence. Section 4 introduces the analytical model. Section 5 presents the quantitative model and its calibration. Section 6 uses the calibrated model to decompose the training gap and evaluate the role of training in human capital accumulation and cross-country income differences. Section 7 concludes.

2 Related literature

Through its focus on the role of firms in driving and funding on-the-job training, our paper relates to the theoretical literature on general training investments, first proposed by Becker (1964) and subsequently developed by others (Acemoglu, 1997; Moen and Rosen, 2004). A central insight of this literature is that firms may have limited incentives to finance general

training, since the portability of skills implies that general human capital gains are quickly reflected in wages. However, subsequent work has shown that firms may still be willing to invest in general training in the presence of labor market frictions, which allow them to capture part of the returns to training (Acemoglu and Pischke, 1999). More recently, a growing literature has examined the determinants of firms' training decisions (Black et al., 1999; Braga, 2018).[2] Empirical studies in this literature have documented a positive correlation between firm size and training investments in the United States (Barron et al., 1987; Frazis et al., 1995) and the United Kingdom (Harris, 1999).

We contribute to this literature in three main ways. First, we use data from over 100 countries to show that the positive relationship between firm size and training is also prevalent in low- and middle-income countries. Second, using administrative firm-level data from China and Mexico, we examine the roles of productivity and the labor share in shaping this relationship. Third, we bring these insights to a macro-development context by modeling training investments in a framework with firm heterogeneity and labor market frictions that generate a hold-up problem varying in severity across firms, and by quantifying the contribution of firms to cross-country training gaps along the development path.[3]

In addition, by highlighting the lower levels of training prevalent in smaller and less productive firms, particularly in developing countries, our paper also relates to the misallocation literature, which studies the productivity losses stemming from the large share of small, unproductive firms documented in many developing economies (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013; Bento and Restuccia, 2017). A large contingent of the misallocation literature has focused on studying the factors driving the large fraction of unproductive firms in developing settings. The factors discussed include financial and credit constraints (Buera et al., 2011; Midrigan and Xu, 2014), policy distortions (Restuccia and

[2] A related literature studies the role of firms and firm-level characteristics in shaping workers' human capital accumulation (see, for example, Gregory, 2019; Arellano-Bover, 2020; Engbom, 2021; Friedrich et al., 2021; Jarosch, 2022; Engbom et al., 2022; Arellano-Bover and Saltiel, 2023). This work documents substantial heterogeneity across firms in their promotion of human capital accumulation and shows that this heterogeneity is an important determinant of life-cycle earnings dynamics.

[3] Our paper also relates to the large labor literature examining the impacts of on-the-job training, and particularly firm-sponsored job-related training, on productivity and wages (see Leuven and Oosterbeek, 2004; Bassanini et al., 2005; Heckman et al., 1999; Kluve, 2010; What Works - Centre for Local Economic Growth, 2016; McKenzie, 2017; Card et al., 2018; Ma et al., 2024, for reviews). This literature documents overwhelmingly positive effects of work-related training on wages and productivity. Some work focuses on disentangling the effects of on-the-job training on wages from its effects on productivity (e.g., Dearden et al., 2006; Conti, 2005; Konings and Vanormelingen, 2015), generally finding that productivity gains substantially exceed wage gains, consistent with firms capturing part of the returns to training.

Rogerson, 2008; Da-Rocha et al., 2017), technology availability and shocks (Porzio, 2017; Poschke, 2018), information imperfections (David et al., 2014), limits to delegation (Akcigit et al., 2016), entrepreneurial and worker human capital (Roys and Seshadri, 2014), and managerial skill (Bhattacharya et al., 2012), among others. We argue that, beyond the static productivity losses emphasized in this literature, the concentration of workers in such firms generates additional dynamic losses by limiting workers' access to on-the-job training and slowing human capital accumulation.

3 Data and empirical evidence on firms, labor market turnover, and training investments

We now present empirical evidence on how training investments vary across firms along the development path. We first describe the data sources and define our measure of on-the-job training. We then present the main empirical findings.

3.1 Data

3.1.1 Cross-country data used to document link between firm size and training

To document the relationship between firm size and training investments, we draw on firm-level data from more than 100 countries. Our primary source is the World Bank Enterprise Survey (WB-ES), which we complement with the European Union Continuing Vocational Training (EU-CVT) enterprise survey. Together, these two datasets cover both developing and developed economies with per-capita GDP levels ranging from $1,000 to $60,000, allowing us to assess how firm size relates to training across all stages of development. In addition, we use cross-country worker-level data from the OECD Programme for the International Assessment of Adult Competencies (PIAAC) to further corroborate the firm size–training relationship.

The WB-ES is a collection of firm-level surveys covering the private manufacturing and service sectors, encompassing approximately 136,000 firms across 140 low- and middle-income countries. The surveys are administered through interviews with establishment owners and top managers, who may consult their firms' accountants or human resources staff when responding to specific questions. Fully state-owned firms are excluded. Firms with fewer than five employees are also typically excluded, although in some country surveys the WB-ES in-

cludes establishments with fewer than five employees. We rely on the two WB-ES waves conducted between 2002–2005 and 2006–2017, as these waves include standardized questions on worker training. Additional details on the WB-ES data are provided in Appendix A.1.

The EU-CVT provides information on European enterprises' investments in continuing vocational training for their workforce, including measures of participation, training intensity, and training expenditures. Due to data availability, our analysis relies on three waves of the EU-CVT conducted in 2005, 2010, and 2015, which cover all EU member states as well as Norway. Additional details on the EU-CVT data are provided in Appendix A.2.

PIAAC is an international survey conducted in over 40 OECD countries between 2011 and 2017. The survey assesses and compares the learning environments, skills, and competencies of adults aged 16 to 65. In total, it covers a sample of approximately 230,000 individuals. PIAAC collects detailed information on workers' participation in skill-related learning activities, as well as on how these skills are used in various settings. Additional details on the PIAAC data are provided in Appendix A.3.

3.1.2 Administrative data used to document link between productivity, labor share, and training

To investigate the drivers of the relationship between firm size and training, particularly the roles of firm productivity and the labor share, we rely on administrative data from Chinese and Mexican firms. The Chinese data come from the Annual Survey of Manufacturing, which contains detailed financial information for all manufacturing firms with revenues exceeding 5 million RMB. We use data from 2005–2007, which include information on expenditures on training fees. The Mexican data come from the Economic Census, which contains detailed financial information for all establishments operating in a permanent location outside of agriculture and forestry, thus excluding peddlers and vending carts, for example. We primarily rely on data from the 2019 census, which report the share of workers who received training in the previous year. Additional details on these data sources are provided in Appendix A.4 and Appendix A.5, respectively.

3.1.3 Cross-country data used to document link between job turnover and training

To examine the role of job turnover in shaping training investments, we combine data on firms' training provision from the EU-CVT with country-level measures of job turnover from

Donovan et al. (2023).[4] As the country-year coverage of the EU-CVT does not fully overlap with the turnover measures in Donovan et al. (2023), we additionally use institutional indicators from the World Bank Worldwide Governance Indicators to proxy for cross-country differences in job turnover. Additional details on these data sources are provided in Appendix A.6.

3.2 Defining on-the-job training

Before turning to the empirical evidence, we define on-the-job training and its key characteristics to ensure consistency across data sources. Following Ma et al. (2024), we adopt the definition of non-formal education and training from ISCED (2011), under which training is defined as any organized and structured learning activity that occurs outside the formal education system.[5]

This definition encompasses two broad types of training: formal and informal training. Formal training is characterized by a structured curriculum and includes classroom instruction, seminars, and workshops, among other activities. Such training is typically conducted outside the immediate work environment and shows a high degree of organization by a trainer or institution. It is also generally broad in scope and not tailored to specific tasks, machinery, or job roles. Informal training, by contrast, refers to task-related learning that takes place in the workplace and often arises through instruction or interaction with coworkers. It includes guided on-the-job training, job rotation, exchanges, and other forms of learning through participation in learning circles.

In this sense, our definition of training includes all organized and structured on-the-job learning activities that take place outside the formal education system. This encompasses several important sources of human capital accumulation, such as participation in seminars or workshops and learning through instruction or interaction with coworkers. Our definition excludes formal schooling, which accounts for less than 10% of adult education activities (Ma et al., 2024). It also excludes informal learning activities such as reading journals, visiting museums, or learning through media in an unstructured or unplanned manner, as these activities are primarily self-directed and often do not involve employers.[6]

[4]Because our firm-level datasets do not contain information on worker turnover (and because turnover is itself an equilibrium outcome influenced by firms' training decisions), we rely on country-level variation in job turnover to assess the role of worker mobility and poaching risk on training.

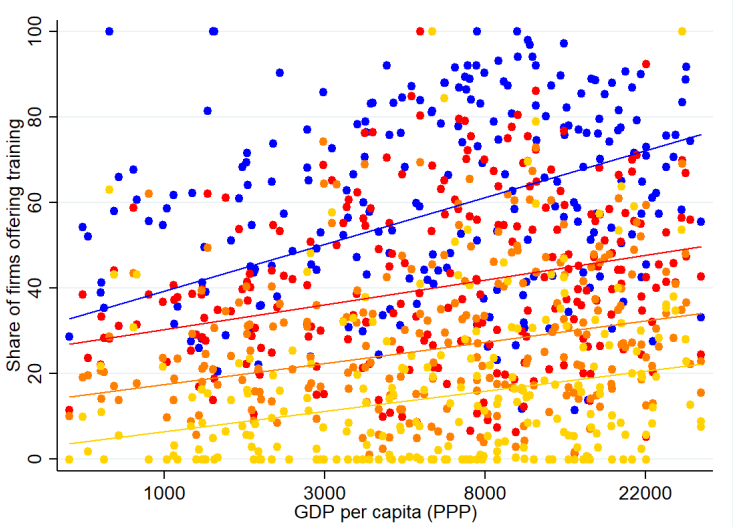
[5]The International Standard Classification of Education (ISCED), adopted by UNESCO, provides uniform and internationally agreed definitions to facilitate comparisons of education systems across countries.

[6]Our definition also excludes learning-by-doing, as we focus on forms of learning that involve a pecuniary

3.3 Link between firm size and training

We first document a strong positive relationship between firm size and the provision of formal training across countries. Figure 3.1 plots the share of firms offering formal training by firm size across countries at different levels of GDP per capita using WB-ES data. While firms of all sizes are more likely to provide training in richer countries, consistent with the existing literature, larger firms are systematically more likely to offer formal training at every level of development.

Figure 3.1: Share of firms offering formal training by firm size



Notes: Each dot in this figure represents the share of firms offering formal training in a specific firm-size category in each country. The firm-size categories considered are 1–5 (gold), 6–20 (orange), 21–100 (red), and 100+ (blue). Data on training comes from the WB-ES. Data on GDP per capita come from the Penn World Tables. Results are weighted using the observations' weights provided in the survey.

To further examine this relationship, Table B.1 presents firm-level regressions in which an indicator for whether a firm provides formal training is regressed on firm size. The estimated relationship is strongly positive and remains highly robust to the inclusion of year, country, and industry fixed effects.[7] Complementing these results, Figure B.3 plots the difference in the share of firms offering training between medium and small firms against the corresponding difference between large and medium firms. Panel (a) presents results from the WB-ES, while

and/or opportunity cost.

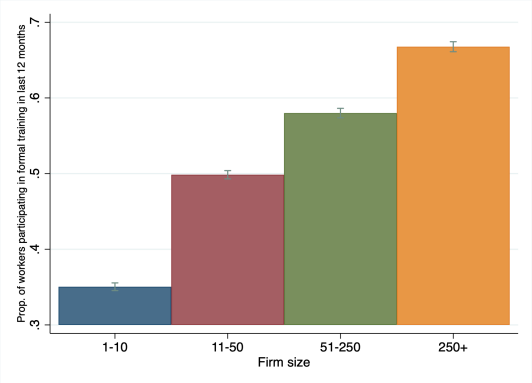
[7]The positive association between firm size and training is robust along several additional dimensions. Using EU-CVT data for 2010, Table B.2 shows that larger firms report higher participation in formal training across nearly all training purposes (e.g., general IT, management, office administration). We also find evidence along the intensive margin: Figure B.1 shows that hours spent in formal training per participant increase with firm size. Finally, Figure B.2 shows that the positive relationship between firm size and training persists when firm size is defined relative to the country-year median number of employees, rather than using absolute size categories.

panel (b) shows analogous evidence from the EU-CVT. In both datasets, 95% of country-year observations lie in the upper-right quadrant, indicating that training provision increases both from small to medium firms and from medium to large firms.

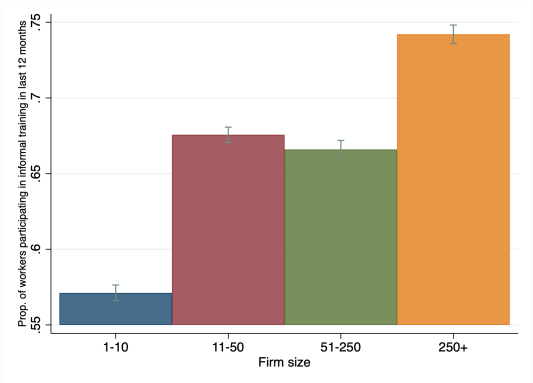
Finally, Figure 3.2 uses worker-level data from PIAAC to show that the positive relationship between firm size and on-the-job training is also present from workers' perspective, and extends to informal training. Using pooled data across all countries in the PIAAC sample, the figure plots the share of workers reporting participation in formal and informal training during the previous 12 months across firm size categories. Consistent with the firm-level evidence, workers in larger firms are more likely to engage in both formal and informal training. Table B.5 shows that these patterns are robust to controlling for demographic characteristics and wages, as well as occupation, industry, and country fixed effects.[8]

Figure 3.2: Share of workers participating in formal and informal training by firm size

(a) Formal training



(b) Informal training



Notes: This table shows the share of workers participating in formal and informal training in the last 12 months in firms of different sizes. Data comes from PIAAC and encompasses 34 countries in the OECD. Results are weighted using the observations' weights provided in the survey. 95% confidence intervals are plotted.

3.4 Link between productivity, labor share, and training

We now investigate the roles of productivity and the labor share in shaping firm-level training investments and the positive correlation between firm size and training. We use administrative firm-level data from the 2005–2007 waves of the Chinese Annual Survey of Manufacturing and the 2019 wave of the Mexican Economic Census, which allow us to measure productivity and labor shares directly.

[8]We also document evidence along the intensive margin in Figure B.4: workers in larger firms spend more hours, on average, in their most recent formal training activity than workers in smaller firms.

3.4.1 Construction of training, labor share, and TFP variables

In the Chinese Annual Survey of Manufacturing, training is measured by per-worker training expenditures, while in the Mexican Economic Census it is measured by the share of employees who received training (both formal and informal) in the past year. The labor share is calculated as the ratio of a firm's payroll to its sales. We construct firm-level measures of total factor productivity using several methods, which we describe below.

1. TFP à la Hsieh and Klenow (2009) We first compute firm-level TFP following the approach of Hsieh and Klenow (2009). This method uses firm-level data on revenue, payroll, and fixed capital and defines TFP as the residual from a constant-returns-to-scale Cobb–Douglas production function in capital and labor. As in Hsieh and Klenow (2009), and in order to control for human capital when estimating TFP, we measure labor input using payroll, which captures both wage per unit of human capital and the number of units of human capital. We consider three alternative measures for the labor share in this Cobb-Douglas function: (i) a labor share fixed at $2/3$, as suggested by cross-country evidence (Gollin, 2002) (HK1); (ii) industry-level average labor shares following Hsieh and Klenow (2009) (HK2);[9] and (iii) firm-level labor shares (HK3).

2. TFP via production function estimation à la Olley and Pakes (1996) and Levinsohn and Petrin (2003) We also compute firm-level TFP using the production function estimation methods developed by Olley and Pakes (1996) (OP) and Levinsohn and Petrin (2003) (LP). These methods combine theoretical insights with data on firms' revenue, payroll, number of employees, investment, intermediate input costs, and fixed capital stock to address the simultaneity bias in productivity estimation, which arises because productivity affects output both directly and indirectly through input choices. The OP method uses investment as a proxy for unobserved productivity shocks, while the LP method instead relies on intermediate inputs, allowing for the presence of plants with zero investment. Implementing these methods requires panel data with all variables deflated by their corresponding price indices.[10] For example, we compute real revenue using three-digit industry-level producer price indices and real investment using the capital expenditures price index. To control for human capital in the TFP estimation, we also include payroll as an input. Finally, to allow

[9] Given substantial labor market distortions in China, we follow Hsieh and Klenow (2009) and proxy China's industry-level labor shares using the corresponding U.S. industry-level labor shares.

[10] Although the Mexican Economic Census reports on-the-job training only in 2019, we also use data from the 2009 and 2014 waves to estimate TFP in the Mexican context.

for heterogeneity in production technologies, we estimate production functions separately by two-digit industry.

Our TFP measures aim to capture physical productivity (TFPQ) rather than revenue-based productivity (TFPR), given the distortions that generate variation in TFPR across firms. As noted by Foster et al. (2008), measuring TFPQ typically requires plant-specific price deflators or physical output data, which are not commonly available. Hsieh and Klenow (2009) address this issue by adopting a constant-elasticity-of-demand framework that separates firms' prices and quantities given an assumed demand elasticity. In contrast, the methods of Olley and Pakes (1996) and Levinsohn and Petrin (2003) produce a traditional productivity measure (TFPT) that lies between TFPQ and TFPR by relying on industry-level, rather than plant-specific, price deflators. Using data from industries with observed physical output, Foster et al. (2008) show that TFPT is strongly correlated with TFPQ.

Additional details on the construction of the training, TFP, labor share, and other relevant variables are provided in Appendix A.4 and Appendix A.5. In Figure A.1 and Figure A.2, we present histograms of these measures, along with firm size, for the Chinese and Mexican datasets, documenting substantial variation in both settings.

3.4.2 Specification and results

We examine the relationship between training, productivity, and the labor share using linear regressions estimated separately on the Chinese and Mexican datasets:

$$Y_{i,t} = \gamma_l \text{Labor Share}_{i,t} + \gamma_p \text{TFP}_{i,t} + \gamma_s \text{Firm size}_{i,t} + \zeta \text{Age}_{i,t} + \alpha_{\text{Ind}(i,t)} + \text{AccessCredit}_{i,t} + \eta_t + \xi_i + \epsilon_{i,t} \quad \text{(Chinese data)},$$

$$Y_{i,t} = \gamma_l \text{Labor share}_{i,t} + \gamma_p \text{TFP}_{i,t} + \gamma_s \text{Firm size}_{i,t} + \zeta \text{Age}_{i,t} + \alpha_{\text{Ind}(i,t)} + \text{AccessCredit}_{i,t} + \epsilon_{i,t} \quad \text{(Mexican data)}.$$

$Y_{i,t}$ is an indicator equal to one if firm i offers training to its workers in period t.[11] $\text{Labor Share}_{i,t}$ and $\text{TFP}_{i,t}$ denote firm i's labor share and log TFP in period t, respectively. To better isolate the relationship between productivity, the labor share, and training, we control for firm size, as well as firm age and industry fixed effects, and additionally include year and firm fixed effects in the Chinese specification.[12] In addition, to account for potential differences in firms' financial constraints that may limit their ability to finance training investments,

[11] We focus on the extensive margin of training in the main text to ensure comparability across surveys. In Appendix B.3, we show that the results are broadly robust to intensive-margin measures, namely per-worker training expenditures in China and the share of workers offered training in Mexico.

[12] Including firm fixed effects allows us to control for time-invariant firm characteristics and exploit within-firm variation in productivity and labor share over time.

we also include controls capturing firms' access to credit at time t. In China, we control for leverage, measured as the ratio of liabilities to capital. In Mexico, where balance-sheet liabilities are unavailable, we control for self-reported difficulty in accessing credit and for whether the firm reports having obtained credit during the previous year.

In Tables 3.1 and 3.2, we present the results of these regressions using the TFP measures constructed following Hsieh and Klenow (2009) and the production-function approaches of Olley and Pakes (1996) and Levinsohn and Petrin (2003), respectively. In both China and Mexico, and for nearly all specifications, we find that firms with higher labor shares are systematically less likely to offer training, while firms with higher TFP are more likely to do so. Specifically, a 10 percentage point increase in the labor share is associated with a 0.2–3 percentage point reduction in the probability that a firm offers training. By contrast, a 10 percent increase in TFP is associated with a 0.02–0.35 percentage point increase in the probability of training provision.[13]

Table 3.1: Correlation between training, TFP (HK), and the labor share

Dep. variable:	China Firm offers training						Mexico Firm offers training		
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)
Labor share	-0.152***	-0.027**	-0.191***	-0.033**	-0.210***	-0.106***	-0.029***	-0.099***	-0.099***
	(0.008)	(0.013)	(0.008)	(0.013)	(0.008)	(0.014)	(0.002)	(0.002)	(0.002)
TFP (HK1)	0.007***	0.010***					0.015***		
	(0.001)	(0.001)					(0.0003)		
TFP (HK2)			0.003***	0.009***				0.002***	
			(0.001)	(0.001)				(0.0002)	
TFP (HK3)					-0.004***	0.008***			0.002***
					(0.001)	(0.001)			(0.0002)
Log firm size	0.078***	0.038***	0.080***	0.037***	0.082***	0.035***	0.093***	0.101***	0.101***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)	(0.0003)	(0.0003)	(0.0003)
Age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y			
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Credit access control	Y	Y	Y	Y	Y	Y			
Firm FE		Y		Y		Y			
Observations	772,757	658,590	768,469	654,824	773,878	660,180	1,560,628	1,557,759	1,560,623
R-squared	0.065	0.704	0.065	0.704	0.065	0.704	0.214	0.213	0.212

Notes: This table shows different specifications in which we regress an indicator of whether the firm provides training to at least some of its workers on firm size, labor share, and different measures of TFP constructed using the methodology of Hsieh and Klenow (2009). The TFP and labor share measures are described in Appendix A.4 and Appendix A.5. Industry FE corresponds to four-digit industries in both China and Mexico. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

[13] Consistent patterns emerge in the cross-country data. In Table B.3, using the WB-ES, we regress an indicator for training provision on the labor share (measured as the ratio of total compensation to sales), firm size, and additional controls. We find that a 10 percentage point increase in the labor share is associated with approximately a 1 percentage point decrease in the likelihood that a firm offers training.

Since larger firms tend to have lower labor shares and higher productivity, these results suggest that part of the positive correlation between training and firm size observed in the cross-country data operates through these channels.[14] Nevertheless, the positive relationship between firm size and training remains large and statistically significant across all specifications, indicating that differences in productivity and labor shares do not fully account for the firm-size gradient in training.

Table 3.2: Correlation between training, TFP (OP & LP), and the labor share

Dep. variable:	China Firm offers training				Mexico Firm offers training	
	(1)	(2)	(3)	(4)	(1)	(2)
Labor share	-0.301***	-0.037***	-0.180***	-0.021	-0.089***	-0.093***
	(0.009)	(0.014)	(0.009)	(0.014)	(0.007)	(0.002)
TFP (OP)	-0.015***	0.006***			0.031***	
	(0.001)	(0.001)			(0.009)	
TFP (LP)			0.005***	0.013***		0.035***
			(0.001)	(0.001)		(0.002)
Log firm size	0.083***	0.036***	0.078***	0.033***	0.14***	0.11***
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.0004)
Age FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y		
Industry FE	Y	Y	Y	Y	Y	Y
Credit access control	Y	Y	Y	Y	Y	Y
Firm FE		Y		Y		
Observations	771,902	657,499	771,903	658,111	122,340	719,120
R-squared	0.065	0.704	0.062	0.703	0.313	0.202

Notes: This table shows different specifications in which we regress an indicator of whether the firm provides training to at least some of its workers on firm size, labor share, and different measures of TFP constructed using the methodologies of Olley and Pakes (1996) and Levinsohn and Petrin (2003). The TFP and labor share measures are described in Appendix A.4 and Appendix A.5. Industry FE corresponds to four-digit industries in both China and Mexico. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

3.5 Link between job turnover and training

Finally, we explore the link between job turnover and training investments, motivated by the hold-up problem and the idea that firms invest in training only when they can retain workers

[14]The positive correlation between firm size and productivity has been documented in a variety of settings (see, for example, Oi and Idson, 1999; Hsieh and Klenow, 2009, 2014; Syverson, 2011). Similarly, labor shares have been shown to decline with firm size across a range of countries and contexts (Karabarbounis and Neiman, 2014; Grullon et al., 2019; Autor et al., 2020; Barkai, 2020; Gouin-Bonenfant, 2022). We provide additional cross-country evidence for this pattern using a pooled sample of firms from over 100 countries in the WB-ES, as shown in Table B.4. Consistent with the existing literature, we find that larger firms tend to have lower labor shares.

long enough to capture the returns from that investment. To do so, we combine country-level variation in job turnover with firm-level training investments from the EU-CVT to assess the role of worker mobility and poaching risk in shaping training provision.

Specifically, we regress the share of formally trained workers in each country-year in the EU-CVT on the predicted probability that a worker remains in the same job after one quarter, controlling for GDP per capita and year fixed effects. The results are reported in Table 3.3. Ideally, we would directly use the cross-country job turnover measures constructed by Donovan et al. (2023) in this regression. However, the timing of these measures does not fully overlap with the EU-CVT. We therefore predict job turnover rates for our country-years of interest by exploiting the documented link between job turnover, the ease of contract termination, and broader institutional quality (Bassanini and Duval, 2009; Haltiwanger et al., 2014). In particular, we regress the turnover measures from Donovan et al. (2023) on a set of institutional indicators from the World Bank Worldwide Governance Indicators (Kaufmann and Kraay, 2024), encompassing voice and accountability, political stability, regulatory quality, rule of law, and control of corruption, and use the predicted turnover measures in the training regression. We provide further details on the job turnover measures, the governance indicators, and their relationship in Appendix A.6.

Table 3.3: Correlation between job turnover and training

Dep. variable:	Proportion of workers exposed to formal training
	(1)
Log (GDP per capita)	7.86***
	(0.78)
Probability same job after quarter	20.4**
	(8.81)
Constant	-72.2***
	(8.13)
Year FE	Y
Observations	208
R-squared	0.617

Notes: This table shows the results from regressing the share of workers exposed to formal training in each country-year in the EU-CVT data on the predicted probability of staying in the same job after a quarter, log GDP per capita (PPP), and year fixed effects. Since the timing of the job turnover rates built by Donovan et al. (2023) does not match that of our training data, we predict job turnover rates for our country-years of interest by regressing job turnover measures on the following institutional measures from the World Bank Worldwide Governance Indicators: voice and accountability, political stability, regulatory quality, rule of law, and control of corruption. The data on GDP per capita comes from the Penn World Table. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In Table 3.3, we find that lower predicted job turnover—i.e., a higher likelihood of workers

remaining in the same job—is associated with a significantly greater share of trained workers. Because job turnover is predicted using institutional indicators, this evidence is suggestive, but it is broadly consistent with firms being less inclined to invest in training when the risk of worker separation is higher.

Since larger firms tend to pay higher wages and experience lower turnover, these results suggest that part of the positive relationship between firm size and training documented earlier operates through the job turnover and hold-up channel.[15] In particular, weaker worker retention reduces small firms' incentives to invest in training, reinforcing the role of the hold-up problem in depressing training investments when separation risk is high.

3.6 Summary and taking stock

Taken together, the results on firm size, firm productivity, the labor share, and job turnover point to four central patterns shaping training investments. First, they support the presence of complementarities between firm productivity and workers' human capital, as more productive firms are systematically more likely to invest in training. Second, the negative relationship between the labor share and training indicates that when a larger share of revenue accrues to workers, firms have weaker incentives to invest in training.[16] Third, even conditional on productivity and the labor share, training remains substantially higher in larger firms, suggesting that large firms have additional incentives to invest in training beyond productivity and revenue-sharing considerations. Finally, higher job turnover depresses training investments, consistent with a hold-up problem whose severity varies across firms depending on their ability to retain workers.

4 Analytical model

We now construct an analytical model that formalizes how the incentives for training vary with firm characteristics along the development path and embeds the mechanisms highlighted in the empirical analysis above. This model also offers economic insights into the mechanisms driving the results in the quantitative model of Section 5.

[15] Although direct evidence jointly linking firm size, training, and turnover is limited, existing studies document that larger firms tend to exhibit lower job turnover across countries (e.g., Haltiwanger et al., 2014), and that higher turnover is associated with reduced training in the United States (e.g., Frazis et al., 2000).

[16] Furthermore, this pattern is difficult to reconcile with a worker-driven training mechanism, which would instead predict a positive relationship between workers' revenue shares and training. Instead, it is consistent with firms playing a central role in training decisions.

The model describes an economy with firm productivity heterogeneity and labor market frictions, following Burdett and Mortensen (1998). Workers live for two periods and accumulate human capital through on-the-job training when matched with firms. Firms post vacancies and wages per efficiency unit to attract workers.[17] Training investments are determined and financed after the worker and firm match. In the second period of life, some workers separate exogenously, while others engage in on-the-job search and switch firms when they receive a higher wage offer. In what follows, we focus on the stationary equilibrium of the model, in which the firm-level distributions of workers' age and human capital are constant over time, and firms with the same productivity level are symmetric.

4.1 Workers

The economy is populated by a continuum of workers, each of whom lives for two periods. Workers are born identical but may accumulate human capital at different rates through on-the-job training. Workers supply one unit of labor inelastically in each period.

Workers have linear utility and maximize the present value of consumption:

$$\max_{\{c^Y, c^O\}} \; c^Y + \frac{c^O}{1+\rho}, \quad s.t. \quad c^Y + \frac{c^O}{1+r} = w^Y + \frac{w^O}{1+r},$$

where superscripts Y and O denote young and old ages, respectively, and $\rho > 0$ governs time preference.[18] The consumption good is the numeraire. In the steady state, $\rho = r$, implying that workers are indifferent between consuming when young or old. The population size of each generation is normalized to one.

4.2 Production

There is a unit measure of firms that are heterogeneous in productivity, $z \sim G(z)$, and produce a homogeneous good used for consumption and for financing training and vacancy costs. Consistent with the literature and our empirical evidence on the positive association between TFP and training, we assume that firm productivity and worker human capital are complementary (Acemoglu and Pischke, 1998; Bagger et al., 2014). Once workers and firms are matched, the output produced by worker i at firm j is given by

$$y_{i,j} = A z_j h_i,$$

[17] This wage posting setup allows the labor share to differ across firms.
[18] Wages w^Y and w^O entering the utility function are net of any training costs paid by workers.

where A denotes economy-wide productivity, z_j denotes firm-specific productivity, and h_i denotes worker i's efficiency units of labor (human capital).

Since we focus on a stationary equilibrium, in what follows we index firms and their choices solely by their specific productivity level, z.

4.3 Job search and matching

Firms post vacancies $v(z)$ at the start of each period, together with a contract specifying the wage rate per efficiency unit $w(z)$. This wage rate determines the labor share in firm z, given by $\beta(z) = w(z)/z$.

Posting vacancies is costly. Firms incur a vacancy posting cost given by $c_v \bar{w} \frac{v(z)^{1+\gamma_v}}{1+\gamma_v}$, which is strictly convex ($\gamma_v > 0$), following Acemoglu and Hawkins (2014), to ensure the coexistence of firms with different productivity levels in equilibrium. We scale vacancy costs by the average wage $\bar{w}$ to reflect the cost of human resources staff. The total number of vacancies is $V = \int v(z)\, dG(z)$, and the distribution of posted wage offers faced by job seekers is $F(w) = \frac{1}{V} \int_{w(z)<w} v(z)\, dG(z)$.

Each successful job match remains in place until it is either exogenously destroyed, endogenously terminated, or the worker retires. At the start of the second period of life, there is an exogenous separation probability δ. Separated workers enter the unemployment pool together with newly born workers. In addition, a fraction η of the remaining employed old workers engage in on-the-job search and switch firms if they receive a higher wage offer. The total mass of job seekers is therefore given by $\tilde{U} = 1 + \delta + \eta(1 - \delta)$, which includes new entrants, exogenously separated workers, and employed workers searching on the job.

For analytical tractability, and since our focus is on characterizing training investments across firms, in this analytical model we assume a simple matching technology, $M(\tilde{U}, V) = \min\{\tilde{U}, V\}$, and assume c_v is small enough such that $V > \tilde{U}$, which ensures full employment in equilibrium. As usual, market tightness is defined by $\theta = \frac{V}{\tilde{U}}$.

4.4 Training investments

A young worker starts with an initial level of human capital $h^Y = 1$ (normalization). Training for s units of time increases next-period human capital according to

$$h^O = h^Y + \zeta L(z)^{\alpha} s^{\gamma_s}.$$

$\zeta > 0$ governs the level of training returns, while $0 < \gamma_s < 1$ governs its diminishing returns. When $\alpha > 0$, training returns depend positively on firm size $L(z)$, capturing economies of scale in training. This formulation reflects both our empirical finding that training is more prevalent in larger firms and existing evidence that larger firms may provide better learning environments. Intuitively, larger firms benefit from greater knowledge diffusion among workers and can more easily support higher-quality training infrastructure due to a larger pool of trainees. This assumption is consistent with recent evidence documenting firm heterogeneity in learning environments (Gregory, 2019; Arellano-Bover and Saltiel, 2021), as well as findings of faster returns to experience in larger firms (e.g., Arellano-Bover, 2024).

Training entails two types of costs. First, there are direct costs equal to $c_s\bar{w}$ per unit of training time, proportional to the average wage $\bar{w}$, capturing expenses such as payments to trainers. Second, training involves an opportunity cost, as each unit of training time reduces production time one-for-one.[19] We assume that training raises general human capital, so its benefits accrue even if the worker subsequently changes firms.[20]

4.4.1 Allocation of training costs and optimal training levels

Firms and workers each pay constant shares of training costs, denoted by μ_F and μ_W, respectively, with $\mu_F + \mu_W = 1$. Given these cost shares, both parties independently choose their desired training levels, and the realized training level is determined by the party with the lower desired level—that is, the party with lower willingness to pay. This reflects the requirement of mutual consent for training investments, even when costs are shared. For example, workers may prefer a six-month training program, but if the firm is only willing to support two months, training is limited to that level.

Given this setup, the training level in firm z is given by

$$s(z) = \min\{s_F(z), s_W(z)\} = \min\left\{ \left(\frac{\zeta\gamma_s L(z)^\alpha MR_F(z)}{(1+r)\mu_F(c_s\bar{w} + Az)} \right)^{\frac{1}{1-\gamma_s}}, \left(\frac{\zeta\gamma_s L(z)^\alpha MR_W(z)}{(1+r)\mu_W(c_s\bar{w} + Az)} \right)^{\frac{1}{1-\gamma_s}} \right\}, \tag{1}$$

where $s_F(z)$ and $s_W(z)$ denote the training levels desired by the firm and the worker, respec-

[19] We assume training costs are linear in training time, reflecting that both trainer expenses and foregone production are likely proportional to time spent training.

[20] We focus on general human capital since firm-specific human capital has been shown to account for a relatively small share of wage growth (Altonji and Shakotko, 1987; Kambourov and Manovskii, 2009; Lazear, 2009), and because questions about firm incentives to invest in training and hold-up concerns arise precisely when training generates transferable skills.

tively, and $MR_F(z)$ and $MR_W(z)$ denote their respective marginal benefits from additional human capital in the next period, as characterized in Appendix C.1.

Workers' gains from additional human capital, $MR_W(z)$, depend on expected wage flows both if they remain with firm z and if they switch employers. By contrast, firms' gains, $MR_F(z)$, accrue only if the match continues, since firms do not benefit from training once the worker leaves. This asymmetry generates a hold-up problem, whereby firms underinvest in training due to the possibility that trained workers separate. As a result, in most cases (including in the calibrated quantitative economy), training investments are determined by firms, which typically prefer lower training levels than workers.[21]

Combining this observation with the expression for $MR_F(z)$ yields the following expression for training in firm z:

$$s(z) = \left(\frac{\zeta \gamma_s L(z)^\alpha MR_F(z)}{(1+r)\mu_F(c_s \bar{w} + Az)} \right)^{\frac{1}{1-\gamma_s}} = \left(\underbrace{\frac{\zeta \gamma_s L(z)^\alpha}{(1+r)\mu_F(c_s \bar{w} + Az)}}_{\text{returns \& costs from training}} \underbrace{(1-\delta)\left[1 - \eta \bar{F}(w(z))\right]}_{\text{prob of keeping worker}} \underbrace{(1-\beta(z))}_{\text{1-labor share}} \underbrace{Az}_{\text{productivity}} \right)^{\frac{1}{1-\gamma_s}}. \tag{2}$$

The first term on the right-hand side captures the net marginal human capital gains from training and its costs. The second term reflects expected job turnover, which reduces firms' incentives to invest in training and captures the hold-up problem. The final two terms capture firms' perceived returns to human capital in production, combining firm productivity with the share of output retained by the firm.

This expression highlights the four channels through which firm heterogeneity shapes training investments in the model. First, higher firm productivity z increases the marginal return to human capital in production. Second, a larger workforce $L(z)$ increases the returns to training in human capital when $\alpha > 0$. Third, a lower labor share $\beta(z)$ allows firms to appropriate a larger share of training returns. Fourth, on-the-job search induces differential worker poaching rates across firms, captured by the term $\eta \bar{F}(w(z))$, causing the severity of the hold-up to vary with wages along the productivity distribution of firms.

[21] To further support this result, Appendix C.2 examines how the model's predictions change under alternative assumptions about how training investments are determined and funded. We show that when firms determine training, the model's predictions align most closely with our empirical findings. By contrast, alternative arrangements, such as workers setting training levels themselves or training chosen to maximize the match value, generate predictions that are inconsistent with the observed results. We also provide supporting evidence that firms play a pivotal role in training decisions and financing using the 2011 and 2016 EU Adult Education Surveys (EU-AES). In Appendix C.3, we document that over 65% of trained workers across firms of all sizes report that their employer financed their training costs at least in part. Further details on the EU-AES data are provided in Appendix A.8.

4.5 Solving the firms' problem

Given young workers' training $s(z)$ as discussed above, a firm with productivity z chooses the wage rate $w(z)$ and the number of vacancies $v(z)$ to maximize the total expected value from hiring:

$$\max_{\{w(z),v(z)\}} \underbrace{\frac{v(z)}{\theta}\frac{1}{1+\eta(1-\delta)+\delta}\left[Az - w(z) - \mu_F(c_s\bar{w}+Az)s(z) + \frac{1}{1+r}MR_F(z)(1+\zeta L(z)^{\alpha}s(z)^{\gamma_s})\right]}_{\text{profits from hiring young workers}}$$
$$+ \underbrace{\frac{v(z)}{\theta}\frac{\eta(1-\delta)+\delta}{1+\eta(1-\delta)+\delta}\frac{\eta(1-\delta)F(w(z))\bar{l}(w(z))+\delta\bar{l}}{\eta(1-\delta)+\delta}(Az-w(z))}_{\text{profits from hiring old workers}} - \underbrace{\frac{c_v\bar{w}v(z)^{1+\gamma_v}}{1+\gamma_v}}_{\text{vacancy costs}} \tag{3}$$
$$\text{s.t.} \quad w(z) \geq b\bar{w},$$

The first term captures the net profits from hiring young workers. The term $\frac{v(z)}{\theta}\frac{1}{1+\eta(1-\delta)+\delta}$ denotes the number of young workers met by a firm posting $v(z)$ vacancies, while the expression in brackets combines current-period profits with expected next-period profits from the hire. It is worth noting that these profits account for expected training costs, implying that firms may partially offload training costs by posting lower wages.

The second term captures the profits from hiring old workers, either by poaching them from other firms or hiring them from unemployment. The term $\frac{v(z)}{\theta}\frac{\eta(1-\delta)+\delta}{1+\eta(1-\delta)+\delta}$ denotes the number of old workers met by the firm. Conditional on receiving an offer from firm z, on-the-job movers accept with probability $F(w(z))$, and bring, on average, $\bar{l}(w(z))$ efficiency units of labor (human capital); while unemployed workers accept with probability one and supply $\bar{l}$ efficiency units.[22]

The final term captures total vacancy posting costs.

We solve $w(z)$ and $v(z)$ using the first-order conditions of Equation (3). The optimal wage satisfies a first-order differential equation subject to the wage constraint $w(z) \geq b\bar{w}$ as in Burdett and Mortensen (1998).[23]

[22] The average efficiency units supplied by on-the-job movers are given by $\bar{l}(w) = 1 + \frac{\int_0^{w(z)} \zeta L(u)^{\alpha}s(u)^{\gamma_s}dF(w(u))}{F(w(z))}$, whereas for unemployed workers they are given by $\bar{l} = 1 + \int \zeta L(z)^{\alpha}s(z)^{\gamma_s}dF(w(z))$.

[23] As shown by Hornstein et al. (2011), search-and-matching models with reasonable unemployment benefits struggle to generate the observed degree of frictional wage dispersion. Given our focus on training decisions, we instead match wage dispersion by assuming the lowest wage to be $w_{\min} = b\bar{w}$, where $\bar{w}$ denotes the average wage and b is a constant. We assume unemployed workers accept all offers, consistent with evidence that offer acceptance rates are close to one (van den Berg, 1990). Because under these assumptions unemployment benefits do not affect any other equilibrium outcomes, we abstract from them in the model.

Intuitively, firms raise wages to poach workers and retain trained employees. However, higher wages also reduce profits. The equilibrium wage distribution reflects the balance between these forces, with more productive firms offering higher wages (yet lower labor shares) and thus being more insulated from poaching, as retaining workers is relatively more valuable. Vacancy posting is determined by equating the marginal gains from hiring to the marginal cost of posting vacancies, which rise with the number of vacancies.

4.6 Equilibrium

To close the model, we assume that firm owners use net profits, defined as revenues net of wages, vacancy costs, and training costs borne by the firm, for consumption. We now define the model's steady-state general equilibrium.

Definition 4.1. The general equilibrium for this economy is a set of:

1. workers' consumption choices $\{c^Y, c^O\}$;
2. firms' wage and vacancy posting decisions $\{w(z), v(z)\}$;
3. training decisions $\{s(z)\}$; and
4. an offer distribution $F(w)$ and labor market tightness θ; such that:

(i) given labor market tightness, the offer distribution, and training decisions, workers' consumption choices solve their utility maximization problems;

(ii) given labor market tightness, the offer distribution, and training decisions, firms' wage and vacancy posting decisions solve the firm's problem;

(iii) given labor market tightness and the offer distribution, firms and workers choose training optimally, and the realized training level is determined by the rule described in Equation (1);

(iv) labor market tightness and the offer distribution are consistent with workers' job transitions and firms' wage and vacancy posting decisions; and

(v) total output equals the sum of consumption, vacancy costs, and training costs.

5 Quantitative model and calibration

We now extend the model for quantitative analysis and take it to the data.

5.1 Setup

We extend the analytical setup to more closely replicate key features of labor markets and economic environments across countries. These extensions are described below.

Workers Workers live for $J > 2$ periods, and human capital accumulated due to training depreciates at rate d each period, consistent with empirical evidence (e.g., Mincer, 1989; Blundell et al., 2021). Human capital is bounded below by the level individuals are endowed with at birth, capturing basic cognitive and physical skills.

Firms Firm productivity is Pareto-distributed, $G(z) = 1 - z^{-\kappa}$, consistent with empirical evidence on the firm size distribution (Axtell, 2001).

Matching function We use the standard matching function $M(\tilde{U}, V) = c_M \tilde{U}^{\psi} V^{1-\psi}$, which generates positive unemployment and can be calibrated to deliver reasonable elasticities of matches with respect to the number of searchers $\tilde{U}$ and vacancies V.

Conditions for simulations The optimality conditions of the quantitative model mirror the intuition from the analytical framework and are presented in Appendix D.1. There, we characterize optimal training choices as functions of firm productivity and worker age. The appendix also details firms' wage-setting and vacancy-posting decisions.

5.2 Model calibration

Our calibration procedure proceeds in two steps. First, we calibrate the model to the U.S. economy, which serves as a benchmark. Second, we recalibrate the model for representative economies at six different income levels to assess how firms shape on-the-job training along the development path. In this second step, we jointly recalibrate a subset of parameters associated with our mechanisms of interest.

5.2.1 Calibrating the model to the United States

We calibrate the model to the U.S. economy using quarterly data. Some parameters are set directly based on values from the literature (external calibration), while the remaining parameters are calibrated internally to match key data moments.

Externally calibrated parameters

We draw some common parameters directly from the literature, as presented in Table 5.1. A period in the model is one quarter. We set the quarterly discount rate ρ to 0.01 such that the annualized interest rate is 0.04. Each individual works for 40 years, and therefore the lifetime length is set to $J = 160$ quarters. The ratio of the lowest wage to the average wage is $b = 0.6$ following Hornstein et al. (2011), who calculate the mean-min ratio of wages to be around 1.7 from U.S. labor data. We choose the elasticity of the number of matches to the number of searchers in the matching function to be $\psi = 0.7$, as estimated by Shimer (2005) using the U.S. labor flow and vacancy data. Since we lack estimates for the U.S., we set the depreciation rate of human capital at 2%, following the estimate by Blundell et al. (2021) based on British data. We set $\gamma_v = 1$, implying quadratic vacancy costs, following Acemoglu and Hawkins (2014) who study the hiring behavior of multi-worker firms in the U.S..

Table 5.1: **Externally calibrated parameters**

Parameter	Model	Source
ρ - Discount rate	0.01	Annualized interest rate of 0.04
J - Number of periods	160	40 years of work
b - Ratio of lowest wage to average wage	0.6	Hornstein et al. (2011)
ψ - Elasticity of matches to searchers	0.7	Shimer (2005)
d - Depreciation rate of human capital	0.02	Blundell et al. (2021)
γ_v - Convexity of vacancy costs	1	Acemoglu and Hawkins (2014)

Internally calibrated parameters

Procedure To calibrate the remaining parameters, we use the method of moments, minimizing the squared differences between model-generated and observed data moments. The internally calibrated parameters encompass: the constant in the matching function, c_M; the on-the-job search intensity, η; the costs per unit time of training as a share of the average wage rate, c_s; the constant in vacancy costs, c_v; the constant in training returns, ζ; the convexity in training returns, γ_s; the shape parameter of Pareto productivity distribution, κ; the exogenous separation rate, δ; the share of training costs borne by the firm, μ_F; and the size gradient of training returns, α.

To calibrate these parameters, we target the following moments of the U.S. economy for the 1994–2007 period (or from the closest available years): the average unemployment rate for the period 1994–2007; the ratio of the number of vacancies to the number of unemployed people

from FRED in 2000 to 2007; the Pareto parameter of the firm employment distribution as estimated by Axtell (2001); the share of employed people remaining in the same job after one quarter, and the share of employed people remaining employed after one quarter, which are taken from Donovan et al. (2023); the share of training time in total working hours; the ratio of training time in firms with 100–499 employees to that of firms with 50–99 employees; the ratios of training costs to wage costs of training, and of training time of old versus young workers, as reported in the 1995 Survey of Employer Provided Training (US-SEPT);[24] and the percent wage growth at 40 years' experience for U.S. workers, as estimated by Lagakos et al. (2018).

Calibration results and validation In Table 5.2, we present a comparison of the targeted model and data moments, showing that the model does fairly well in matching the targeted moments. We report the calibrated parameter values in Table 5.3, which are reasonable compared with the literature. Our calibration implies a monthly separation rate of 2.3%. Using the CPS, Shimer (2012) finds this to be 2–4% for all U.S. workers in the period 1994–2007. γ_s captures the diminishing returns of human capital investments (in terms of effective hours) in producing new human capital, and its calibrated value, γ_s =0.27, is close to the estimates in the literature. For instance, Imai and Keane (2004) find this parameter to be 0.22 in the U.S., while Manuelli and Seshadri (2014) estimate this parameter to be 0.48.

We now turn to training returns, captured in the model by the term $\zeta L^{\alpha}(z)s^{\gamma_s}$. Based on this specification and our calibrated parameter values, providing a young worker with a full quarter of training (480 working hours, corresponding to $s = 1$) increases hourly wages by 7.5 percent as a weighted average across firms. This magnitude falls within the range of empirical estimates of U.S. training returns reviewed by Leuven (2004) and Bassanini et al. (2005). For instance, Frazis and Loewenstein (2005) find that 60 hours of formal training raise wages by 3–5 percent. This is consistent with our calibration: in an average-sized firm, 60 hours of training (approximately 12.5 percent of a quarter) implies an increase in hourly wages of about 4 percent.

[24]The 1995 US-SEPT was conducted by the Bureau of Labor Statistics (BLS) and collected information from employers and randomly selected employees in establishments with 50 or more workers. The employer portion of the survey focuses on the intensity and costs of employer-provided formal training. The employee portion of the survey focuses on the time that employees spent on training. This survey provides a sample of 1,062 establishments and over 1,000 employees covering all nine major industry classifications across all 50 states. For further details on this data please see Appendix A.7. We define old workers as those aged 55 and above in the data. Since the early retirement age of U.S. workers is age 62, in the model we classify older workers as individuals in the final seven years of their working life.

Table 5.2: **Moments in the model vs data**

Moments	Data	Model
Moments: labor market		
Unemployment rate (%)	6.5	6.6
Ratio of #Vacancies to #Unemployed	0.55	0.55
Pareto parameter of firm size distribution	1.06	1.10
Share of employed people remaining in the same job after one quarter	0.88	0.88
Share of employed people remaining employed after one quarter	0.94	0.94
Moments: training intensity		
Average training intensity (% time)	2.20	2.23
Ratio of training intensity in firms with 100–499 to 50–99 employees	1.19	1.22
Ratio of training costs to wage costs of training	0.24	0.24
Ratio of old to young workers' training time	0.51	0.50
Percent wage increase of 40 years' experience (%)	89	90

Notes: The sources of the moments are described in the main text.

Table 5.3: **Calibrated parameters**

Parameter	Value
c_M - Constant in matching function	0.85
η - On-the-job search intensity	0.30
c_s - Ratio of training costs per time to wage	0.27
c_v - Constant in vacancy function	0.40
ζ - Constant in training function	0.08
γ_s - Convexity of training function	0.27
κ - Parameter of Pareto productivity dist	8.36
δ - Exogenous separation rate	0.07
$\mu_F(z)$ - Share of training costs paid by firm	0.31
α - Training returns due to larger firm size	0.10

5.2.2 Cross-country calibration

Procedure We recalibrate the model to match representative economies at six different levels of GDP per capita (5%, 10%, 20%, 40%, 60%, and 80% of the U.S. level). Most baseline parameters are held fixed at their U.S. values, while we jointly recalibrate a subset of parameters that govern the mechanisms highlighted in both the empirical analysis and the analytical model, and that shape firms' incentives to invest in training and, in turn, aggregate training outcomes. We group the recalibrated parameters into three blocks:[25]

[25] Because changes in these parameters also affect equilibrium labor shares—which in turn govern training returns—we additionally adjust b in each recalibrated economy so that labor shares are consistent with

- **Firm productivity and size distribution**: common productivity A, the Pareto tail parameter κ, and vacancy convexity γ_v;
- **Labor market turnover**: the separation rate δ and on-the-job search intensity η;
- **Training returns**: the level of training returns ζ and the elasticity of training returns with respect to firm size α.

These three parameter blocks correspond to the mechanisms highlighted in the empirical analysis and formalized in the analytical model. First, to capture productivity–training complementarities and firms' ability to appropriate training returns, we vary parameters governing the level and shape of the firm productivity distribution, together with vacancy posting costs that discipline firm size. In the model, firms' own productivity governs the marginal return to human capital in production, while the shape of the productivity and size distribution disciplines wage posting and thus labor shares and the fraction of training returns firms can appropriate. Second, to capture the role of job turnover and poaching risk, we vary separation rates and on-the-job search intensity, which govern expected match duration and the severity of the hold-up problem. Third, to capture differences in learning environments and how these vary with firm size, we vary parameters governing the level and scale of training returns, which shape the effectiveness with which training time is transformed into human capital, both on average and across firms.

The moments jointly targeted for each representative economy include the aggregate training level;[26] the relative training intensity of above-median firms compared to below-median firms (where the median is evaluated using employment size); the relative value added per worker of above-median firms compared to below-median firms; the relative size of above-median firms compared to below-median firms; the share of workers who remain employed after one quarter; the share of workers who remain in the same job after one quarter; and GDP per capita relative to the United States. We compute the first four moments using the WB-ES data, and obtain the worker retention rates from Donovan et al. (2023). Because these moments are country-specific, we proceed as follows: for each moment, we first estimate a regression of country-level values on log GDP per capita, and then use the fitted relationship to predict the corresponding moments at different income levels (5%, 10%, 20%, 40%, 60%,

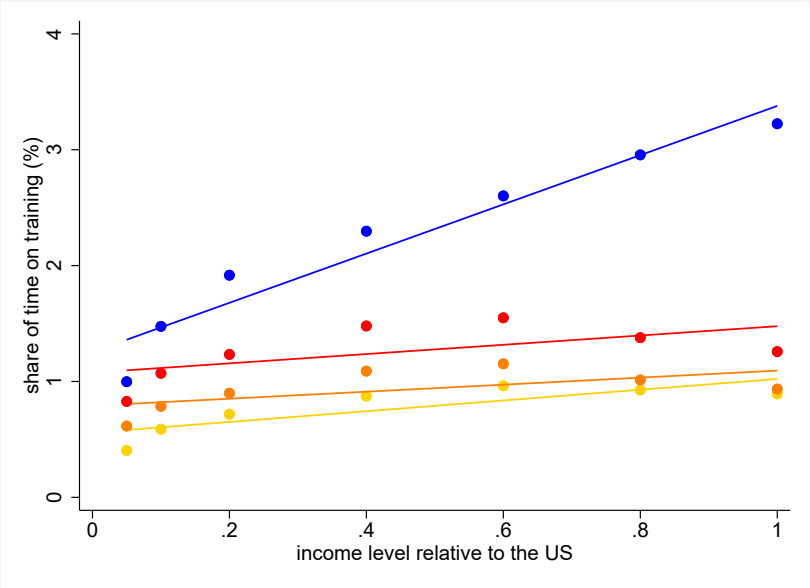
the empirical relationship between labor share and GDP documented in Gollin (2002).

[26] The aggregate level of training primarily serves to discipline country-specific training returns. Accordingly, training returns can be interpreted as a "residual" that captures the component of aggregate training not explained by other model factors, analogous to the role of total factor productivity in development accounting.

and 80% of the U.S. level).

Calibration results and validation Figure D.1 reports the values of the recalibrated parameters in different economies. We find that the level parameter governing training returns is broadly similar across countries, while richer countries exhibit a higher elasticity of training returns with respect to firm size, suggesting stronger economies of scale in training in high-income economies. Consistent with the development accounting literature, aggregate productivity also increases with income. We also find that the Pareto shape parameter of the productivity distribution is lower in poorer countries, implying greater productivity dispersion, in line with recent evidence in the literature (Restuccia, 2025). In addition, vacancy posting convexity is higher in poorer countries, consistent with the well-documented barriers to firm growth in developing economies (e.g., Guner et al., 2008; Hsieh and Klenow, 2014). Finally, in line with evidence of higher job-to-unemployment and job-to-job flows in poorer countries (Donovan et al., 2023), we find that these countries exhibit higher job destruction rates and stronger on-the-job search intensity.

Figure 5.1: Average time spent on training by firm size



Notes: Each dot in this figure represents the average training time per worker for a specific firm-size category in each representative country in the model. As in the data, firm-size categories are 1–5 (gold), 6–20 (orange), 21–100 (red), and 100+ (blue). Because firm size in the model is not directly comparable to firm size in the data, we map each size category to its corresponding percentile in the WB-ES firm-size distribution for each representative country and use these percentiles to construct the corresponding model moments.

Figure 5.1 shows average training time by firm size for each representative country, analogous to Figure 3.1 in the empirical analysis. Consistent with the empirical evidence, the model

predicts that larger firms provide more training on average in all countries, and that training intensity increases with the level of economic development.

6 Quantitative results

In this section, we use the quantitative model to examine how the three channels—training returns, firm productivity and the size distribution, and labor market turnover—(i) shape training across countries at different income levels, (ii) interact with one another, and (iii) shape cross-country differences in human capital and income through their effects on training.

6.1 Effects of alternative channels on training

Our first quantitative exercise examines how our three channels contribute to the training gap between each representative country and the United States. To this end, we change the parameters associated with each channel one at a time in each economy while holding all other parameters fixed, and evaluate the resulting effects on average training time.

Table 6.1 reports the results of this exercise. The second column presents the baseline average training time among workers in each representative economy. Consistent with Figure 5.1, an average worker in the representative country with income equal to 5% of the U.S. level devotes only 0.84% of working time to training, whereas an average worker in the representative country at 80% of the U.S. income level devotes nearly three times as much time (2.13% of working time).

Table 6.1: Decomposition of training differences across countries

Income (U.S. = 1)	Training time	U.S. training returns	U.S. productivity dist.	U.S. labor market turnover
0.05	0.0084	9.11%	19.84%	163.74%
0.10	0.0118	-16.68%	14.22%	145.29%
0.20	0.0148	-19.66%	7.72%	104.41%
0.40	0.0181	-15.13%	1.91%	53.14%
0.60	0.0201	-5.35%	-5.71%	24.38%
0.80	0.0213	-6.36%	-1.81%	14.49%

Notes: Training time is measured as the share of time devoted to training in the recalibrated baseline model for each representative economy. Columns (3)–(5) report the percentage change in training time resulting from setting each corresponding group of parameters to its U.S. level, holding all other parameters fixed.

Columns (3)–(5) report the effects of changing the parameters associated with training returns, firm productivity and size distribution, and labor market turnover, respectively, to

match the United States. We find that reducing labor market turnover to U.S. levels substantially increases average training time across all economies, with particularly large effects in poorer countries. This reflects the impact of lower job destruction rates and weaker on-the-job search intensity, which reduce worker turnover, mitigate the hold-up problem, and strengthen firms' incentives to invest in training. In addition, lower separation rates also increase employment concentration in more productive firms, as these firms are less likely to lose workers while still poaching workers from less productive firms. Since more productive firms provide more training, this compositional shift further amplifies training investment. Quantitatively, lowering labor market turnover to U.S. levels raises training time by 163.74% in the poorest country (5% of U.S. income) and by 14.49% in the country with income equal to 80% of the U.S. level.

Training returns and the productivity distribution play relatively smaller roles. For the poorest country, adjusting training returns to U.S. levels increases training time by 9.11%, whereas for richer countries the same adjustment leads to a decline in training. This finding is consistent with Ma et al. (2024), which documents no clear relationship between training returns and the level of economic development. Changes in the productivity and firm size distribution shift employment toward larger firms. This generates two opposing effects. First, because larger firms face lower poaching risk, this reallocation strengthens firms' incentives to provide training. Second, since countries of lower income levels exhibit weaker economies of scale in training than the U.S., moving toward the U.S. employment structure lowers aggregate training returns. In poorer countries, where small firms are more prevalent, the first force dominates, leading to positive effects of these distributional changes, whereas in richer countries the second force dominates, resulting in muted or even negative effects.

6.2 Interaction effects

In the previous subsection, we examined the effects of changing each channel in isolation, and highlighted the importance of labor market turnover in explaining the training gap across countries. We now study how the two other channels interact with labor market turnover, and jointly shape training outcomes across countries.

Table 6.2 reports the impact of reducing labor market turnover to U.S. levels on training time in each country's baseline model, as well as under alternative scenarios in which one or both of the other channels are also set to U.S. levels. We find that adjusting training returns to U.S. levels substantially amplifies the training gains from reducing labor market

turnover in all countries, with particularly large effects in poorer economies. The underlying mechanism is that, given the steeper firm-size gradient of training returns in the United States, the reallocation of employment toward larger and more productive firms induced by lower labor market turnover strengthens the response of training. By contrast, the interaction effects between labor market turnover and the productivity and firm size distribution are relatively modest, as lower labor market turnover alone already shift employment toward larger firms.

Table 6.2: Interaction effects of different channels on training across countries

	Impact of U.S. labor market turnover on training time under alternative scenarios			
Income (U.S. = 1)	Baseline	U.S. training returns	U.S. productivity dist.	Both U.S. returns & dist.
0.05	163.74%	254.71%	169.37%	240.08%
0.10	145.29%	218.31%	144.39%	207.06%
0.20	104.41%	139.41%	102.18%	135.90%
0.40	53.14%	67.98%	52.13%	67.23%
0.60	24.38%	24.91%	22.64%	24.01%
0.80	14.49%	15.51%	14.43%	15.47%

Notes: This table reports the percentage change in training time relative to the baseline resulting from setting the parameters governing labor market turnover to U.S. values in each country, holding all other parameters fixed, as well as under alternative scenarios in which one or both of the other channels are also set to U.S. values.

These interaction effects underscore the potential gains from coordinated policy interventions. In poorer countries, promoting sustained job attachment together with economies of scale in training at larger firms amplifies the benefits of lower labor market turnover by allowing firms to better exploit the increased employment concentration induced by reduced turnover.

6.3 Human capital and income gains due to training

Finally, we assess how each channel, through its impact on training, contributes to cross-country differences in human capital and income. In our model, the parameters governing each channel affect income both directly (through productivity, firm size, or labor market dynamics) and indirectly, by altering firms' training decisions. To isolate the component operating through training, we compare two counterfactual scenarios for each channel. In the first, the parameters associated with the channel are adjusted and training is allowed to respond endogenously. In the second, the same parameters are adjusted, but training levels at each firm are held fixed at their baseline values. The difference between these

two scenarios isolates the contribution of training. Table 6.3 reports the percent changes in income from this training channel relative to the baseline model.

Table 6.3: Impact of training on human capital and income across countries

Panel A: Human capital gains due to training

Income (U.S. = 1)	U.S. training returns	U.S. productivity dist.	U.S. labor market turnover
0.05	-1.84%	5.67%	19.29%
0.1	-6.13%	6.10%	21.80%
0.2	-7.01%	4.74%	16.47%
0.4	-6.88%	3.28%	10.20%
0.6	-3.99%	0.17%	4.85%
0.8	-2.63%	0.19%	2.13%

Panel B: Income gains due to training

Income (U.S. = 1)	U.S. training returns	U.S. productivity dist.	U.S. labor market turnover
0.05	-1.98%	5.91%	17.23%
0.1	-5.95%	6.24%	19.69%
0.2	-6.57%	4.75%	14.75%
0.4	-6.20%	3.30%	9.12%
0.6	-3.66%	0.46%	4.21%
0.8	-2.36%	0.24%	1.83%

Notes: This table reports the percentage change in human capital and income induced by training resulting from setting each corresponding group of parameters to its U.S. level, holding all other parameters fixed.

Similar to our previous results, we find that reducing labor market turnover to U.S. levels has the largest impact on human capital and income, particularly in low- and middle-income economies, while matching the productivity and firm size distribution yields smaller but still positive effects. In the poorest economy, the training gains induced by reducing labor market turnover and by aligning the productivity and firm size distributions with U.S. levels raise human capital by 19.29% and 5.67%, respectively, generating corresponding income gains of 17.23% and 5.91%. These findings suggest that reforms in developing countries aimed at improving labor market efficiency and addressing the "missing middle firms" problem can generate substantial additional income gains through increased training, providing a further rationale for their effectiveness.

Finally, setting training returns to U.S. levels reduces human capital and income in all economies. As shown in Table 6.1, for all but the poorest economy this decline occurs because training itself falls. For the poorest economy, however, human capital and income decline despite an increase in average training time. This outcome reflects a reallocation of training

toward larger firms and away from smaller firms, where training yields higher marginal gains in human capital. Given high labor market turnover, the resulting underinvestment in human capital among small firms quickly propagates through the economy, offsetting the gains from increased training in larger firms.

7 Conclusions

This paper studies how firms shape on-the-job training along the development path. Using harmonized cross-country data and administrative firm-level evidence, we document that training opportunities are systematically lower in smaller firms across countries, and that differences in productivity, labor shares, and job turnover are key determinants of this pattern. These findings highlight the importance of firm characteristics in shaping investments in on-the-job human capital and suggest that firms constitute a central, yet underexplored, channel behind cross-country training gaps.

To interpret these patterns and assess their aggregate implications, we develop a general equilibrium model in which firm heterogeneity interacts with labor market frictions to determine firms' incentives to invest in training. The model highlights several mechanisms through which firm characteristics affect training decisions, including complementarities between productivity and human capital, economies of scale in training at larger firms, differences in the share of training returns firms are able to appropriate, and differential exposure to labor market turnover and the hold-up problem. Quantitatively, we show that differences in labor market turnover are the dominant driver of cross-country training gaps and, through their effect on training, translate into sizable differences in human capital accumulation and income across countries.

Our findings have important implications for understanding the determinants of on-the-job learning and career advancement. First, they underscore the central role firms play in shaping workers' human capital and productivity trajectories, highlighting the need to integrate firm behavior more explicitly into both data collection efforts and models of human capital formation. Second, our results provide direct evidence of the importance of the hold-up problem in training decisions, suggesting that policies aimed at expanding access to training and improving worker productivity in the developing world should focus on firms and on reducing the constraints that limit their ability and willingness to invest in training. Future research could build on this framework by examining how the effectiveness of training subsidies varies across firms with different characteristics, such as management practices,

technology adoption, industries, and locations.

References

Abowd, J. M., K. L. McKinney, and N. L. Zhao (2018): "Earnings Inequality and Mobility Trends in the United States: Nationally Representative Estimates from Longitudinally Linked Employer-Employee Data," *Journal of Labor Economics*, 36. 2

Acemoglu, D. (1997): "Training and Innovation in an Imperfect Labour Market," *The Review of Economic Studies*, 64, 445, iSBN: 00346527. 2, 4, 60

Acemoglu, D. and W. Hawkins (2014): "Search with Multi-worker Firms," *Theoretical Economics*, 9, 583–628. 18, 24

Acemoglu, D. and J.-S. Pischke (1998): "Why Do Firms Train?" *The Quarterly Journal of Economics*, 113, 79–119. 2, 17, 60

——— (1999): "Beyond Becker: Training in Imperfect Labour Markets," *The Economic Journal*, 109, 112–142. 2, 5

Akcigit, U., H. Alp, and M. Peters (2016): "Lack of Selection and Limits to Delegation: Firm Dynamics in Developing Countries," Tech. rep., publication Title: NBER Working Paper Series Volume: 21905. 6

Altonji, J. G. and R. A. Shakotko (1987): "Do Wages Rise with Job Seniority?" *Review of Economic Studies*, 54, 437–459. 19

Arellano-Bover, J. (2020): "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size," . 5

——— (2024): "Career Consequences of Firm Heterogeneity for Young Workers: First Job and Firm Size," *Journal of Labor Economics*, 42, 549–589. 19

Arellano-Bover, J. and F. Saltiel (2021): "Differences in On-the-Job Learning across Firms," *Working Paper*. 19

——— (2023): "Differences in On-the-Job Learning Across Firms," *Working Paper*. 5

Autor, D., D. Dorn, L. F. Katz, C. Patterson, and J. Van Reenen (2020): "The Fall of the Labor Share and the Rise of Superstar Firms," *The Quarterly Journal of Economics*, 135, 645–709. 14

Axtell, R. L. (2001): "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818–1820. 23, 25

Bagger, J., F. Fontaine, F. Postel-Vinay, and J. M. Robin (2014): "Tenure, Experience, Human Capital, and Wages: A Tractable Equilibrium Search Model of Wage Dynamics," *American Economic Review*, 104, 1551–1596. 2, 17, 64

Barkai, S. (2020): "Declining Labor and Capital Shares," *The Journal of Finance*, 75, 2421–2463. 14

Barron, J. M., D. A. Black, and M. A. Loewenstein (1987): "Employer Size: The Implications for Search, Training, Capital Investment, Starting Wages, and Wage Growth," *Journal of Labor Economics*, 5, 76–89. 5

Barth, E., A. Bryson, J. C. Davis, and R. Freeman (2016): "It's Where You Work:

Increases in the Dispersion of Earnings across Establishments and Individuals in the United States," *Journal of Labor Economics*, 34. 2

Bassanini, A., A. Booth, G. Brunello, M. De Paola, and E. Leuven (2005): "Workplace Training in Europe," *IZA Discussion Paper Series*. 1, 5, 25

Bassanini, A. and R. Duval (2009): "Unemployment, Institutions, and Reform Complementarities: Re-Assessing the Aggregate Evidence for OECD Countries," *Oxford Review of Economic Policy*, 25, 40–59. 15

Becker, G. S. (1964): *Human Capital*, Chicago: University of Chicago Press. 2, 4

Ben-Porath, Y. (1967): "The Production of Human Capital Over the Life Cycle," *Journal of Political Economy*, 75, 352–365. 60

Bento, P. and D. Restuccia (2017): "Misallocation , Establishment Size , and Productivity Misallocation, Establishment Size , and Productivity," *American Economic Journal: Macroeconomics*, 9, 267–303. 5

Bhattacharya, D., N. Guner, and G. Ventura (2012): "Distortions , Endogenous Managerial Skills and productivity differences," *Review of Economic Dynamics*, 16, 11–25. 6

Black, D. A., B. J. Noel, and Z. Wang (1999): "On-the-Job Training, Establishment Size , and Firm Size: Evidence for Economies of Scale in the Production of Human Capital," *Southern Economic Journal*, 66, 82–100. 1, 5

Blundell, R. W., M. Costa Dias, D. Goll, and C. Meghir (2021): "Wages, Experience and Training of Women," *Journal of Labor Economics*, 39, 275–315. 23, 24

Braga, B. (2018): "Earnings Dynamics: The Role of Education throughout a Worker's Career," *Labour Economics*, 52, 83–97, publisher: Elsevier B.V. 1, 5

Buera, F. J., J. P. Kaboski, and Y. Shin (2011): "Finance and Development : A Tale of Two Sectors," *American Economic Review*, 101, 1964–2002. 5

Burdett, K. and D. T. Mortensen (1998): "Wage Differentials, Employer Size, and Unemployment," *International Economic Review*, 39, 257–273. 2, 17, 21, 67

Card, D., J. Kluve, and A. Weber (2018): "What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations," *Journal of the European Economic Association*, 16, 894–931. 5

Conti, G. (2005): "Training, Productivity and Wages in Italy," *Labour Economics*, 12, 557–576. 5

Da-Rocha, J.-M., M. M. Tavares, and D. Restuccia (2017): "Policy distortions and aggregate productivity with endogenous establishment-level productivity," Tech. rep., publication Title: NBER Working Paper Series Volume: 23339. 6

David, J. M., H. A. Hopenhayn, and V. Venkateshwaran (2014): "Information, Misallocation and Aggregate Productivity," 943–1005. 6

Dearden, L., H. Reed, and J. Van Reenen (2006): "The Impact of Training on Productivity and Wages: Evidence from British Panel Data," *Oxford Bulletin of Economics and Statistics*, 68, 397–421. 5

Donovan, K., W. J. Lu, and T. Schoellman (2023): "Labor Market Dynamics and Development," *The Quarterly Journal of Economics*, 138, 2287–2325. 8, 15, 25, 27, 28,

50, 51

Engbom, N. (2021): "Labor Market Fluidity and Human Capital Accumulation," *Manuscript*. 5

Engbom, N., C. Moser, and J. Sauermann (2022): "Firm Pay Dynamics," *NBER Working Paper Series*. 5

Foster, L., J. Haltiwanger, and C. Syverson (2008): "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability," *American Economic Review*, 98, 394–425. 12

Frazis, H., M. Gittleman, and M. Joyce (2000): "Correlates of Training: An Analysis Using Both Employer and Employee Characteristics," *ILR Review*, 53, 443–462. 16

Frazis, H. and M. A. Loewenstein (2005): "Reexamining the Returns to Training: Functional Form, Magnitude, and Interpretation," *The Journal of Human Resources*, 40, 453–476. 25

Frazis, H. J., D. E. Herz, and M. W. Horrigan (1995): "Employer-provided Training: Results from a New Survey," *Monthly Lab. Rev.*, 118, 3. 5

Friedrich, B., L. Laun, C. Meghir, and L. Pistaferri (2021): "Earnings Dynamics and Firm-Level Shocks," *NBER Working Paper Series*. 1, 5

Gollin, D. (2002): "Getting Income Shares Right," *Journal of Political Economy*, 110, 458–474, iSBN: 00223808. 11, 27, 45, 48

Gouin-Bonenfant, E. (2022): "Productivity Dispersion, Between-Firm Competition, and the Labor Share," *Econometrica*, Forthcoming. 14

Gregory, V. (2019): "Firms as Learning Environments: Implications for Earnings Dynamics and Job Search," *Manuscript*, 1–57. 1, 5, 19

Grullon, G., Y. Larkin, and R. Michaely (2019): "Are US Industries Becoming More Concentrated?" *Review of Finance*, 23, 697–743. 14

Guner, N., G. Ventura, and Y. Xu (2008): "Macroeconomic implications of size-dependent policies," *Review of Economic Dynamics*, 11, 721–744. 28

Haltiwanger, J., S. Scarpetta, and H. Schweiger (2014): "Cross country differences in job reallocation: The role of industry, firm size and regulations," *Labour Economics*, 26, 11–25. 15, 16

Harris, R. I. (1999): "The Determinants of Work-related Training in Britain in 1995 and the Implications of Employer Size," *Applied Economics*, 31, 451–463. 5

Heckman, J. J., R. J. Lalonde, and J. A. Smith (1999): "The Economics and Econometrics of Active Labor Market Programs," *Handbook of Labor Economics*, 3, 1865–2097. 5

Hornstein, A., P. Krusell, and G. L. Violante (2011): "Frictional Wage Dispersion in Search Models: A Quantitative Assessment," *American Economic Review*, 101, 2873–98. 21, 24

Hsieh, C. and P. Klenow (2014): "The Life Cycle of Plants in India and Mexico," *The Quarterly Journal of Economics*, 129, 1035–1084. 14, 28

Hsieh, C.-T. and P. J. Klenow (2009): "Misallocation and Manufacturing TFP in China and India," *Quarterly Journal of Economics*, 124. 5, 11, 12, 13, 14, 45, 47, 48, 57

Imai, S. and M. Keane (2004): "Intertemporal Labor Supply and Human Capital Accumulation," *International Economic Review*, 45, 601–641. 25

Jarosch, G. (2022): "Searching for Job Security and the Consequences of Job Loss," *NBER Working Paper Series*. 1, 5

Kambourov, G. and L. Manovskii (2009): "Occupational Specificity of Human Capital," *International Economic Review*, 50, 63–115. 19

Karabarbounis, L. and B. Neiman (2014): "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 129, 61–103. 14

Kaufmann, D. and A. Kraay (2024): *The Worldwide Governance Indicators: Methodology and 2024 Update*, Washington, DC: World Bank. 15, 50

Kluve, J. (2010): "The Effectiveness of European Active Labor Market Programs," *Labour Economics*, 17, 904–918, publisher: Elsevier B.V. 5

Konings, J. and S. Vanormelingen (2015): "The Impact of Training on Productivity and Wages: Firm-Level Evidence," *The Review of Economics and Statistics*, 97, 485–497. 5

Lagakos, D., B. Moll, T. Porzio, N. Qian, and T. Schoellman (2018): "Life Cycle Wage Growth across Countries," *Journal of Political Economy*, 126. 25

Lazear, E. P. (2009): "Firm-Specific Human Capital: A Skill-Weights Approach," *Journal of Political Economy*, 117, 699–726. 19

Leuven, E. (2004): "A Review of the Wage Returns to Private Sector Training," *Working Paper*. 25

Leuven, E. and H. Oosterbeek (2004): "Evaluating the Effect of Tax Deductions on Training," *Journal of Labor Economics*, 22, 461–488. 1, 5

Levinsohn, J. and A. Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 70, 317–341. 11, 12, 13, 14, 45, 47, 48, 58

Ma, X., A. Nakab, and D. Vidart (2024): "Human Capital Investment and Development: The Role of On-the-job Training," *Journal of Political Economy Macroeconomics*, 2. 1, 5, 8, 30

Manuelli, R. E. and A. Seshadri (2014): "Human Capital and the Wealth of Nations," *American Economic Review*, 104, 2736–2762. 25, 60

McKenzie, D. (2017): "How Effective are Active Labor Market Policies in Developing Countries? A Critical Review of Recent Evidence," *World Bank Research Observer*, 32, 127–154. 5

Midrigan, V. and D. Y. Xu (2014): "Finance and misallocation: Evidence from plant-level data," *American Economic Review*, 104, 422–458, arXiv: 1011.1669v3 ISBN: 9788578110796. 5

Mincer, J. (1989): "Human Capital and the Labor Market: A Review of Current Research," *Educational Researcher*, 18. 23

Moen, E. and A. Rosen (2004): "Does Poaching Distort Training?" *The Review of Economic Studies*, 71, 1143–1162. 2, 4, 60

Oi, W. Y. and T. L. Idson (1999): "Firm Size and Wages," *Handbook of Labor Economics*,

3rd Part, 2165–2214. 14

Olley, G. S. and A. Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263–1297. 11, 12, 13, 14, 45, 47, 48, 58

Pischke, J.-S. (2001): "Continuous Training in Germany," *Journal of Population Economics*, 14, 523–548. 2

Porzio, T. (2017): "Cross-Country Differences in the Optimal Allocation of Talent and Technology," *Job Market Paper*. 6

Poschke, M. (2018): "The Firm Size Distribution across Countries and Skill-Biased Change in Entrepreneurial Technology," *American Economic Journal: Macroeconomics*, 10, 1–41. 6

Restuccia, D. (2025): "On Productivity and Distortions across Countries," Working Paper 34573, National Bureau of Economic Research. 28

Restuccia, D. and R. Rogerson (2008): "Policy distortions and aggregate productivity with heterogeneous establishments," *Review of Economic Dynamics*, 11, 707–720, iSBN: 1094-2025. 5

——— (2013): "Misallocation and productivity," *Review of Economic Dynamics*, 16, 1–10, publisher: Elsevier Inc. ISBN: 1094-2025. 5

Roys, N. and A. Seshadri (2014): "Economic development and the structure of production," Tech. rep., publication Title: Working Paper ISBN: 09535314 (ISSN) ISSN: 14695758. 6

Shimer, R. (2005): "The Cyclical Behavior of Equilibrium Unemployment and Vacancies," *American Economic Review*, 95, 25–49. 24

——— (2012): "Reassessing the Ins and Outs of Unemployment," *Review of Economic Dynamics*, 15. 25

Song, J., D. Price, F. Guvenen, N. Bloom, and T. V. Wachter (2019): "Firming Up Inequality," *Quarterly Journal of Economics*. 2

Syverson, C. (2011): "What Determines Productivity?" *Journal of Economic Literature*, 49, 326–365. 14

van den Berg, G. J. (1990): "Search Behaviour, Transitions to Non-participation and the Duration of Unemployment," *The Economic Journal*, 100, 842–865. 21

What Works - Centre for Local Economic Growth (2016): "Evidence Review 1: Employment Training," Tech. rep., London School of Economics. 5

Online Appendix

A Data sources

A.1 World Bank Enterprise Survey (WB-ES)

The World Bank Enterprise Survey (WB-ES) is a collection of firm-level surveys covering private-sector establishments in manufacturing and services across more than 140 low- and middle-income countries.[27] The survey is administered by private contractors through face-to-face interviews on behalf of the World Bank. The WB-ES typically conducts between 1,200 and 1,800 interviews in large economies, around 360 in medium-sized economies, and approximately 150 in small economies. The survey employs a stratified random sampling design in which firms are grouped by size, sector, and geographic region, with random sampling within each stratum to ensure representativeness. Firm size categories are defined as 5–19 employees (small), 20–99 employees (medium), and 100 or more employees (large). Geographic regions are selected to collectively capture the majority of economic activity within each country. Because small and medium-sized firms constitute the majority of establishments in most economies, the WB-ES oversamples large firms. To ensure representativeness, all results are weighted using the observations' weights provided in the survey.

The WB-ES defines formal training as follows: "Formal training has a structured and defined curriculum. It may include classroom work, seminars, lectures, workshops, and audio-visual presentations and demonstrations. It does not include training to familiarize workers with equipment and machinery on the shop floor, training aimed at familiarizing workers with the establishment's standard operating procedures, or employee orientation at the beginning of a worker's tenure. In-house training may be conducted by other non-supervisory workers of the establishment, the establishment's supervisors or managers, or the establishment's training centers." Using this definition, we construct the following variables:

- Establishments offering formal training: An indicator equal to one if the enterprise had formal training programs for its permanent full-time employees during the last completed fiscal year.
- Labor share: The ratio of total employee compensation to total sales of the enterprise.
- Firm size: The total number of employees in the enterprise.

[27] Industries covered include ISIC Rev. 3.1 codes 15–37, 45, 50–52, 55, 60–64, and 72.

Table A.1: Countries in the WB-ES

Country	Year(s)	Country	Year(s)
Afghanistan	2008, 2014	Dem.Rep.Congo	2006, 2010, 2013
Albania	2002, 2005, 2007, 2013	Ecuador	2003, 2006, 2010, 2017
Algeria	2002	Egypt	2004, 2013, 2016
Angola	2006, 2010	El Salvador	2003, 2006, 2010, 2016
A.and Barbuda	2010	Eritrea	2002, 2009
Argentina	2006, 2010, 2017	Estonia	2002, 2005, 2009, 2013
Armenia	2002, 2005, 2009, 2013	Eswatini	2006, 2016
Azerbaijan	2002, 2005, 2009, 2013	Ethiopia	2002, 2011, 2015
Bahamas	2010	Fiji	2009
Bangladesh	2002, 2007, 2013	FYR Macedonia	2002, 2005, 2009, 2013
Barbados	2010	Gabon	2009
Belarus	2002, 2005, 2008, 2013	Gambia	2006, 2018
Belize	2010	Georgia	2002, 2005, 2008, 2013
Benin	2004, 2009, 2016	Germany	2005
Bhutan	2009, 2015	Ghana	2007, 2013
Bolivia	2006, 2010, 2017	Greece	2005
Bos. and Her.	2002, 2005, 2009, 2013	Grenada	2010
Botswana	2006, 2010	Guatemala	2003, 2006, 2010, 2017
Brazil	2003, 2009	Guinea	2006, 2016
Bulgaria	2002,'04,'05,'07,'09,'13	Guinea-Bissau	2006
Burkina Faso	2006, 2009	Guyana	2004, 2010
Burundi	2006, 2014	Honduras	2003, 2006, 2010, 2016
Cambodia	2003, 2013, 2016	Hungary	2002, 2005, 2009, 2013
Cameroon	2006, 2009, 2016	India	2002, 2006, 2014
Cape Verde	2006, 2009	Indonesia	2003, 2009, 2015
Cen. Af. Rep.	2011	Iraq	2011
Chad	2009, 2018	Ireland	2005
Chile	2004, 2006, 2010	Israel	2013
China	2002, 2003, 2012	Ivory Coast	2009, 2106
Colombia	2006, 2010, 2017	Jamaica	2005, 2010
Congo	2009	Jordan	2006, 2013
Costa Rica	2005, 2010	Kazakhstan	2002, 2005, 2009, 2013
Croatia	2002, 2005, 2007, 2013	Kenya	2003, 2007, 2013
Czech Republic	2002, 2005, 2009, 2013	Kosovo	2009, 2013
Djibuti	2013	Kyrgystan	2002, 2003,'05,'09,'13
Dominica	2010	Laos	2006, 2009, 2009, 2012
Dom. Republic	2005, 2010, 2016	Latvia	2002, 2005, 2009, 2013

Table A.1 - continued

Country	Year(s)	Country	Year(s)
Lebanon	2006, 2013	Serbia	2003, 2009, 2013
Lesotho	2003, 2009, 2016	Ser. and Mon.	2002, 2005
Liberia	2009, 2017	Sierra Leone	2009, 2017
Lithuania	2002,'04,05,'09,'13	Slovakia	2002, 2005, 2009, 2013
Madagascar	2005, 2009, 2013	Slovenia	2002, 2005, 2009, 2013
Malawi	2005, 2009, 2014	Solomon Islands	2015
Malaysia	2002, 2015	South Africa	2003, 2007
Mali	2003, 2007, 2010, 2016	South Korea	2005
Mauritania	2006, 2014	South Sudan	2014
Mauritius	2005, 2009	Spain	2005
Mexico	2006, 2010	Sri Lanka	2004, 2011
Micronesia	2009	St. K. and Nevis	2010
Moldova	2002, 2003,'05,'09,'13	Sudan	2014
Mongolia	2004, 2009, 2013	Suriname	2010
Montenegro	2003, 2009, 2013	Swaziland	2006
Morocco	2004, 2013	Sweden	2014
Mozambique	2007	Syria	2003
Myanmar	2014, 2016	Tajikistan	2002, 2003, 05, 08, 13
Namibia	2006, 2014	Tanzania	2003, 2006, 2013
Nepal	2009, 2013	Thailand	2004, 2016
Nicaragua	2003, 2006, 2010, 2016	Timor-Leste	2009, 2015
Niger	2005, 2009, 2017	Togo	2009, 2016
Nigeria	2007, 2014	Tonga	2009
Oman	2003	Tri. and Tob.	2010
Pakistan	2002, 2007, 2013	Tunisia	2013
Panama	2006, 2010	Turkey	2002, 2005, 2008, 2013
P. New Guinea	2015	Uganda	2003, 2006, 2013
Paraguay	2006, 2010, 2017	Ukraine	2002, 2005, 2008, 2013
Peru	2002, 2006, 2010, 2017	Uruguay	2006, 2010, 2017
Philippines	2003, 2009, 2015	Uzbekistan	2002, 2003, 05, 08, 13
Poland	2002,03,05,09,13	Vanuatu	2009
Portugal	2005	Venezuela	2006, 2010
Romania	2002, 2005, 2009, 2013	Vietnam	2005,
Russia	2002, 2005, 2009, 2012	W.B. and Gaza	2006, 2013
Rwanda	2006, 2011	Yemen	2010, 2013
Samoa	2009	Zambia	2007, 2013
Senegal	2003, 2007, 2014	Zimbabwe	2011, 2016

A.2 European Union Continuing Vocational Training Survey (EU-CVT)

The EU Continuing Vocational Training Survey (CVT) collects information on enterprises' investment in continuing vocational training for their staff. The information collected includes participation, time spent, and costs of CVT investments. Member states were asked to develop their own survey methods, such as written surveys, telephone interviews and direct personal interviews. In our analysis, we use data from 3 of the 5 waves of the EU-CVT: CVTS3 (2005), CVTS4 (2010) and CVTS5 (2015), which cover EU member states and Norway. Continuing vocational training refers to educational activities which are planned in advance, directly or indirectly financed at least partially by the enterprise, and geared towards the acquisition of new competences or the development and improvement of existing ones. Unstructured learning and initial vocational training (IVT) are excluded from CVT.

CVT measures and activities cover both CVT courses and other forms of CVT. CVT courses are clearly separated from the active workplace (instruction takes place in locations assigned for learning such as classrooms or training centers); show a high degree of organization by a trainer or training institution; and are designed for a group of learners (e.g. a curriculum exists). Two distinct types of CVT courses are identified: internal and external CVT courses. CVT courses are considered to be "formal training". Other forms of CVT are typically connected to the workplace, but they can also include participation (instruction) in conferences and trade fairs, among others. These are often characterized by self-organization by the individual learner or by a group of learners and are typically tailored to the workers' needs. The following types of other forms of CVT are identified: guided-on-the-job training; job rotation, exchanges, secondments or study visits; participation in conferences, workshops, trade fairs and lectures; participation in learning or quality circles; and self-directed learning/e-learning. These other forms of CVT are considered to be "informal training".

We capture enterprises that provide formal and informal training, along with time spent in formal courses, and firm size, in the following way:

- Enterprises offering training: Enterprises that provided CVT courses (formal training) or other forms of CVT (informal training) to their employees during the reference year.
- Firm size: Number of persons employed, which is defined as the total number of persons who work at the enterprise, excluding persons holding an apprenticeship or training

contract.

- Hours spent in formal training courses per participant: Average number of hours spent in CVT courses in the last year by workers who participated.

A.3 OECD Program for the International Assessment of Adult Competencies (PIAAC)

The Program for the International Assessment of Adult Competencies (PIAAC) is an international survey conducted by the Organization for Economic Cooperation and Development (OECD). The survey aims to assess and compare the learning environments, skills, and competencies of adults aged 16 to 65 in more than 40 OECD countries. 24 countries participated in round 1 of the survey, which collected data from 1 August 2011 to 31 March 2012. Round 2 of the assessment included 9 participating countries, with data collection taking place from April 2014 to the end of March 2015. Finally, round 3 included participation from 6 countries, with data collection taking place from July to December 2017.

PIAAC collects information about workers' learning investments in skills, along with information on how adults utilize these skills in various settings, namely home, work, and the wider community. In addition, PIAAC measures workers' proficiency in three key domains: literacy, numeracy, and problem-solving in technology-rich environments. In every country, PIAAC provides methodological documents and guidelines to facilitate the proper collection of data and ensure harmony in the definitions and concepts across countries.

We limit our sample to individuals who are currently employed and exclude military personnel. After these refinements and data construction for our main variables of interest, the countries included in our analysis from each round of the survey encompass:

- Round 1 (2011–2012): Austria, Belgium (Flanders), Canada, Czech Republic, Denmark, Estonia, Finland, France, Germany, Ireland, Italy, Japan, Korea, Netherlands, Norway, Poland, Russia, Slovakia, Spain, Sweden, United States, and the UK.
- Round 2 (2014–2015): Chile, Greece, Israel, Lithuania, New Zealand, Slovenia, and Turkey.
- Round 3 (2017): Ecuador, Hungary, Kazakhstan, Mexico, and Peru.

We capture workers who engage in formal and informal training in the following way:

- Workers engaging in formal training: Workers who reported participating in courses,

seminars, workshops, private lessons or other courses led by either coworkers or outsiders in the last 12 months.

- Workers engaging in informal training: Workers who reported learning skills from coworkers more than once a month in their current work.
- Hours spent in formal training: Number of hours spent in formal training in the last 12 months.
- Firm size: Number of people currently working for the same employer.

A.4 Chinese Annual Survey of Manufacturing

The Chinese Annual Survey of Manufacturing is an administrative dataset with detailed information on the demographics and balance sheets of manufacturing firms in China. On average, 200 to 400 thousand firms are surveyed in each sample year, representing more than 90% of China's manufacturing output. Data from the survey is available for the years 1998–2007 and 2011–2013. This survey contains detailed financial information from all manufacturing firms with revenues exceeding a certain threshold. In the early period (1998–2007), the database is skewed towards state-owned enterprises (SOEs) and large private firms with sales more than 5 million RMB. In the later period (2011-2013), only firms with sales higher than 20 million RMB are covered. The financial information contained in the survey includes sales, employment, payroll, capital stock, investment, cost of intermediate inputs, location, and industry classification in each sample year. In addition, the surveys for the years 2005–2007 contain information on expenditures in workers' formal training fees.

We limit our analysis to firms with at least one paid employee, and a positive value of sales, value added, payroll, and fixed capital. The variables we construct using this data are:

- Per-worker expenditures in training fees: Ratio between total training expenditures and employment in every firm.
- Labor share: Ratio between payroll and value added in each firm; winsorized at the 99th percentile.
- Firm size: Total number of employees in the firm.
- Credit access (leverage): Ratio between liabilities and capital in each firm
- TFP (HK1, labor share = $^2/_3$): Measure of TFPQ given by $(P_{i,s}Y_{i,s})^{\frac{\sigma}{\sigma-1}}/K_{i,s}^{\alpha}(w_{i,s}L_{i,s})^{1-\alpha}$,

where $P_{i,s}Y_{i,s}$, $K_{i,s}$ and $w_{i,s}L_{i,s}$ capture the value added, book value of fixed capital, and payroll, respectively, for each firm i operating in industry s (measured at the 4-digit level). The elasticity of substitution between the varieties produced in each industry, σ, which is used to separate price and quantity from the value added in the construction of TFPQ, is set to 3 following Hsieh and Klenow (2009), while the labor share $1-\alpha$ is given by $2/3$ following Gollin (2002). This TFP measure is winsorized at the 1st and 99th percentiles.

- TFP (HK2, industry-level labor share): Measure of TFPQ given by $(P_{i,s}Y_{i,s})^{\frac{\sigma}{\sigma-1}}/K_{i,s}^{\alpha}(w_{i,s}L_{i,s})^{1-\alpha}$, where $P_{i,s}Y_{i,s}$, $K_{i,s}$ and $w_{i,s}L_{i,s}$ capture the value added, book value of fixed capital, and payroll, respectively, for each firm i operating in industry s (measured at the 4-digit level). The elasticity of substitution between the varieties produced in each industry, σ, which is used to separate price and quantity from the value added in the construction of TFPQ, is set to 3 following Hsieh and Klenow (2009), while the labor share $1-\alpha$ is given by the average labor share in industry s following Hsieh and Klenow (2009). This TFP measure is winsorized at the 1st and 99th percentiles.

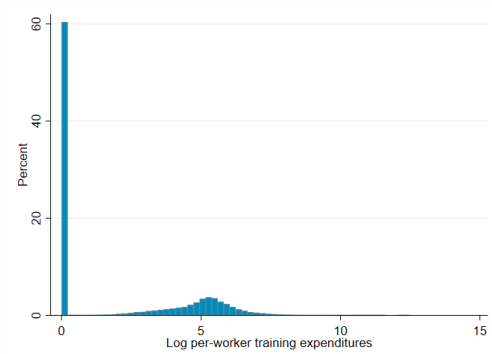
- TFP (HK3, firm-level labor share): Measure of TFPQ given by $(P_{i,s}Y_{i,s})^{\frac{\sigma}{\sigma-1}}/K_{i,s}^{\alpha}(w_{i,s}L_{i,s})^{1-\alpha}$, where $P_{i,s}Y_{i,s}$, $K_{i,s}$ and $w_{i,s}L_{i,s}$ capture the value added, book value of fixed capital, and payroll, respectively, for each firm i operating in industry s (measured at the 4-digit level). The elasticity of substitution between the varieties produced in each industry, σ, which is used to separate price and quantity from the value added in the construction of TFPQ, is set to 3 following Hsieh and Klenow (2009), while the labor share $1-\alpha$ is given by the labor share prevalent in each firm. This TFP measure is winsorized at the 1st and 99th percentiles.

- TFP (OP) and TFP (LP): Measures of TFPT estimated using the methodologies of Olley and Pakes (1996) and Levinsohn and Petrin (2003), respectively. We implement both approaches using value added as the outcome variable and estimate them separately by 2-digit industry. In both specifications, we include capital and firm age as state variables, and payroll and firm size as freely adjustable inputs. The two methods differ in the proxy used for unobserved productivity. In the OP specification, we use log investment (measured as net purchases of fixed capital) and restrict the sample to firms with positive investment. In the LP specification, we use log intermediate input expenditures as the proxy and restrict the sample to firms with positive intermediate inputs. All variables are deflated using the corresponding price index (base December

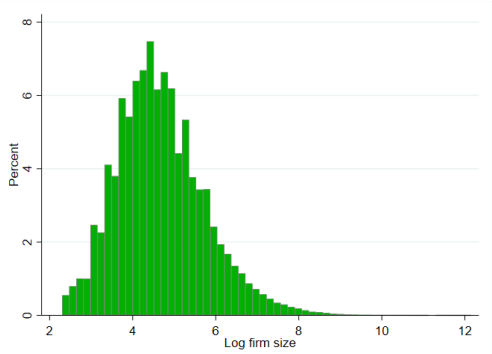
1998).

Figure A.1: Histograms of key variables in Chinese data

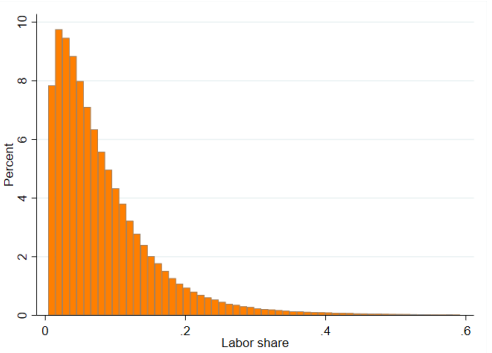
(a) Log per-worker training expenditures



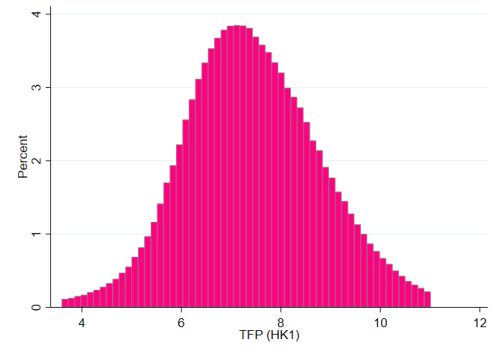
(b) Log firm size



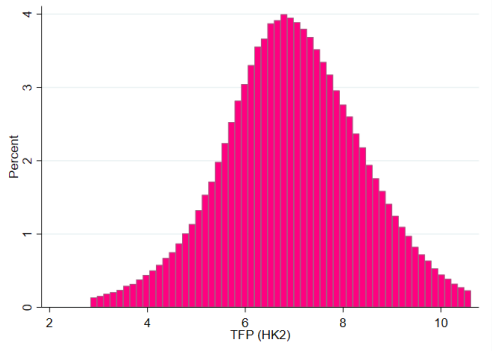
(c) Labor share



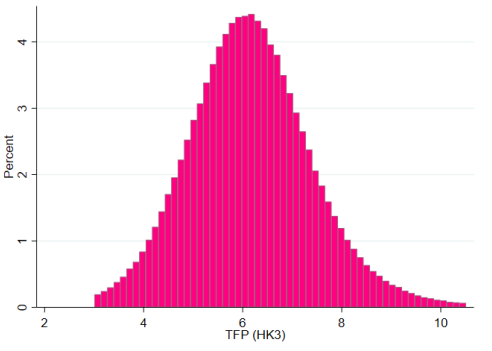
(d) TFP (HK1)



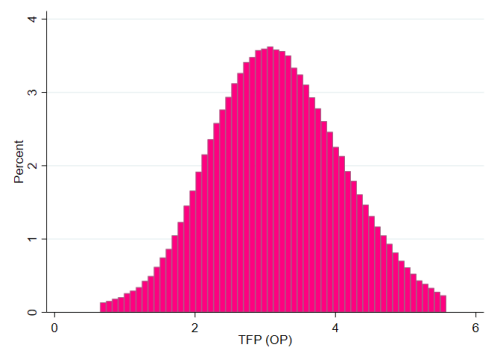
(e) TFP (HK2)



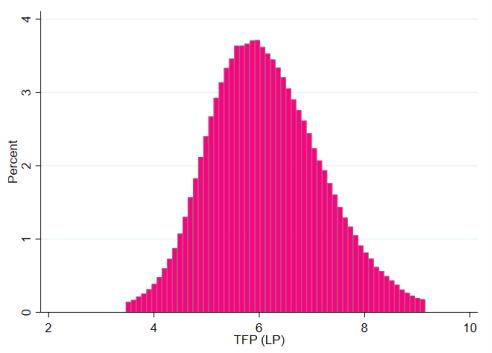
(f) TFP (HK3)



(g) TFP (OP)



(h) TFP (LP)



Notes: These figures plot histograms of the training, labor share, log firm size, and different TFP measures in the Chinese data. The construction of these variables is described in detail in Appendix A.4. The TFP and labor share measures plotted are winsorized at the 1st and 99th percentiles.

A.5 Mexican Economic Census

The Mexican Economic Census is an administrative dataset with detailed information on the demographics and balance sheets of all economic units outside of agriculture and forestry that operate in a permanent location delimited by buildings or other fixed installations.[28] The census is conducted every five years, and data is available for 1989–2019. In 2019 the census surveyed more than 6 million economic units.

The financial information contained in each census includes sales, employment, payroll, capital stock, investment, cost of intermediate inputs, location, and industry classification. For our main regressions, we rely on data from 2019, which contains information on the share of workers that received training (both formal and informal) in the past year. However, and given that the production function estimation methods of Olley and Pakes (1996) and Levinsohn and Petrin (2003) require panel data, we also include data from the economic censuses of 2009 and 2014, which contain information allowing us to link establishments across census waves.

We limit our analysis to firms with at least one paid employee, and a positive value of sales, value added, payroll, and fixed capital. The variables we construct using this data are:

- Share of employees trained: Ratio between the number of workers that received training in the past year and the total number of workers.
- Labor share: Ratio between payroll and sales in each firm; winsorized at the 99th percentile.
- Firm size: Total number of employees in the firm.
- Credit access (self-reported): Binary indicators capturing whether the firm reports difficulty accessing credit and whether it obtained credit during the previous year.
- TFP (HK1, labor share = $2/3$): Measure of TFPQ given by $(P_{i,s}Y_{i,s})^{\frac{\sigma}{\sigma-1}}/K_{i,s}^{\alpha}(w_{i,s}L_{i,s})^{1-\alpha}$, where $P_{i,s}Y_{i,s}$, $K_{i,s}$ and $w_{i,s}L_{i,s}$ capture the revenue, book value of fixed capital, and payroll, respectively, for each firm i operating in industry s (measured at the 4-digit level). The elasticity of substitution between the varieties produced in each industry, σ, which is used to separate price and quantity from revenue in the construction of TFPQ, is set to 3 following Hsieh and Klenow (2009), while the labor share $1-\alpha$ is

[28] A few economic units fitting this description are excluded in some waves due to multiple difficulties. For example, in 2019 political associations and homes with domestic workers, among others, are excluded.

given by $^2/_3$ following Gollin (2002). This TFP measure is winsorized at the 1st and 99th percentiles.

- TFP (HK2, industry-level labor share): Measure of TFPQ given by $(P_{i,s}Y_{i,s})^{\frac{\sigma}{\sigma-1}}/K_{i,s}^{\alpha}(w_{i,s}L_{i,s})^{1-\alpha}$, where $P_{i,s}Y_{i,s}$, $K_{i,s}$ and $w_{i,s}L_{i,s}$ capture the revenue, book value of fixed capital, and payroll, respectively, for each firm i operating in industry s (measured at the 4-digit level). The elasticity of substitution between the varieties produced in each industry, σ, which is used to separate price and quantity from revenue in the construction of TFPQ, is set to 3 following Hsieh and Klenow (2009), while the labor share $1-\alpha$ is given by the average labor share in industry s following Hsieh and Klenow (2009). This TFP measure is winsorized at the 1st and 99th percentiles.

- TFP (HK3, firm-level labor share): Measure of TFPQ given by $(P_{i,s}Y_{i,s})^{\frac{\sigma}{\sigma-1}}/K_{i,s}^{\alpha}(w_{i,s}L_{i,s})^{1-\alpha}$, where $P_{i,s}Y_{i,s}$, $K_{i,s}$ and $w_{i,s}L_{i,s}$ capture the revenue, book value of fixed capital, and payroll, respectively, for each firm i operating in industry s (measured at the 4-digit level). The elasticity of substitution between the varieties produced in each industry, σ, which is used to separate price and quantity from revenue in the construction of TFPQ, is set to 3 following Hsieh and Klenow (2009), while the labor share $1-\alpha$ is given by the labor share prevalent in each firm. This TFP measure is winsorized at the 1st and 99th percentiles.

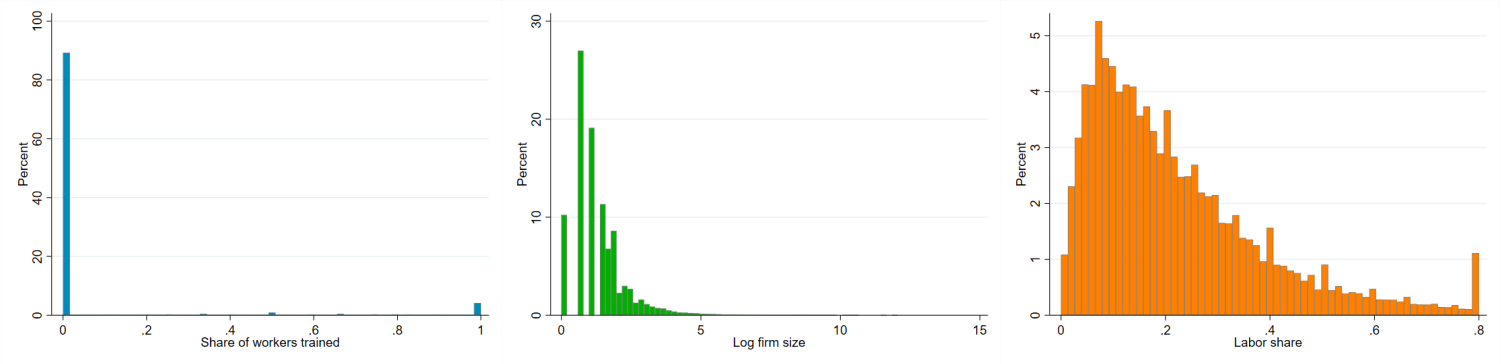
- TFP (OP) and TFP (LP): Measures of TFPT estimated using the methodologies of Olley and Pakes (1996) and Levinsohn and Petrin (2003), respectively. We implement both approaches using revenue as the outcome variable and estimate them separately by 2-digit industry. In both specifications, we include capital and firm age as state variables, and payroll and firm size as freely adjustable inputs. In the OP specification, we additionally include intermediate inputs as a freely adjustable input given the use of revenue as the outcome, while in the LP specification intermediate inputs serve as the proxy for unobserved productivity. The OP specification instead uses log investment (measured as net purchases of fixed capital) as the proxy and restricts the sample to firms with positive investment, while the LP specification restricts the sample to firms with positive intermediate inputs. All variables are deflated using the corresponding price index (base June 2012).

Figure A.2: Histograms of key variables in Mexican data

(a) Share of workers trained

(b) Log firm size

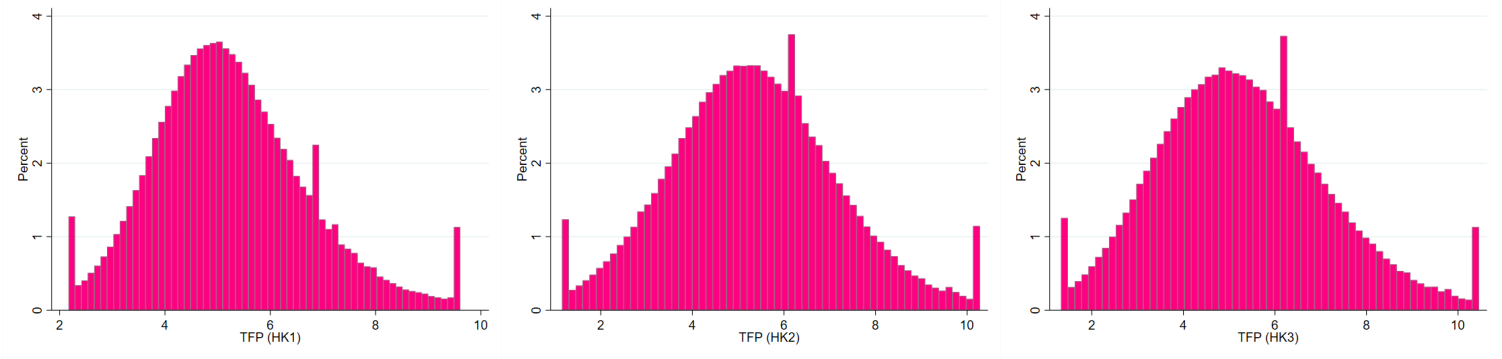
(c) Labor share



(d) TFP (HK1)

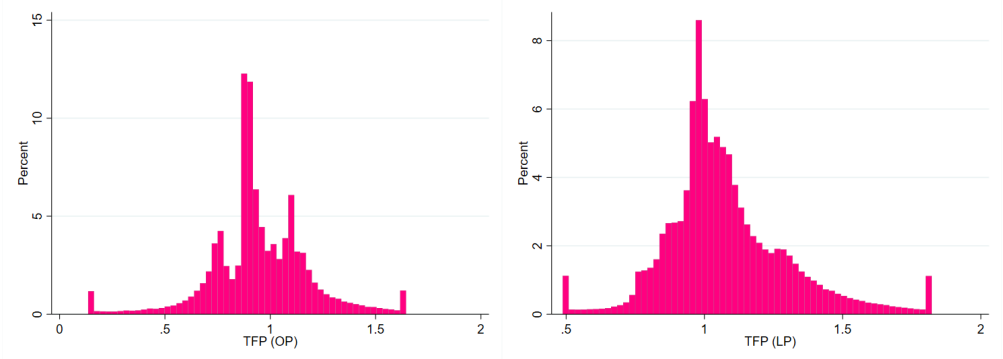
(e) TFP (HK2)

(f) TFP (HK3)



(g) TFP (OP)

(h) TFP (LP)



Notes: These figures plot histograms of the training, labor share, log firm size, and different TFP measures in the Mexican data. The construction of these variables is described in detail in Appendix A.5. The TFP and labor share measures plotted are winsorized at the 1st and 99th percentiles.

A.6 Job turnover measures from Donovan et al. (2023) and World Bank Worldwide Governance Indicators

We construct the predicted job turnover rates used in Section 3.5 by combining two data sources.

- **Job turnover rates from Donovan et al. (2023)**: Donovan et al. (2023) construct a cross-country dataset of labor market flows using harmonized microdata from rotating panel labor force surveys covering approximately 80 million individuals across 49 countries. The data allow workers to be tracked across consecutive quarters, enabling the measurement of quarterly employment transitions. We use their measures of the probability that a worker remains in the same job after one quarter across countries and years, which they kindly shared with us.
- **World Bank Worldwide Governance Indicators (WGI) (Kaufmann and Kraay, 2024)**: The WGI provide annual country-level measures of institutional quality for over 200 countries during 1996–2018 across multiple dimensions, including voice and accountability, political stability, government effectiveness, regulatory quality, rule of law, and control of corruption. The indicators aggregate information from more than 30 underlying sources, including household and firm surveys, commercial data providers, non-governmental organizations, and public-sector institutions, reflecting the views and experiences of citizens, firms, and experts worldwide.

We regress the observed turnover measures from Donovan et al. (2023) on five of the WGI institutional indicators: voice and accountability, political stability, regulatory quality, rule of law, and control of corruption.[29] We then use the fitted values from this regression to predict job turnover for the country–year observations in our EU-CVT sample. The results of this prediction regression are reported below.

[29] We do not include government effectiveness, as this variable was not statistically significant at the 10% level when all variables were included in the regression.

Table A.2: Relationship between job turnover and World Bank Worldwide Governance Indicators

Dep. variable:	Probability same job after quarter
	(1)
Voice and accountability	0.11***
	(0.018)
Political stability	0.040***
	(0.0055)
Regulatory quality	-0.016*
	(0.0095)
Rule of law	0.11***
	(0.014)
Control of corruption	-0.13***
	(0.0074)
Constant	0.80***
	(0.010)
Observations	381
R-squared	0.65

Notes: This table shows the results from regressing the probability of staying in the same job after a quarter in each country-year built by Donovan et al. (2023) on five of the WGI institutional indicators: voice and accountability, political stability, regulatory quality, rule of law, and control of corruption. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

A.7 U.S. Survey of Employer Provided Training Data

The Bureau of Labor Statistics (BLS) in the United States has conducted two surveys of employer provided training (US-SEPT). The first US-SEPT, conducted in 1994, focused on the existence and types of formal training programs provided or financed by establishments. The second US-SEPT, conducted in 1995, collected training information from both employers and randomly selected employees. Due to data availability, we focus on the 1995 wave of the survey, which studies establishments with 50 or more workers. The employer portion of the survey focuses on the intensity and costs of employer-provided formal training. The employee portion of the survey focuses on the time that employees spent on both formal and informal training. This survey provides a sample of 1,062 establishments and over 1,000 employees covering all nine major industry classifications across all 50 states.

The micro-level data for this survey are not available for researchers outside the BLS. Thus, we rely on aggregate statistics on the ratio of training in firms of different sizes, the share of training time relative to total working hours, and different types of training costs, for the

calibration of the quantitative model presented in Section 5.

A.8 European Union Adult Education Survey (EU-AES)

The EU-AES is a worker-level survey that collects information on participation in education and learning activities, including on-the-job training, with the specific goal of understanding adult education patterns. The AES is one of the main data sources for the EU lifelong learning statistics and covers approximately 666,000 adults aged 25–64. This data was collected as a mandatory survey in 2011, and 2017 in 27, and 28 EU member states, respectively.[30]

We leverage information from the EU-AES on both whether workers engaged in training during the previous year and how this training was financed—whether by employers, public institutions, workers themselves, or a combination of these sources—to further support our findings on the central role of firms in training decisions and financing.

B Robustness of empirical results

B.1 Robustness of results using cross-country data

B.1.1 Results using firm-level data (WB-ES and EU-CVT)

Table B.1: Correlation between firm size and training (WB-ES)

Dep. variable	Firm offers formal training				
	(1)	(2)	(3)	(4)	(5)
Log firm size	0.11***	0.11***	0.10***	0.10***	0.11***
	(0.0027)	(0.0028)	(0.0027)	(0.0027)	(0.0029)
Constant	0.022**	0.13	-0.056**	0.18	0.21
	(0.0092)	(0.12)	(0.022)	(0.12)	(0.40)
Year FE		Y		Y	Y
Country FE			Y	Y	Y
Industry FE					Y
Observations	93,297	93,297	93,297	93,297	87,573
R-squared	0.083	0.098	0.167	0.172	0.183

Notes: This table shows different specifications in which we regress a dummy variable indicating if the firm offers formal training to its workers on firm size. Data come from the WB-ES. Results are weighted using the observations' weights provided in the survey. Industry FE correspond to two-digit industries. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

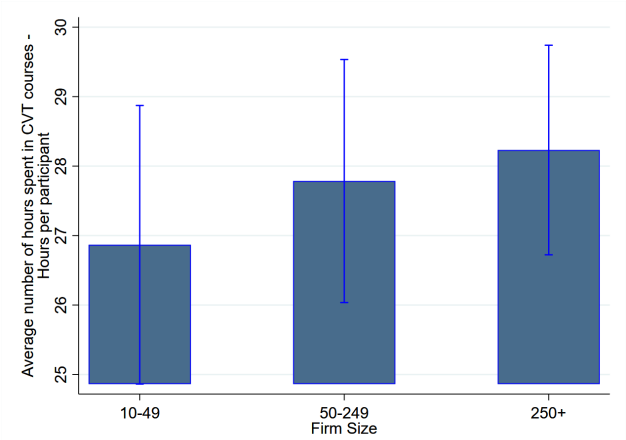
[30]The survey was also collected in 2007, but was voluntary.

Table B.2: Share of firms offering formal training by purpose and firm size (EU-CVT)

	Average by firm size in 2010			
	All	10-49	50-249	250+
General IT	27.3	23.7	34.5	54.7
Professional IT	16.9	14.5	21	37.5
Management	32	26.2	43.7	74.3
Team working	32.5	29	38.3	61.6
Customer handling	38.5	35.4	44.1	62.7
Problem solving	30.1	28.5	31.2	50
Office administration	26.9	24.3	32.3	45.1
Foreign language	15.3	11	24	46.9
Technical or job-specific	69	67.2	73.2	81.2
Oral or written communication	14.7	12.7	16.9	36.5
Numeracy and/or literacy	7	6.7	6.5	14.7
Other skills and competences	11	11.2	10.4	10.3

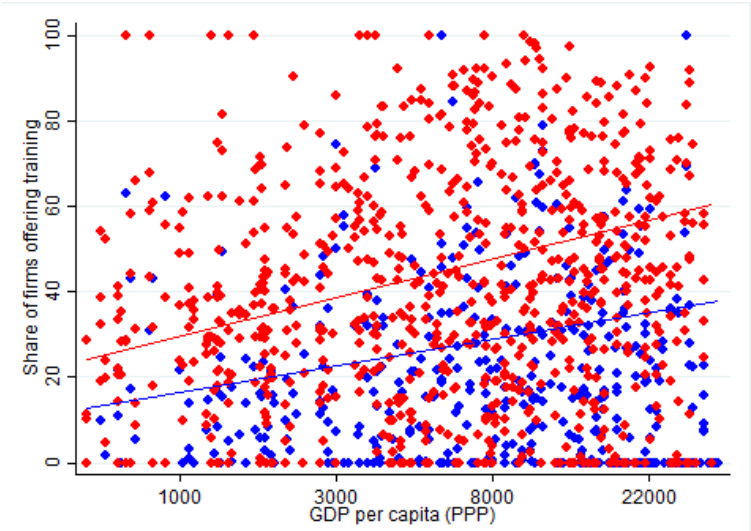
Notes: This table presents the average shares of firms reporting formal training for different purposes by firm size in the EU-CVT in 2010.

Figure B.1: Hours spent in formal training courses per participant by firm size (EU-CVT)



Notes: We plot the average number of hours spent in formal training (CVT) courses in the last year by each worker who participated, grouped by firm size categories. The bars represent half of the standard deviation of the mean hours spent in training for each group. Data comes from the EU-CVT. 95% confidence intervals are plotted.

Figure B.2: Share of firms offering formal training by relative firm size

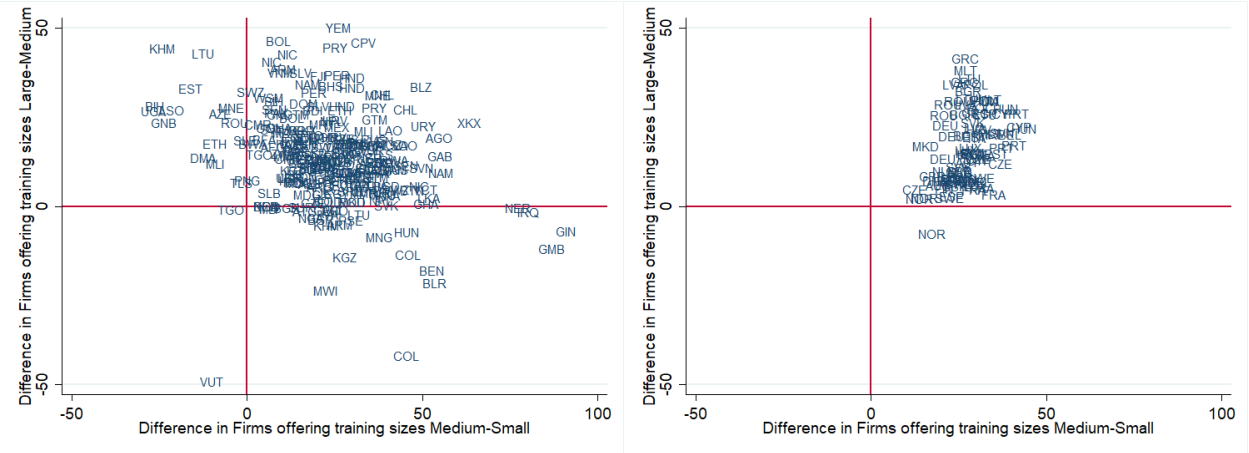


Notes: Each dot represents the share of firms in a specific firm size category offering formal training in each country. Red dots represent firms with more employees than the median number of employees in that year and country, while blue dots represent those with fewer. Data on training comes from the WB-ES. Data on GDP per capita comes from the Penn World Tables. Results are weighted using the observations' weights provided in the survey.

Figure B.3: Difference in share of firms offering formal training by firm size

(a) WB-ES

(b) EU-CVT



Notes: We plot the share of firms offering formal training between medium and small firms, against the difference in the share of firms offering training between large and medium firms. The firm sizes considered are: 2-20 (Small), 21-100 (Medium), and 100+ (large). Data comes from the WB-ES (Panel (a)), and the EU-CVT (Panel (b)).

Table B.3: Correlation between training and the labor share (WB-ES)

Dep. variable:	Firm offers formal training					
	(1)	(2)	(3)	(4)	(5)	(6)
Labor share	-0.14***	-0.080***	-0.076***	-0.11***	-0.10***	-0.10***
	(0.019)	(0.018)	(0.018)	(0.020)	(0.020)	(0.020)
Log firm size		0.11***	0.11***	0.11***	0.11***	0.11***
		(0.0027)	(0.0028)	(0.0029)	(0.0029)	(0.0029)
Constant	0.37***	0.044***	0.18	0.17	0.23	0.23
	(0.0057)	(0.010)	(0.12)	(0.37)	(0.40)	(0.40)
Year FE			Y		Y	Y
Country FE				Y	Y	Y
Industry FE						Y
Observations	92,012	92,012	92,012	92,012	92,012	87,295
R-squared	0.002	0.009	0.031	0.133	0.137	0.157

Notes: This table shows different specifications in which we regress a dummy variable indicating whether the firm provides formal training to at least some of its workers on firm size, and labor share using data from the WB-ES. The labor share measure is described in Appendix A.1. Results are weighted using the observations' weights provided in the survey. Industry FE correspond to two-digit industries. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

Table B.4: Correlation between firm size and the labor share (WB-ES)

Dep. variable	Labor share (in percentage points)				
	(1)	(2)	(3)	(4)	(5)
Log firm size	-1.10***	-1.13***	-1.00***	-0.98***	-1.34***
	(0.12)	(0.14)	(0.11)	(0.12)	(0.12)
Constant	25.6***	66.7***	28.2***	66.0***	36.1***
	(0.43)	(11.2)	(1.57)	(11.3)	(7.67)
Year FE		Y		Y	Y
Country FE			Y	Y	Y
Industry FE					Y
Observations	111,375	111,375	111,375	111,375	100,196
R-squared	0.004	0.009	0.049	0.055	0.105

Notes: This table shows different specifications in which we regress the labor share on firm size using data from the WB-ES. The labor share measure is described in Appendix A.1. Results are weighted using the observations' weights provided in the survey. Industry FE correspond to two-digit industries. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

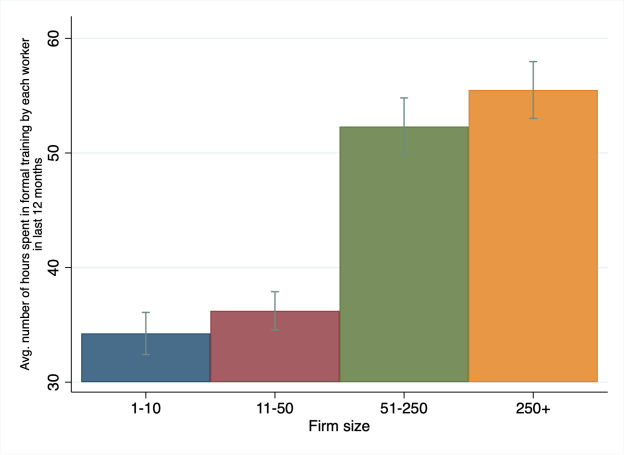
B.2 Results using worker-level data (PIAAC)

Table B.5: Correlation between formal and informal training and firm size

	(1)	(2)	(3)
Dep. variable	Formal training	Informal training	Hours of formal training
Firm of 11–50 workers	0.09***	0.07***	3.70
	(0.01)	(0.01)	(2.60)
Firm of 51–250 workers	0.15***	0.06***	16.79***
	(0.01)	(0.01)	(3.55)
Firm of 250+ workers	0.21***	0.09***	20.95***
	(0.01)	(0.01)	(4.26)
Constant	0.26**	0.72***	10.65
	(0.10)	(0.05)	(13.18)
Age FE	Y	Y	Y
Country FE	Y	Y	Y
Demographic controls	Y	Y	Y
Worker type FE	Y	Y	Y
Industry FE	Y	Y	Y
Occupation FE	Y	Y	Y
Wage controls	Y	Y	Y
Observations	55,502	56,342	56,393
R-squared	0.22	0.12	0.04

Notes: This table shows regressions of variables indicating participation in formal and informal training, along with the number of hours spent in formal training in the last 12 months. The omitted firm size category is firms with 1-10 employees. Demographic controls include educational attainment level and gender. Worker type categories include private employee, government employee, non-profit employee and self-employed. Industry and occupation categories are at the 2-digit level (ISIC rev. 4 and ISCO 2008, respectively). Wage controls include the current hourly wage. Results are weighted using the observations' weights provided in the survey. Robust standard errors are in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure B.4: Hours spent in formal training by firm size



Notes: We plot the average number of hours spent by each worker in the most recent formal training activity of the last 12 months in firms of different sizes. Data comes from PIAAC and encompasses 34 countries in the OECD. Results are weighted using the observations' weights provided in the survey. 95% confidence intervals are plotted.

B.3 Robustness of results using Chinese and Mexican data

Table B.6: Correlation between training (intensive margin), TFP (HK), and the labor share

Dep. variable:	China Log per-worker training expenditures						Mexico Share of workers trained		
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)
Labor share	-0.769***	-0.014	-1.066***	-0.053	-1.750***	-0.669***	0.020***	-0.046***	-0.045***
	(0.042)	(0.066)	(0.042)	(0.066)	(0.039)	(0.068)	(0.002)	(0.001)	(0.001)
TFP (HK1)	0.113***	0.083***					0.015***		
	(0.002)	(0.004)					(0.0002)		
TFP (HK2)			0.083***	0.077***				0.003***	
			(0.002)	(0.004)				(0.0001)	
TFP (HK3)					0.033***	0.073***			0.003***
					(0.003)	(0.004)			(0.0001)
Log firm size	0.262***	-0.034***	0.275***	-0.046***	0.297***	-0.063***	0.052***	0.059***	0.059***
	(0.003)	(0.011)	(0.003)	(0.011)	(0.003)	(0.011)	(0.0003)	(0.0002)	(0.0002)
Age FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y	Y	Y			
Industry FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Credit access control	Y	Y	Y	Y	Y	Y			
Firm FE		Y		Y		Y			
Observations	772,494	658,297	768,209	654,531	773,612	659,887	1,560,628	1,557,759	1,560,623
R-squared	0.061	0.708	0.061	0.709	0.060	0.709	0.209	0.207	0.206

Notes: This table shows different specifications in which we regress log per-worker training expenditures (China) and the share of employees trained (Mexico) on firm size, labor share, and different measures of TFP constructed using the methodology of Hsieh and Klenow (2009). The TFP and labor share measures are described in Appendix A.4 and Appendix A.5. As many firms report zero training expenditures, we apply a log1plus transformation to per-worker training expenditures. Industry FE corresponds to four-digit industries in both China and Mexico. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

Table B.7: Correlation between training (intensive margin), TFP (OP & LP), and the labor share

Dep. variable:	China Log per-worker training expenditures				Mexico Share of workers trained	
	(1)	(2)	(3)	(4)	(1)	(2)
Labor share	-1.670***	-0.087	-0.932***	0.039	-0.013***	-0.034***
	(0.044)	(0.070)	(0.043)	(0.067)	(0.005)	(0.002)
TFP (OP)	0.002	0.056***			0.024***	
	(0.004)	(0.006)			(0.006)	
TFP (LP)			0.122***	0.110***		0.031***
			(0.004)	(0.006)		(0.002)
Log firm size	0.307***	-0.052***	0.235***	-0.073***	0.079***	0.060***
	(0.003)	(0.011)	(0.004)	(0.011)	(0.0007)	(0.0003)
Age FE	Y	Y	Y	Y	Y	Y
Year FE	Y	Y	Y	Y		
Industry FE	Y	Y	Y	Y	Y	Y
Credit access control	Y	Y	Y	Y	Y	Y
Firm FE		Y		Y		
Observations	771,638	657,209	771,639	657,822	122,340	719,120
R-squared	0.060	0.709	0.058	0.707	0.251	0.149

Notes: This table shows different specifications in which we regress log per-worker training expenditures (China) and the share of employees trained (Mexico) on firm size, labor share, and different measures of TFP constructed using the methodologies of Olley and Pakes (1996) and Levinsohn and Petrin (2003). The TFP and labor share measures are described in Appendix A.4 and Appendix A.5. As many firms report zero training expenditures, we apply a log1plus transformation to per-worker training expenditures. Industry FE corresponds to four-digit industries in both China and Mexico. Robust standard errors in parentheses. *p <0.05, **p <0.01, ***p <0.001.

C Additional results from analytical model

C.1 Firms' and workers' benefits from training

In a firm with productivity z, a wage distribution of offers $F(w(u))$ from poaching firms indexed by u, and labor share $\beta(z)$, the marginal benefits of additional human capital in the next period for workers and firms are respectively given by

$$MR_W(z) = \underbrace{(1-\delta)\left[1-\eta\bar{F}(w(z))\right]}_{\text{prob of worker staying in firm}} \underbrace{\beta(z)Az}_{\text{worker's share of value}} + \underbrace{(1-\delta)\eta\int \mathbf{1}_{\{w(u)>w(z)\}}\beta(u)Auf(w(u))w'(u)du}_{\text{expected worker's share of value after job-to-job transitions}} + \underbrace{\delta\int \beta(u)Auf(w(u))w'(u)du}_{\text{expected worker's share of value after exogenous separations}} \tag{C.1}$$

and

$$MR_F(z) = \underbrace{(1-\delta)\left[1-\eta\bar{F}(w(z))\right]}_{\text{prob of worker staying in firm}} \underbrace{(1-\beta(z))\,Az}_{\text{firm's share of value}}. \tag{C.2}$$

where $\bar{F}(w(z)) = 1 - F(w(z))$ is the probability of obtaining an offer with higher wage than $w(z)$.

To obtain these, we consider workers' and firms' optimization problems. A young worker's present value of income in firm z is:

$$\underbrace{\beta(z)Az}_{\text{current wage}} - \underbrace{\mu_W\left(Az + c_s\bar{w}\right)}_{\text{worker's per-unit training costs}} \times \underbrace{s}_{\text{training level}} + \frac{1}{1+\rho}\Bigg\{\delta \underbrace{\int \beta(u)Auf(w(u))w'(u)du}_{\text{U back to a firm}} \times \underbrace{h'}_{\text{next-period human capital}}$$

$$+ (1-\delta)\Bigg[\underbrace{\left[1-\eta\bar{F}(w(z))\right]\beta(z)Az}_{\text{if stay in current firm}} + \underbrace{\eta\int \mathbf{1}_{\{w(u)>w(z)\}}\beta(u)Auf(w(u))w'(u)du}_{\text{if move to new firm}}\Bigg] \times \underbrace{h'}_{\text{next-period human capital}}\Bigg\},$$

where $h' = 1 + \zeta L(z)^{\alpha} s^{\gamma_s}$ is the next-period human capital. The expression $\mathbf{1}_{\{w(u)>w(z)\}}$ represents the indicator function that takes the value 1 if the wage in the poaching firm u is greater than the wage in the current firm z, and 0 otherwise. The equation $f(w(u))w'(u) = \frac{\partial F(w(u))}{\partial u}$ represents the density function of vacancies in poaching firms, where $f(w(u))$ is the density of wage offers and $F(w(u))$ is the cumulative distribution function of wage offers. Taking the first-order condition of the above equation with regard to h' and multiplying by $(1+\rho)$ (as we consider the next-period value), we can obtain $MR_W(z)$.

The firm's present value of income is:

$$\underbrace{(1-\beta(z))Az}_{\text{current profits}} - \underbrace{\mu_F\left(Az + c_s\bar{w}\right)}_{\text{firm's per-unit training costs}} \times \underbrace{s}_{\text{training level}} +$$

$$\frac{1}{1+\rho} \times \underbrace{(1-\delta)\left[1-\eta\bar{F}(w(z))\right](1-\beta(z))\,Az}_{\text{future profits, from workers who stay}} \times \underbrace{h'}_{\text{next-period human capital}}.$$

Taking the first-order condition of the above equation with regard to h' and multiplying by $(1+\rho)$ (as we consider the next-period value), we can obtain $MR_F(z)$.

C.2 Alternative training decision rules

We now consider how the analytical model's predictions change under alternative assumptions about how training investments are determined and funded.

For this, consider an analogous setup to that of the analytical model in Section 4, but where there are four possible cases concerning how training is determined and its costs

allocated:[31]

1. Firms bear all training costs and fully determine training, which follows Acemoglu and Pischke (1998).
2. Workers bear all costs and fully determine training, as typical in the literature studying post-schooling human capital accumulation (Ben-Porath, 1967; Manuelli and Seshadri, 2014).
3. Firms and workers choose training levels to maximize the joint match value, and the shares of the training costs allocated to each party correspond to the shares of this value they each receive (Acemoglu, 1997; Moen and Rosen, 2004).
4. As in the main text, firms and workers each pay constant shares of training costs, μ_F and μ_W, with the actual training level determined by the party with the lower desired training level.

Table C.1 presents the optimal training levels under the different scenarios of training determination. We now describe each of the scenarios respectively.

Table C.1: Optimal training levels in different scenarios

Scenario	Share of training costs paid		Optimal training level
	Firm ($\mu_F(z)$)	Worker ($\mu_W(z)$)	
1. Firms pay	1	0	$\left(\frac{\zeta\gamma_s L(z)^\alpha MR_F(z)}{(1+r)(c_s\bar{w}+Az)}\right)^{\frac{1}{1-\gamma_s}}$
2. Workers pay	0	1	$\left(\frac{\zeta\gamma_s L(z)^\alpha MR_W(z)}{(1+r)(c_s\bar{w}+Az)}\right)^{\frac{1}{1-\gamma_s}}$
3. Maximize match value	$\frac{MR_F(z)}{MR_W(z)+MR_F(z)}$	$\frac{MR_W(z)}{MR_W(z)+MR_F(z)}$	$\left(\frac{\zeta\gamma_s L(z)^\alpha (MR_W(z)+MR_F(z))}{(1+r)(c_s\bar{w}+Az)}\right)^{\frac{1}{1-\gamma_s}}$
4. Constant cost shares	μ_F	μ_W	$\min\left\{\left(\frac{\zeta\gamma_s L(z)^\alpha MR_F(z)}{(1+r)\mu_F(c_s\bar{w}+Az)}\right)^{\frac{1}{1-\gamma_s}}, \left(\frac{\zeta\gamma_s L(z)^\alpha MR_W(z)}{(1+r)\mu_W(c_s\bar{w}+Az)}\right)^{\frac{1}{1-\gamma_s}}\right\}$

Training determined by firms. When the firm pays all training costs, or when the firm pays a large enough share of the training costs in the scenario with constant cost shares ($\mu_F > \mu_W \frac{MR_F(z)}{MR_W(z)}$), training levels are chosen by the firm to maximize their returns net of costs which implies that the level of training is given by

[31] Note that in general, training costs are allocated between workers and firms via two components: implicit or wage-cut costs, which arise since firms can offload some of the expected training costs by posting lower wages, and explicit (or out-of-pocket) costs, which arise after the match is formed. We examine the allocation of the explicit cost component.

$$s(z) = \left(\frac{\zeta \gamma_s L(z)^{\alpha} MR_F(z)}{(1+r)\mu_F(c_s \bar{w} + Az)} \right)^{\frac{1}{1-\gamma_s}} = \left(\underbrace{\frac{\zeta \gamma_s L(z)^{\alpha}}{(1+r)\mu_F(c_s \bar{w} + Az)}}_{\text{returns \& costs from training}} \underbrace{(1-\delta)\left[1-\eta \bar{F}(w(z))\right]}_{\text{prob of keeping worker}} \underbrace{(1-\beta(z))}_{\text{1-labor share}} \underbrace{Az}_{\text{productivity}} \right)^{\frac{1}{1-\gamma_s}}. \tag{C.3}$$

This expression matches that discussed in the main text.

Training determined by workers. When the worker pays all training costs, or when the worker pays a large enough share of the training costs in the scenario with constant cost shares ($\mu_W > \mu_F \frac{MR_W(z)}{MR_F(z)}$), training levels are chosen by the worker to maximize their returns net of costs, which implies that the level of training is given by

$$\begin{aligned} s(z) &= \left(\frac{\zeta \gamma_s L(z)^{\alpha} MR_W(z)}{(1+r)\mu_W(c_s \bar{w} + Az)} \right)^{\frac{1}{1-\gamma_s}} \\ &= \Bigg[\underbrace{\frac{\zeta \gamma_s L(z)^{\alpha}}{(1+r)\mu_W(c_s \bar{w} + Az)}}_{\text{returns \& costs from training}} \Bigg(\underbrace{(1-\delta)\left[1-\eta \bar{F}(w(z))\right]}_{\text{prob of worker staying in firm}} \underbrace{\beta(z)Az}_{\text{worker's share of value}} + \underbrace{(1-\delta)\eta \int \mathbf{1}_{\{w(u)>w(z)\}} \beta(u) Auf(w(u))w'(u)du}_{\text{expected worker's share of value after J-J transitions}} \\ &\quad + \underbrace{\delta \int \beta(u) Auf(w(u))w'(u)du}_{\text{worker's share of value after exogenous separations}} \Bigg) \Bigg]^{\frac{1}{1-\gamma_s}}. \end{aligned} \tag{C.4}$$

This equation indicates that the higher the share of match value captured by the worker, either through the labor share at the current firm or in outside options, the greater the worker's incentive to invest in human capital. The worker's chosen training level also increases with the productivity of both the current and potential employers, as higher productivity raises the returns to human capital investment.

Training determined to maximize the match value. When the training level is chosen to maximize the joint match value of workers and firms, the level of training is given by

$$\begin{aligned} s(z) &= \left(\frac{\zeta \gamma_s L(z)^{\alpha} (MR_W(z) + MR_F(z))}{(1+r)(c_s \bar{w} + Az)} \right)^{\frac{1}{1-\gamma_s}} \\ &= \Bigg[\underbrace{\frac{\zeta \gamma_s L(z)^{\alpha}}{(1+r)(c_s \bar{w} + Az)}}_{\text{returns \& costs from training}} \Bigg(\underbrace{(1-\delta)\left[1-\eta \bar{F}(w(z))\right]}_{\text{prop of keeping match}} \underbrace{Az}_{\text{firm's+worker's share of value}} + \underbrace{(1-\delta)\eta \int \mathbf{1}_{\{w(u)>w(z)\}} \beta(u) Auf(w(u))w'(u)du}_{\text{expected worker's share of value after J-J transitions}} \\ &\quad + \underbrace{\delta \int \beta(u) Auf(w(u))w'(u)du}_{\text{worker's share of value after exogenous separations}} \Bigg) \Bigg]^{\frac{1}{1-\gamma_s}}. \end{aligned} \tag{C.5}$$

In this scenario, training reflects both the firm's and the worker's share of the match value, as well as the value workers expect to receive following job-to-job or exogenous transitions. Because gains to both parties are taken into account, the resulting training level is higher than in cases where training is determined by only one side.

C.2.1 Confronting alternate scenarios with empirical evidence

Armed with these results, we now confront the model's training predictions under each scenario with our empirical findings. Although firm productivity, firm size, and the labor share are jointly determined in the model, we consider their effects separately, mirroring our empirical strategy of including all three variables in the same regression to isolate their individual contributions.[32]

Effect of productivity on training. First, when firms determine the training level, and holding firm size and labor share constant, training rises with firm productivity. This stems from the supermodularity of the production function which increases the returns to human capital acquisition in more productive firms, but also from an alleviation of the hold-up problem in these settings. Since wages increase with firm productivity, workers will be less likely to be poached when working in more productive firms, encouraging these firms to increase training investments.

By contrast, when the worker chooses the training level, or when training maximizes joint match value, training may decrease with firm productivity. In particular, although the returns from training also increase with productivity in these cases due to the supermodularity in production, this increase is slow since workers' returns from training incorporate the benefits after leaving the firm. As such, this increase may not be fast enough to compensate workers for the loss of compensated time when training, which is relatively higher in more productive firms, and may therefore cause training to decrease with productivity in these scenarios.[33] Thus, when workers choose the training level, whether training increases with z or not is dictated by the shape of the firm productivity distribution governing the benefits from leaving the firm, and the size of the loss of time in training.

Effect of firm size on training. Second, in all cases, if $\alpha > 0$, training increases with firm size holding firm productivity and the labor share constant. This reflects that the returns to training in human capital accumulation rise with firm size and that the hold-up problem is less severe for larger firms.

[32] The formal mathematical derivation of these results is available upon request.

[33] This opportunity cost of training is also present when the firm chooses training, but is not enough to cause training to decrease in this case.

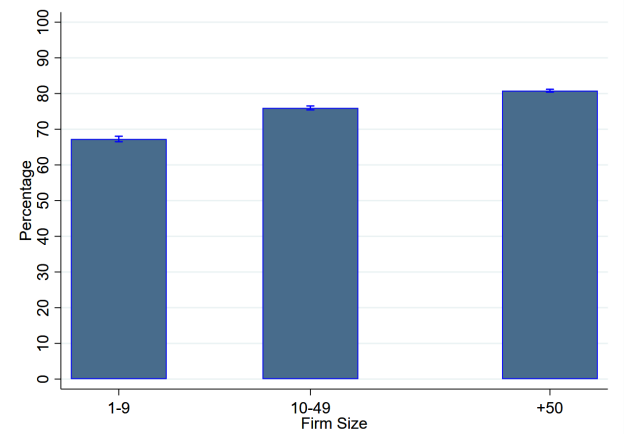
Effect of the labor share on training. Third, when firms determine the training level, and holding the productivity level and firm size constant, training decreases with the labor share. This arises because the labor share is inversely correlated with the returns the firm receives from the match. In the case where the worker chooses the training level, a higher labor share will induce higher learning investments since the worker will reap a higher portion of the match value. In the case of joint internal efficiency, the labor share matters only through its influence on job-to-job transitions. In particular, in this case a higher labor share also increases optimal training level, as it reduces the incidence of job-to-job transitions which reduce the benefits firms' receive from training.

Summary These results suggest that only cases in which firms determine the level of training are consistent with our empirical findings.

C.3 Further evidence on importance of firms for training decisions and financing

We use worker-level data from the Adult Education Surveys conducted in the EU in 2011 and 2016 (EU-AES), which contain information on both whether workers engaged in training during the previous year and how this training was financed, to document that firms fully or partly finance a majority of training activities and thus play a pivotal role in training decisions. Further details on the EU-AES are provided in Appendix A.8.

Figure C.1: Share of trained workers reporting employer-financed training by firm size



Notes: We plot the share of workers who report that their training was financed at least in part by their employer, among those who participated in training during the previous 12 months, by firm size category. Data comes from European Union Adult Education Surveys (EU-AES) conducted in 2011 and 2016, and is weighted using the weights provided in the survey. 95% confidence intervals are plotted.

In Figure C.1, we plot the share of workers who report that their training was financed at least in part by their employer, among those who participated in training during the previous 12 months, by firm size category. We find that across firms of all sizes, over 65% of workers who participated in training during the previous year report that their employer paid for their training expenses at least in part.

This evidence underscores the central role of firms in financing training and is consistent with our modeling assumption that firms are key decision-makers in training investments.

D Additional results from quantitative model

D.1 Conditions for simulations

Workers' value With linear utility, workers' utility is determined by the discounted income flows that are earned with current and future human capital accumulation. We denote the value for a worker of age a and human capital h in a firm with productivity z as $W^a(h,z)$. We denote the value of an unemployed worker by $W^a(h)$, and following Bagger et al. (2014), we assume that unemployment is equivalent to employment in the least productive firm: $W^a(h) = W^a(h, z_{\min})$. This assumption solves the complication of allowing for heterogeneous reservation wages for workers of different human capital levels and ages. With $\theta = \frac{V}{\tilde{U}}$ capturing the labor market tightness, we denote $q(\theta) = \frac{M}{V}$ as the vacancy-filling rate and $\frac{M}{\tilde{U}} = q(\theta)\theta$ as the job-finding rate.

First, note that in the last period of their lifetime ($a = J$), workers have no incentive to accumulate human capital. Thus, we can obtain $W^J(h,z) = w(z)h$. For younger workers ($a < J$), we can solve for their respective values through backward induction:

$$W^a(h,z) = \underbrace{w(z)h - \mu_W(c_s\bar{w} + Az)s^a(h,z)}_{\text{wage income net of training costs}} + \underbrace{\frac{\delta}{1+\rho}W^{a+1}(h')}_{\text{value if being separated exogenously in the next period}}$$
$$+ \underbrace{\frac{1-\delta}{1+\rho}\left[(1-\eta\theta q(\theta))\,W^{a+1}(h',z) + \eta\theta q(\theta)\int \max\{W^{a+1}(h',u), W^{a+1}(h',z)\}dF(w(u))\right]}_{\text{value if staying or transitioning from job to job in the next period}}.$$

The term h', given by $h' = e^a(h,z) = 1 + (1-d)(h-1) + \zeta L^\alpha(z)(s^a(h,z))^{\gamma_s}$, denotes human capital in the next period, and $s^a(h,z)$ is the optimal training level as described below. To simplify our notation, we use $e^a(h,z)$ to represent workers' human capital evolution. We

use $e_U(h) = 1 + (1-d)(h-1)$ to denote the next-period's human capital for unemployed workers.

Firms' value The firm's value of matching with a worker of age a and human capital level h (after hiring) is denoted by $F^a(h,z)$, and follows:

$$F^a(h,z) = \underbrace{(Az - w(z))\, h - \mu_F\,(Az + c_s\bar{w})\, s^a(z,h)}_{\text{revenue net of wage and training costs}}$$
$$+ \underbrace{\frac{1-\delta}{1+\rho}\left(1 - \eta\theta q(\theta)\int \mathbf{1}_{\{W^{a+1}(h',u) > W^{a+1}(h',z)\}} dF(w(u))\right) F^{a+1}(h',z)}_{\text{value if the worker stays in the firm in the next period}},$$

where $h' = e^a(h,z)$ is defined as above.

Employment distribution We use $N^a(h,z)$ to denote the measure of workers of age a and human capital h in all firms with productivity z right before job search happens. Similarly, we use $U^a(h)$ to denote the measure of unemployed workers right before job search happens. The number of searchers is the sum of the unemployed and on-the-job searchers,

$$\tilde{U} = \sum_{a=1}^{J}\left[\int U^a(h)dh + \eta\int\int N^a(h,z)dhdz\right].$$

For the entering cohort endowed with human capital level of 1, the number of unemployed searchers is given by $U^1(1) = 1$ and $U^1(h) = 0\ \forall\ h > 1$, with null existing employment $N^1(h,z) = 0\ \forall\ h$.

The following equations characterize the evolution of these measures, accounting for human capital formation, job search, and exogenous and endogenous job separations,

$$N^{a+1}(h',z) = \underbrace{(1-\delta)\int_{h'=e^a(h,z)}\left[1 - \eta\theta q(\theta)\int \mathbf{1}_{\{W^a(h,u) > W^a(h,z)\}} dF(w(u))\right] N^a(h,z)dh}_{\text{workers that stay in the last-period job search and are not exogenously separated this period}}$$
$$+ \underbrace{(1-\delta)\theta q(\theta) f(w(z)) w'(z)\left[\int_{h'=e^a(h,z)} U^a(h)dh + \eta\int_{h'=e^a(h,z)}\int N^a(h,y)\mathbf{1}_{\{W^a(h,z) > W^a(h,y)\}} dydh\right]}_{\text{last-period hires that are not exogenously separated this period}};$$

$$U^{a+1}(h') = \left(\underbrace{\int \frac{\delta}{1-\delta} N^{a+1}(h', z) dz}_{\text{exog separations this period}} + \underbrace{(1 - \theta q(\theta)) \int_{h'=e_U(h)} U^a(h) dh}_{\text{last-period unemployed searchers still w/o jobs}} \right).$$

Training According to the first-order condition of the firm's value, the firms' optimal training level, $s_F^a(h, z)$ is given by

$$s_F^a(h, z) = \left(\frac{\zeta \gamma_s L^\alpha(z)}{\mu_F \left(Az + c_s \bar{w} \right)} \frac{\partial F^a(h, z)}{\partial h'} \right)^{1/(1-\gamma_s)},$$

where $\frac{\partial F^a(h,z)}{\partial h'}$ captures the firms' return of an extra efficiency unit of human capital in the next period, which can be derived as:

$$\frac{\partial F^a(h, z)}{\partial h'} = \sum_{a'=a+1}^{J} (1-d)^{a'-a-1} \left\{ \prod_{\tau=a}^{a'-1} \left[\frac{1-\delta}{1+\rho} \left(1 - \eta \theta q(\theta) \int \mathbf{1}_{\{W^{\tau+1}(h^{\tau+1}, u) > W^{\tau+1}(h^{\tau+1}, z)\}} dF(w(u)) \right) \right] \right\} \times Az \times \frac{Az - w(z)}{Az}$$

where $h^{\tau+1} = e^\tau(h^\tau, z)\ \forall\ \tau \geq a$ with $h^a = h$. Thus, $s_F^a(h, z)$ can be decomposed as:

$$\underbrace{\log s_F^a(h, z)}_{\text{training time in firm } z \text{ for age } a} = \underbrace{\frac{1}{1-\gamma_s} \log \frac{\zeta \gamma_s}{\mu_F}}_{\text{constant}} + \underbrace{\frac{1}{1-\gamma_s} \log \frac{L(z)^\alpha}{Az + c_s \bar{w}}}_{\text{returns \& costs from training}} + \underbrace{\frac{1}{1-\gamma_s} \log(1 - \beta(z))}_{\text{1-labor share}} + \underbrace{\frac{1}{1-\gamma_s} \log Az}_{\text{productivity}}$$
$$+ \underbrace{\frac{1}{1-\gamma_s} \log \sum_{a'=a+1}^{J} (1-d)^{a'-a-1} \left\{ \prod_{\tau=a}^{a'-1} \left[\frac{1-\delta}{1+\rho} \left(1 - \eta \theta q(\theta) \int \mathbf{1}_{\{W^{\tau+1}(h^{\tau+1}, u) > W^{\tau+1}(h^{\tau+1}, z)\}} dF(w(u)) \right) \right] \right\}}_{\text{probability of keeping workers}} \tag{D.1}$$

where $\beta(z) = \frac{w(z)}{Az}$ is the labor share. This resembles the expression in Equation C.3, except that it accounts for multiple periods.

The workers' optimal training level, $s_W^a(h, z)$, is given by

$$s_W^a(h, z) = \left(\frac{\zeta \gamma_s L^\alpha(z)}{\mu_W \left(Az + c_s \bar{w} \right)} \frac{\partial W^a(h, z)}{\partial h'} \right)^{1/(1-\gamma_s)},$$

where $\frac{\partial W^a(h,z)}{\partial h'}$ captures the workers' return of an extra efficiency unit of human capital in the next period.

The optimal training level is given by $s^a(h, z) = \min\{s_W^a(h, z), s_F^a(h, z)\}$.

Vacancy and wage determination Each firm maximizes its total value from hiring by choosing the number of vacancies $v(z)$ and wage rate $w(z)$:

$$\max_{v(z),w(z)} \underbrace{\sum_{a=1}^{J} \frac{q(\theta)}{\tilde{U}} \left[\eta \int \int \mathbf{1}_{\{W^a(h,z)>W^a(h,y)\}} N^a(h,y) F^a(h,z) dy dh + \int U^a(h) F^a(h,z) dh \right]}_{\text{benefits from hiring on-the-job or unemployed searchers by posting a vacancy}} v(z) - \underbrace{c_v \bar{w} \frac{v(z)^{1+\gamma_v}}{1+\gamma_v}}_{\text{vacancy costs}}.$$

The optimality condition for vacancies is given by:

$$\underbrace{c_v \bar{w} v(z)^{\gamma_v}}_{\text{marginal costs of a vacancy}} = \underbrace{\sum_{a=1}^{J} \frac{q(\theta)}{\tilde{U}} \left[\eta \int \int \mathbf{1}_{\{W^a(h,z)>W^a(h,y)\}} N^a(h,y) F^a(h,z) dy dh + \int U^a(h) F^a(h,z) dh \right]}_{\text{marginal benefits from hiring on-the-job or unemployed searchers by posting a vacancy}}.$$

The differential equation of wages can be obtained by totally differentiating the right-hand side of the above equation with regard to $w(z)$, as firms choose wages to maximize the value of each vacancy:

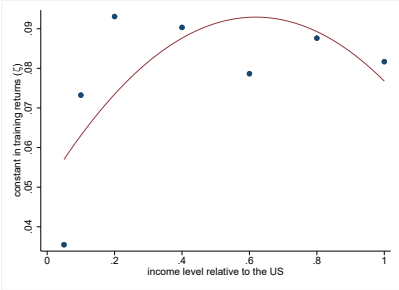
$$\sum_{a=1}^{J} \frac{q(\theta)}{\tilde{U}} \left[\eta \int \int \mathbf{1}_{\{W^a(h,z)>W^a(h,y)\}} N^a(h,y) \frac{\partial F^a(h,z)}{\partial w(z)} dy dh + \eta \int \int \frac{\partial \mathbf{1}_{\{W^a(h,z)>W^a(h,y)\}}}{\partial w(z)} N^a(h,y) F^a(h,z) dy dh \right.$$
$$\left. + \int U^a(h) \frac{\partial F^a(h,z)}{\partial w(z)} dh \right] = 0.$$

This differential equation can be evaluated numerically. Combined with the lowest wage $b\bar{w}$, we can iterate the wage structure $w(z)$ until convergence. When the model abstracts from human capital, the wage differential equation can be analytically written in a similar way as in Burdett and Mortensen (1998).

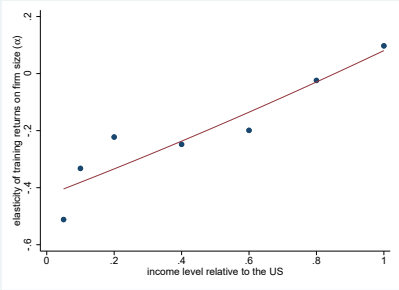
D.2 Cross-country calibration

Figure D.1: Cross-country calibration results

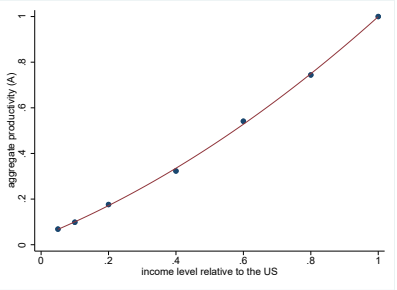
(a) Constant in training returns



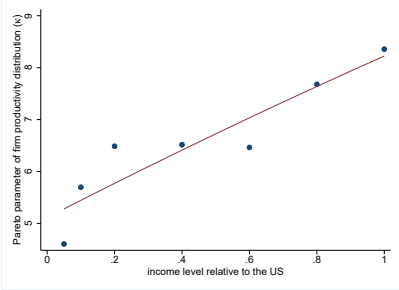
(b) Elasticity of returns to size



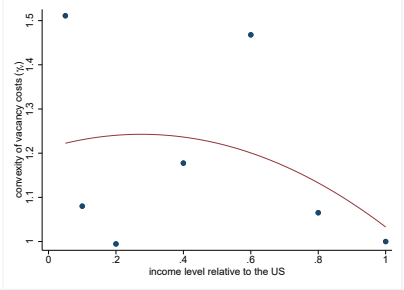
(c) Aggregate productivity



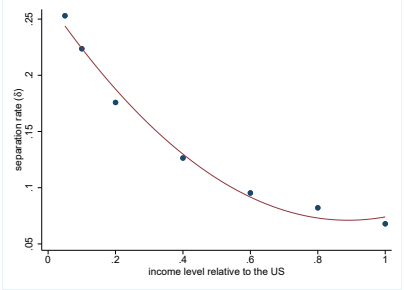
(d) Pareto tail of productivity dist



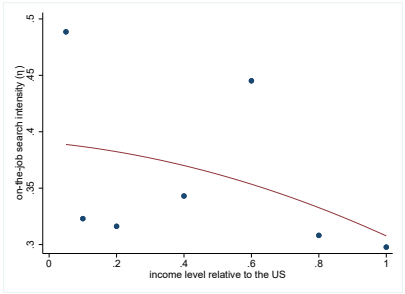
(e) Vacancy convexity



(f) Separation rate



(g) On-job-search intensity



Notes: This figure plots the calibrated parameter values for different income levels (5%, 10%, 20%, 40%, 60%, 80%, and 100% of the U.S. level) in the model.